ORIGINAL ARTICLE

Lung Cancer Incidence and Mortality with Extended Follow-up in the National Lung Screening Trial

The National Lung Screening Trial Research Team*

Received 21 July 2018; revised 29 May 2019; accepted 31 May 2019
Available online - 28 June 2019

ABSTRACT

Introduction: The National Lung Screening Trial (NLST) randomized high-risk current and former smokers to three annual screens with either low-dose computed tomography (LDCT) or chest radiography (CXR) and demonstrated a significant reduction in lung cancer mortality in the LDCT arm after a median of 6.5 years' follow-up. We report on extended follow-up of NLST subjects.

Methods: Subjects were followed by linkage to state cancer registries and the National Death Index. The number needed to screen (NNS) to prevent one lung cancer death was computed as the reciprocal of the difference in the proportion of patients dying of lung cancer across arms. Lung cancer mortality rate ratios (RRs) were computed overall and adjusted for dilution effect, with the latter including only deaths with a corresponding diagnosis close enough to the end of protocol screening.

Results: The median follow-up times were 11.3 years for incidence and 12.3 years for mortality. In all, 1701 and 1681 lung cancers were diagnosed in the LDCT and CXR arms, respectively (RR = 1.01, 95% confidence interval [CI]: 0.95–1.09). The observed numbers of lung cancer deaths were 1147 (with LDCT) versus 1236 (with CXR) (RR = 0.92, 95% CI: 0.85–1.00). The difference in the number of patients dying of lung cancer (per 1000) across arms was 3.3, translating into an NNS of 303, which is similar to the original NNS estimate of around 320. The dilution-adjusted lung cancer mortality RR was 0.89 (95% CI: 0.80–0.997). With regard to overall mortality, there were 5253 (with LDCT) and 5366 (with CXR) deaths, for a difference across arms (per 1000) of 4.2 (95% CI: –2.6 to 10.9).

Conclusion: Extended follow-up of the NLST showed an NNS similar to that of the original analysis. There was no overall increase in lung cancer incidence in the LDCT arm versus in the CXR arm.

Keywords: Lung cancer; Screening; Low-dose CT; Incidence; Mortality

**A complete list of members of the National Lung Screening Trial Research Team is provided in the Appendix.*

Disclosure: Dr. Aberle reports grants from American College of Radiology Imaging Network (ACRIN), grants from American College of Radiology Imaging Network (ACRIN), during the conduct of the study; other from 2016 Veracyte Advisory Board, non-financial support from 2016: Moffitt Cancer Center, grants from 2016: Kaiser Radiology Symposium, grants and other from 2016: MCL Consortium Meeting, non-financial support from 2016: ECOG-ACRIN Semi Annual Meeting, grants from 2016 DECAMP Consortium Meeting, other from 2016: American Society of Clinical Oncology (ASCO) Annual Meeting, non-financial support from 2016: National Academy of Sciences Workshop, non-financial support from 2016: Siemens Annual UCLA Research Meeting, personal fees from 2016: Veracyte Advisory Board Meeting, personal fees and non-financial support from 2016: International Lung Cancer Congress (ILCC) Meeting, grants and non-financial support from 2016: MCL Consortium Meeting, non-financial support from 2016: Harvard University Chest Imaging Course, personal fees from 2016: Torrance Medical Center Grand Rounds, personal fees and non-financial support from 2016: Yale Grand Rounds, CT, other from 2016: ASCO, non-financial support from 2016: NCI Workshop, Washington DC, grants and other from 2017: MCL Houston, personal fees and non-financial support from 2017: National Jewish Health, other from 2017: World Congress Thoracic Imaging, non-financial support from 2017: American College of Radiology, other from 2017: EDRN Steering Committee Meeting, personal fees from 2017: NIH Section Review, grants from 2017: Kaiser Watch the spot meeting, Pasadena, other from 2018: Grand Rounds WLAVA, personal fees from 2018: NIH Section Review, San Diego, grants from 2018: DECAMP 2018 Meeting, personal fees from 2018: International Symposium on "Clinical update in Respiratory Medicine", grants from 2018: EDRN, Bethesda, grants from 2018: MCL, Bethesda, non-financial support from 2018: AIMBE College of Fellows, non-financial support from 2018: Cleveland Clinic Visiting Professor, non-financial support from 2018: SPORE Workshop - Lung Cancer, non-financial support from 2018: IASLC, Toronto, grants from 2018: MCL Steering Committee, personal fees from 2018: Cancer Research UK (CRUK), other from 2018: RSNA, outside the submitted work. Dr. Sicks reports a grant to her institution from the National Cancer Institute (grant U10CA18-820-0151) during the study. Dr. Chiles reports a grant to her institution from the National Institutes of Health (grant CA 80098) during the study. The remaining authors declare no conflict of interest.

Address for correspondence: Paul Pinsky, PhD, National Cancer Institute, 9609 Medical Center Dr., Bethesda, MD 20892. E-mail: pp4f@nih.gov

Published by Elsevier Inc. on behalf of International Association for the Study of Lung Cancer.

ISSN: 1556-0864

https://doi.org/10.1016/j.jtho.2019.05.044

Introduction

Lung cancer is the leading cause of cancer death worldwide.[1] Early detection and treatment through screening with low-dose computed tomography (LDCT) has been investigated as a potential means of reducing lung cancer deaths for more than two decades.[2,3] In 2011, a large U.S. study, the randomized National Lung Screening Trial (NLST), reported a significant (20%) reduction in lung cancer mortality in high-risk current and former smokers who were screened annually (three times) with LDCT versus with chest radiography (CXR).[4,5] Other small randomized trials, primarily in Europe, have reported mixed results in terms of a reduction in lung cancer mortality but were substantially underpowered.[6-10] Recently, the other large LDCT screening trial, NELSON (in Europe), reported preliminarily on its findings. Through a 10-year study follow-up period and after four rounds of LDCT screening, NELSON reported a 26% reduction in lung cancer mortality in men (risk ratio = 0.74, 95% confidence interval [CI]: 0.60–0.91) and a 39% reduction in women (risk ratio = 0.61, 95% CI: 0.35–1.04) in the LDCT arm versus in the control (nonscreening) arm.[11]

The median follow-up in the NLST as originally reported was 6.5 years, or about 4.5 years after the final scheduled screen.[5] After the original trial report, an extended follow-up study of the NLST cohort was undertaken, utilizing passive linkages to state cancer registries and the National Death Index (NDI). An additional 5 years of data are now available for lung cancer incidence, and an additional 6 years of data are available for mortality.

The primary objective of the NLST extended follow-up study was to ascertain whether the originally reported reduction in lung cancer mortality in the LDCT arm versus in the CXR arm was maintained. With a follow-up of 4 to 5 years after the final screen in the original report, it is possible that earlier detection with LDCT only delayed lung cancer death instead of preventing it. With 6 additional years of mortality follow-up, it is now possible to observe whether lung cancer deaths were in fact prevented by LDCT screening (at least for a decade) rather than merely delayed. A secondary objective of the extended follow-up was to further assess overdiagnosis in the trial. A modest but statistically significant increase in lung cancer incidence in the LDCT arm, possibly signaling overdiagnosis, was observed with the original follow-up period.[12] With longer follow-up, it is of interest to see whether this increase has been preserved.

There are potential issues, however, with examining the extended follow-up data for lung cancer mortality. With follow-up now well beyond the period of trial screening, there is the potential, or even likelihood, of some dilution of the screening effect.[13-16] Specifically, patients in whom cancer did not develop until after the last scheduled screen might not have benefited from the trial screenings; therefore, deaths in such patients would only serve to add noise to the estimates, roughly an equal number of deaths in each arm. Therefore, in analyzing these data, we have used various methods that attempt to control for a dilution effect, including examining the difference in lung cancer deaths across arms in addition to the rate ratio (RR) and examining the RR adjusted for dilution by considering time of diagnosis.[13-15] This latter method, which is well known in the mammography screening trial literature, includes only those cancer deaths for which the corresponding time of cancer diagnosis is close enough to the end of protocol screening in the trial.

Methods

A more detailed description of the NLST has been published previously.[4] Briefly, men and women aged 55 to 74 years who had a history of a minimum of 30 pack-years of cigarette smoking and who were either current smokers or had quit within the past 15 years were enrolled from 2002 to 2004 at 33 medical institutions across the United States. Exclusion criteria included a previous lung cancer diagnosis, a computed tomography (CT) scan in the prior 18 months, unexplained weight loss in the year before enrollment, and hemoptysis. Participants were randomized to an LDCT or single-view CXR arm, with three annual protocol screens for each modality.

Participants were actively followed for lung cancer incidence and all-cause mortality until December 31, 2009. During that time, medical records were abstracted for those with a positive screening test result or lung cancer diagnosis. Vital status was assessed through annual or semiannual questionnaires and by linkage with the NDI. Institutional review boards at each center approved the study, and each person provided written consent to participate in the study.

After the active follow-up period, participants were followed only passively through linkages with state cancer registries and the NDI. Linkages were performed by each participating registry and the NDI by using probabilistic linkage methods. Linkages were conducted with cancer registries in the state of the screening center (the center's "home state" registry) as well as in some neighboring states. For logistical reasons, not all home state registries participated in the linkage effort. In addition, some screening centers did not have participants' names available for linkage purposes, which precluded performing registry linkage for some registries. All centers

but one were able to link with the NDI. The personally identifiable information that the NLST had available for linkage included Social Security number, full name (for some screening centers), date of birth, and sex.

For centers with home state cancer registry linkage (22 of 33, comprising 87.6% of trial participants), lung cancer incidence follow-up was through the end of 2014; otherwise, it was through the end of 2009. Mortality follow-up was through the end of 2015 for centers with NDI linkage (comprising 97.8% of trial participants) and through the end of 2009 for the one center without NDI linkage. See Supplementary Table 1 for a summary of linkage efforts by screening center. For assessing mortality due to lung cancer, deaths in the original analysis period were evaluated by a death review panel.[4] For the current analysis, the death panel classification was used for those deaths, whereas the underlying cause of death from the NDI linkage was used for subsequent deaths.

Quantitative Methods

Rates (lung cancer incidence, lung cancer mortality, and all-cause mortality) were calculated as the number of events divided by the corresponding person-time; RRs were computed as the LDCT arm rate divided by the CXR arm rate. Person-time for incidence ended at the end of incidence follow-up, date of lung cancer diagnosis, or date of death, whichever came first. Person-time for mortality ended at the end of mortality follow-up or death, whichever came first. In addition to rates and RRs, for each event type we computed the proportion of subjects in each arm with the event and the difference in those proportions across arms. Note that unlike the RR, the expected difference in proportions is not affected by dilution because an equal number of events in each arm occurring beyond the time when screening could have an effect cancel each other out on average. The number needed to screen (NNS) to prevent one lung cancer death was calculated as the reciprocal of the difference in the proportion dying of lung cancer across arms. Potential interactions of several risk factors with trial arm—specifically, age, sex, and smoking status (current versus former smoker)—were assessed by using Poisson regression. The distribution of lung cancer cases by histologic type and stage was analyzed using chi-square tests. The overdiagnosis rate was calculated as the difference in lung cancer cases across arms divided by the number of LDCT screen–detected cases.

Analysis Adjusted for Dilution

To derive the dilution-adjusted lung cancer mortality RR, the cutoff time for cancer diagnosis must first be determined; only those lung cancer deaths (in each arm) for which the corresponding diagnosis is before this cutoff time are included in the RR computation. One proposed method is to assess when in study time the cumulative cancer incidence across arms first becomes equalized.[13,14] If screening results in overdiagnosis, incidence will never become equalized across arms. In the NLST, although there was overdiagnosis based on the original data, most of the overdiagnosed cases were identifiable by histologic type.[12] Almost all cases classified as bronchioloalveolar carcinoma (BAC) were overdiagnosed, and BAC represented most of the overdiagnosed cases.[12] Therefore, to define the cutoff time for the dilution-adjusted analysis, incidence across arms was examined by study year, excluding BAC cases, and the cutoff time was defined as the (end of the) earliest study year for which there was no significant difference in cumulative incidence across arms. As a sensitivity analysis, we also examined dilution-adjusted RRs using alternative study year cutoff times.

Analyses Using Calendar Time versus Study Time

Lung cancer mortality results for the NLST based on the original data were first reported with use of a cutoff date of January 15, 2009, in accordance with the interim analysis plan and to account for time lags associated with the end point verification process; a cutoff of December 31, 2009, was used for all-cause mortality.[5] Lung cancer mortality results were subsequently reported using all events through the later cutoff date (December 31, 2009).[17] Because subjects were enrolled in the NLST over roughly a 2-year period, these calendar time cutoffs resulted in a range of times on study for the original analysis, with medians of 5.5 years (interquartile range 5.2–5.9) and 6.5 years (interquartile range 6.1–6.9) for the earlier and later dates, respectively. From a scientific standpoint, analyses based on study time cutoffs, in which all subjects have essentially the same time on study, are more meaningful because they allow assessment of all events within a given time after randomization and protocol screens. The extended follow-up data allow us now to compute lung cancer mortality RRs based on study time cutoffs with median follow-up times similar to those in the original analyses.

Results

In total, 26,722 and 26,730 participants were randomized to the LDCT and CXR arms, respectively. Baseline participant demographics and smoking history were similar across arms (Table 1).

The median follow-up times for incidence and mortality were similar across arms. For incidence, the median (25th percentile/75th percentile) follow-up times were 11.3 years (9.0/11.8 years) in the LDCT arm and 11.3 years (8.9/11.8 years) in the CXR arm; for mortality, the

median (25th percentile/75th percentile) follow-up time was 12.3 years (11.9/12.8 years) in each arm.

Lung Cancer Incidence

Figure 1*A* and *B* shows cumulative lung cancer incidence by arm. There were 1701 lung cancer cases in the LDCT arm versus 1681 in the CXR arm, giving an RR of 1.01 (95% CI: 0.95–1.09) (see Fig. 1*A*). For all cases excluding BAC, the RR was slightly less than 1 (RR = 0.97, 95% CI: 0.90–1.04), whereas there was a significant increase in BAC cases in the LDCT arm (RR = 2.6, 95% CI: 1.9–3.7) (see Fig. 1*B*). As seen in Figure 1*A*, the excess cumulative number of cases in the LDCT arm versus in the CXR arm peaks around year 3, the end of the screening phase of the trial, and declines thereafter. The overall lung cancer rates per 10,000 person-years were 63.8 and 62.9 in the LDCT and CXR arms, respectively. The overdiagnosis rates were 3.1% (20 of 649) overall and 79% (75 of 95) for BAC.

Lung Cancer Characteristics

Table 2 shows the distribution of histologic type and stage by arm. With the exception of BAC, the histologic type distribution was generally similar across arms. In terms of stage, a significantly higher proportion of LDCT arm cases versus CXR arm cases were stage I (39.6% versus 27.5% [$p < 0.0001$]) (excluding BAC, the stage I proportions were 37% versus 27% [$p < 0.0001$]). Conversely, a significantly lower proportion of cases in the LDCT arm versus in the CXR arm were stage IV (27.5% versus 35.5% [$p < 0.0001$]).

Lung Cancer Mortality, Stage IV Disease, and All-Cause Mortality

Table 3 shows lung cancer mortality rates across arms. There were 1147 deaths (42.9 per 1000 subjects) due to lung cancer in the LDCT arm versus 1236 (46.2 per 1000) in the CXR arm. The difference across arms (CXR minus LDCT) in the number of subjects (per 1000) dying of lung cancer was 3.3 (95% CI: −0.2 to 6.8, $p = 0.06$), which translates into an NNS of 303. The RR for lung cancer mortality was 0.92 (95% CI: 0.85–1.00, $p = 0.05$). The lung cancer mortality RR was lower for women (RR = 0.86) than for men (RR = 0.97), lower for current smokers (RR = 0.88) than for former smokers (R = 1.01), and lower for subjects aged 55 to 64 years at entry (RR = 0.86) than for those aged 65 to 74 years at entry (RR = 1.01) (see Table 3). However, the interactions of trial arm by sex and by smoking status were not statistically significant, indicating that there was no statistical difference in the RRs by sex or smoking status. The interaction of trial arm by age was borderline significant ($p = 0.051$).

Table 1. Demographics and Smoking History

Characteristic	LDCT (n = 26,722)	CXR (n = 26,730)
Men, n (%)	15,769 (59.0)	15,761 (59.0)
Women, n (%)	10,953 (41.0)	10,969 (41.0)
Non-Hispanic white, n (%)	23,953 (89.6)	23,949 (89.6)
Non-Hispanic black, n (%)	1187 (4.4)	1174 (4.4)
Hispanic, n (%)	479 (1.8)	456 (1.7)
Asian, n (%)	546 (2.0)	525 (2.0)
American Indian/Native Alaskan, n (%)	87 (0.4)	97 (0.4)
Native Hawaiian/Pacific Islander, n (%)	83 (0.3)	81 (0.3)
Other/unknown, n (%)	387 (1.5)	448 (1.7)
Current smoker, n (%)	12,860 (48.1)	12,900 (48.3)
Former smoker, n (%)	13,862 (51.9)	13,830 (51.7)
Median pack-years (25th percentile/ 75th percentile)	48 (39/66)	48 (39/66)
Age at randomization, y, n (%)		
55-59	11,442 (42.8)	11,423 (42.7)
60-64	8170 (30.6)	8199 (30.7)
65-69	4756 (17.8)	4761 (17.8)
70-74	2354 (8.8)	2347 (8.8)

LDCT, low-dose computed tomography; CXR, chest radiography.

For the analysis adjusted for dilution, the cumulative incidence RR across arms (excluding BAC) first became nonsignificant at study year 6, with an RR of 1.07 ($p = 0.13$). Therefore, for the dilution-adjusted analysis, only those deaths with diagnosis through study year 6 were included. There were 578 such lung cancer deaths in the LDCT arm versus 646 in the CXR arm, giving a lung cancer mortality RR of 0.89 (95% CI: 0.80–0.997, $p = 0.043$) (see Table 3). The difference in the number dying of lung cancer (per 1000) across arms based on the dilution-adjusted analysis was 2.5 (95% CI: 0.001–5.1, $p = 0.05$), giving an NNS of 394. A pattern similar to that in the overall analysis of the RR was observed, with the RR being lower in women (RR = 0.80) than in men (RR = 0.95), in current smokers (RR = 0.84) than in former smokers (RR = 0.99) and in younger (RR = 0.85) versus older (RR = 0.94) subjects, though none of these interactions with trial arm were statistically significant. Supplementary Table 2 shows cumulative incidence RRs for various alternative study time cutoffs and the corresponding dilution-adjusted RRs.

There were 468 stage IV cases in the LDCT arm versus 597 in the CXR arm, giving a RR of 0.79 (95% CI: 0.70–0.89) (see Table 3). Considering the same study period as in the dilution-adjusted analysis (through study year 6), there were 245 (LDCT) versus 344 (CXR) cases (RR = 0.72, 95% CI:0.61–0.84). There were no significant interactions by sex, age, or smoking status, either for all stage IV cases or for stage IV cases through year 6.

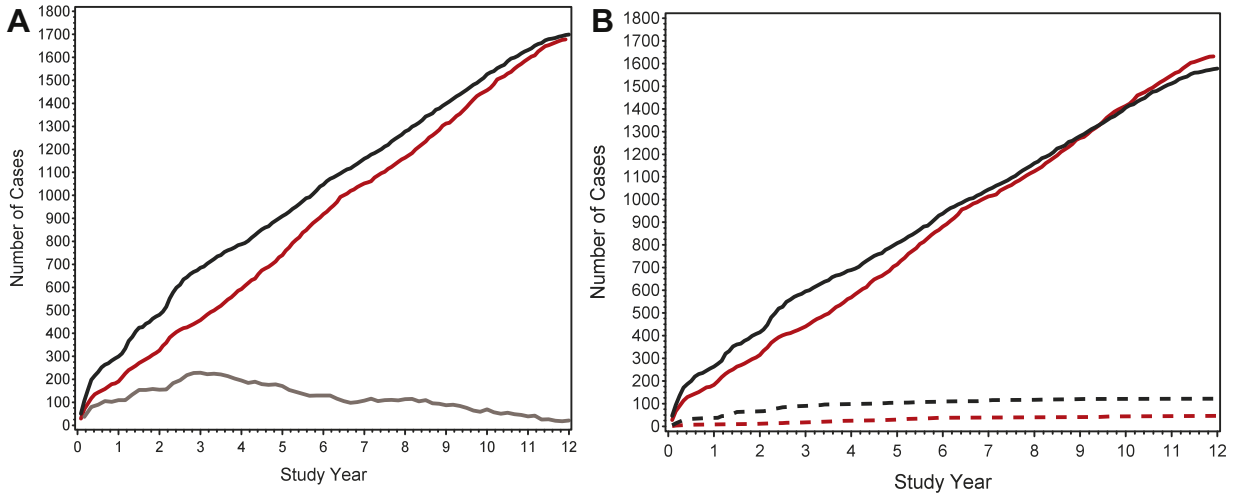

Figure 1. Cumulative lung cancer cases by arm. (*A*) All lung cancers. Black represents the low-dose computed tomography (LDCT) arm, red represents the chest radiography (CXR) arm. Gray line represents excess of cases in the LDCT arm over the CXR arm. (*B*) All cases excluding bronchioloalveolar carcinoma are indicated by solid lines, bronchioloalveolar carcinoma cases are indicated by dotted lines. Black represents the LDCT arm, red represents the CXR arm.

Table 2. Histologic Type and Stage of Lung Cancers by Arm

	LDCT Arm	CXR Arm	
Variable	n (%)	n (%)	*p* Value[a]
All	1701	1681	
Histologic type			
All NSCLC	1397 (82.1)	1343 (79.9)	0.28
BAC	121 (7.1)	46 (2.7)	<0.0001
Adenocarcinoma	608 (35.7)	598 (35.6)	0.76
Squamous	416 (24.5)	395 (23.5)	0.45
Large cell	56 (3.3)	53 (3.2)	0.77
Other NSCLC	196 (11.5)	251 (14.9)	0.009
SCLC	245 (14.4)	291 (17.3)	0.05
Carcinoid	12 (0.7)	7 (0.4)	
Unknown	47 (2.8)	40 (2.4)	
Stage[b]			
I[c]	673 (39.6)	462 (27.5)	<0.0001
IA	523	326	
1B	148	134	
II[c]	145 (8.5)	153 (9.1)	0.65
IIA	91	80	
IIB	43	66	
III[c]	298 (17.5)	321 (19.1)	0.36
IIIA	204	216	
IIIB	84	94	
IV	468 (27.5)	597 (35.5)	<0.0001
Occult	5	4	
Unknown	112 (6.6)	143 (8.5)	

Note: International Classification of Diseases for Oncology, Third Edition, code 8000 is considered unknown and code 8010 is considered other NSCLC.
[a]For the difference in proportion of cases.
[b]Based on the American Joint Committee on Cancer sixth edition for cases through 2009 and (primarily) the seventh edition for cases from 2010 onward.
[c]Includes some cases without distinction between Stage IA and 1B, Stage IIA and IIB or Stage IIIA and IIIB.
LDCT, low-dose computed tomography; CXR, chest radiography; BAC, bronchioloalveolar carcinoma.

Figure 2*A* and *B* show lung cancer deaths over time for the overall and dilution-adjusted analyses, as well as for stage IV cases over time.

Table 4 shows lung cancer mortality RRs for comparable time periods for the originally reported and extended follow-up data, based on calendar time and study time cutoffs, respectively. For similar median follow-up time, the RRs were similar. For example, at a median of 5.5 years' follow-up for both the calendar and study time cutoffs, the RRs were 0.80 and 0.81, respectively (see Table 4). Going from 6 to 7 study years, however, the RR increased substantially, from 0.81 to 0.86, a result of the greater number of lung cancer deaths in the LDCT than in the CXR arm in study year 7 (see Fig. 2*A*).

Overall mortality results by arm are shown in Table 3. The overall mortality RR was 0.97 (95% CI: 0.94–1.01), with a difference across arms in the number dying (per 1000) of 4.2 (95% CI: −2.6 to 10.9, $p = 0.18$). The distribution of causes of death was similar across arms (Supplementary Table 3).

Discussion

In this extended follow-up analysis of the NLST, the difference in the proportion dying of lung cancer across arms (CXR minus LDCT) was 3.3 per 1000, which translates into an NNS to prevent one lung cancer death of 303. This difference of 3.3 per 1000 was similar to that observed in prior analyses of the original trial data, based either on the January 15, 2009, cutoff (3.2 per 1000) or the December 31, 2009, cutoff (3.1 per 1000), and the NNS of 303 was similar to the earlier reported NNS values of around 320.[5,17] The stability of this difference over time

Table 3. Lung Cancer Mortality, Stage IV Incidence and Overall Mortality by Arm

Outcome	LDCT, n (per 1000 subjects)	CXR, n (per 1000 subjects)	Difference across Arms per 1000 Subjects (95% CI) (CXR Minus LDCT)	RR (95% CI)	*p* Value Interaction[a]
All lung cancer deaths					
All subjects	1147 (42.9)	1236 (46.2)	3.3 (-0.2 to 6.8)	0.92 (0.85-1.00)	
Men	733 (46.5)	755 (47.9)	1.4 (-3.3 to 6.1)	0.97 (0.87-1.07)	0.17
Women	414 (37.8)	481 (43.9)	6.1 (0.8-11.3)	0.86 (0.75-0.98)	
Current smoker	724 (56.3)	818 (63.4)	7.1 (1.3-12.9)	0.88 (0.80-0.97)	0.12
Former smoker	423 (30.5)	418 (30.2)	-0.3 (-4.3 to 3.8)	1.01 (0.88-1.15)	
Age 55-64 y at randomization	641 (32.7)	739 (37.7)	5.0 (1.3-8.6)	0.86 (0.78-0.96)	0.051
Age 65-74 y at randomization	506 (71.2)	497 (69.9)	-1.3 (-9.7 to 7.2)	1.01 (0.90-1.15)	
Lung cancer deaths: dilution-adjusted analysis[b]					
All subjects	578 (21.6)	646 (24.2)	2.5 (0.001-5.1)	0.89 (0.80-0.997)	
Men	373 (23.7)	390 (24.7)	1.1 (-2.3 to 4.5)	0.95 (0.83-1.10)	0.14
Women	205 (18.7)	256 (23.3)	4.6 (0.8-8.4)	0.80 (0.66-0.96)	
Current smoker	356 (27.7)	423 (32.8)	5.1 (0.9-9.3)	0.84 (0.73-0.97)	0.16
Former smoker	222 (16.0)	223 (16.1)	0.1 (-2.9 to 3.1)	0.99 (0.82-1.19)	
Age 55-64 y at randomization	310 (15.8)	362 (18.4)	2.6 (0.1-5.2)	0.85 (0.73-0.99)	0.39
Age 65-74 y at randomization	268 (37.7)	284 (40.0)	2.3 (-4.1 to 8.6)	0.94 (0.80-1.11)	
Stage IV cases					
All subjects	468 (17.5)	597 (22.3)	4.8 (2.5-7.2)	0.79 (0.70-0.89)	
Men	303 (19.2)	365 (23.2)	3.9 (0.8-7.1)	0.83 (0.71-0.97)	0.24
Women	165 (15.1)	232 (21.2)	6.1 (2.6-9.6)	0.71 (0.58-0.87)	
Current smoker	297 (23.1)	386 (29.9)	6.8 (2.9-10.7)	0.77 (0.66-0.90)	0.69
Former smoker	171 (12.3)	211 (15.3)	2.9 (0.2-5.7)	0.81 (0.66-0.99)	
Age 55-64 y at randomization	278 (14.2)	367 (18.7)	4.5 (2.0-7.0)	0.76 (0.65-0.89)	0.48
Age 65-74 y at randomization	190 (26.7)	230 (32.4)	5.6 (0.1-11.2)	0.83 (0.69-1.01)	
Stage IV cases through year 6					
All Subjects	245 (9.2)	344 (12.9)	3.7 (1.9-5.5)	0.71 (0.60-0.84)	
Men	165 (10.5)	214 (13.6)	3.1 (0.7-5.5)	0.77 (0.63-0.95)	0.21
Women	80 (7.3)	130 (11.9)	4.5 (2.0-7.1)	0.62 (0.47-0.82)	
Current smoker	153 (11.9)	221 (17.1)	5.2 (2.3-8.2)	0.70 (0.57-0.86)	0.66
Former smoker	92 (6.6)	123 (8.9)	2.3 (0.2-4.3)	0.75 (0.57-0.98)	
Age 55-64 y at randomization	140 (7.1)	207 (10.5)	3.4 (1.6-5.3)	0.68 (0.55-0.84)	0.46
Age 65-74 y at randomization	105 (14.8)	137 (19.3)	4.5 (0.3-8.8)	0.77 (0.60-0.99)	
Overall mortality (all subjects)	5253 (196.6)	5366 (200.7)	4.2 (-2.6 to 10.9)	0.97 (0.94-1.01)	
Overall mortality excluding lung cancer deaths (all subjects)	4106 (153.7)	4130 (154.5)	0.9 (-5.3 to 7.0)	0.99 (0.95-1.03)	

[a]All deaths with a corresponding lung cancer diagnosis within 6 years of randomization were included.
[b]*p* Value for interaction of trial arm by age, sex, or smoking status for the RR.
LDCT, low-dose computed tomography; CXR, chest radiography; CI, confidence interval; RR, rate ratio.

indicates that LDCT screening did not just delay lung cancer death by a few years but also prevented it, or at least delayed it for more than a decade.

In contrast to the stability over time of the difference in lung cancer deaths across arms, and by extension the NNS, the RR for lung cancer mortality changed

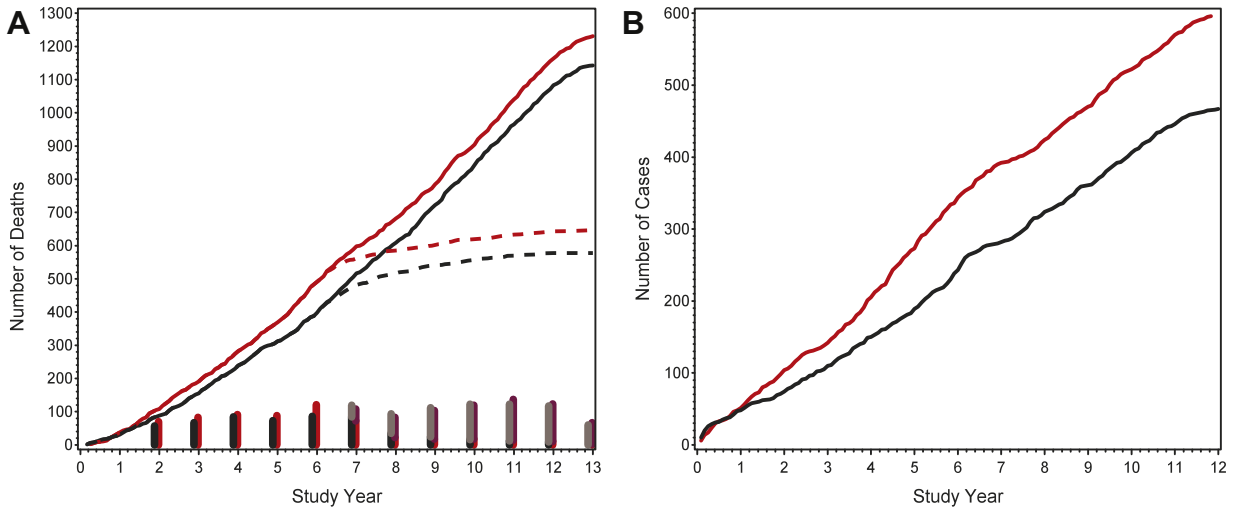

Figure 2. (*A*) Lung cancer deaths by trial arm. Black represents the low-dose computed tomography arm, red represents the chest radiography arm; solid lines are all deaths and dotted lines show deaths for the dilution-adjusted analysis. Vertical bars show the number of deaths for each study year. Black and gray bars and red and purple bars represent the chest radiography and low-dose computed tomography arms, respectively. The total height of each bar represents all deaths, the height of the black or red segment represents the number of deaths for the dilution-adjusted analysis. (*B*) Stage IV lung cancers cases by trial arm. Black represents the LDCT arm, red represents the CXR arm.

substantially over time. The RR derived from the original data increased from 0.80 to 0.84 on the basis of a relatively small (about 1-year) difference in the calendar time cutoff. With extended follow-up well beyond the end of protocol screening, the RR would be expected to move toward the null due because of dilution of the screening effect, and this was in fact observed, with an RR of 0.92. However, for the dilution-adjusted analysis, the RR was 0.89, showing a smaller reduction in mortality than in earlier analyses. Although this RR was adjusted for dilution, dilution may still have affected the estimate, as the 4-year postscreening window for diagnosis likely included some cancers with short lead times whose outcome could not have been affected by screening. Although mortality RR estimates from trials are an important public health tool for assessing screening benefits, they are problematic because a standard screening trial, with several rounds of screening and some additional years of follow-up, does not match up exactly with screening as performed in the

Table 4. Lung Cancer Mortality Results for Comparable Median Follow-up Periods Based on Calendar Time versus Study Time Cutoffs

Time Period for Inclusion of Lung Cancer Deaths	Median (25th Percentile/ 75th Percentile) Years of Follow-up for Mortality	LDCT	CXR	RR (95% CI)	Difference per 1000 subjects
Study time cutoff		Lung Cancer Deaths, n	Lung Cancer Deaths, n		
Through study year					
5.0	5.0 (5.0/5.0)	312	370	0.84 (0.72-0.98)	2.2
5.5	5.5 (5.5/5.5)	347	427	0.81 (0.71-0.93)	3.0
6.0	6.0 (6.0/6.0)	398	491	0.81 (0.71- 0.92)	3.5
6.5	6.5 (6.5/6.5)	457	550	0.83 (0.73-0.94)	3.5
7.0	7.0 (7.0/7.0)	517	600	0.86 (0.76-0.97)	3.1
Calendar time cutoff					
Through January 15, 2009	5.5 (5.2/5.9)	356	443	0.80 (0.73-0.93)	3.3
Through December 31, 2009	6.5 (6.1/6.9)	469	552	0.84 (0.75-0.96)	3.1

Note: Includes only those lung cancer deaths occurring in the given time periods. This is in contrast to the dilution analyses, in which deaths can occur any time but diagnoses have to occur during certain time periods.
LDCT, low-dose computed tomography; CXR, chest radiography; RR, rate ratio; CI, confidence interval.

population setting. As seen here, small changes in follow-up time can lead to nontrivial changes in RR. With the original NLST findings, modeling efforts attempted to extrapolate trial results to the population screening setting.[18,19] Additional modeling efforts incorporating these extended follow-up data, and the results of the NELSON trial, may prove useful for informing both the population and individual perspectives, the latter of which is most appropriate for shared decision making.

The *p* values for the lung cancer mortality RR and difference in proportions hovered around the 0.05 level. *p* Values were not emphasized because the null hypothesis of no lung cancer mortality difference across arms had already been rejected by the original analysis. The dilution effect of adding (roughly) equal numbers of events in each arm, in addition to moving the RR but not the difference in proportions toward the null, also increases the SD of both the RR and the difference in proportions and thus tends to increase the associated *p* value. For example, counting only lung cancer deaths occurring within 6.5 years of randomization, there were 457 and 550 in the LDCT and CXR arms, respectively, giving a difference in proportions of 3.5 (per 1000) and a corresponding SD of 1.1. For total follow-up, approximately equal numbers of deaths were added (690 and 686 in the LDCT and CXR arms, respectively), giving a similar difference in proportions (3.3) but a substantially inflated SD of 1.8, which caused the *p* value to increase from 0.003 to 0.06. The additional nearly 700 deaths in each arm also caused the RR to increase toward the null, from 0.83 to 0.92, and the *p* value to increase from 0.003 to 0.05.

The reduction in stage IV disease across arms was greater than the reduction in lung cancer deaths. Because stage IV cases have a high case fatality rate, the difference in deaths from stage IV cancers across arms ($n = 140$) was similar to the difference in overall stage IV cases across arms ($n = 129$). However, the difference in lung cancer deaths across arms was only 89, because the 140 fewer deaths from stage IV cancers in the LDCT arm were partially offset by an excess in the LDCT arm of 31 deaths from stage I to III and 20 deaths from unknown stage cancers. Therefore, some of the difference across arms in stage IV cases may have been the result of earlier (at a time when metastases were not clinically apparent) diagnosis in the LDCT arm of tumors that eventually progressed to metastatic disease in spite of early diagnosis and treatment.

In contrast to what was observed with the original follow-up, in this extended follow-up analysis there was no statistically significant reduction in all-cause mortality in the LDCT arm versus in the CXR arm. However, as already described, for the same difference in proportions, the *p* value is substantially higher in the extended follow-up than in original analysis because of the extra noise associated with the dilution effect. For all-cause mortality, the differences across arms in the proportion dying were 4.6 per 1000 in the original analysis and a similar 4.2 per 1000 here, indicating that the difference in all-cause mortality was essentially sustained. The *p* value, though, increased from 0.02 to a nonsignificant 0.18 in the extended follow-up. Therefore, the current lack of a statistically significant effect for all-cause mortality should not be taken to negate the original significant finding; it is more likely related to use of the "incorrect window" for follow-up (i.e., too long a period after screening).[16] In addition, with respect to non–lung cancer mortality, the original RR was 0.96, with a corresponding nonsignificant *p* value of 0.29; therefore, this is not inconsistent with the currently observed nonsignificant RR of 0.99 for non–lung cancer mortality.

As with the originally reported results, in the updated analysis there was an observed lower RR for lung cancer mortality (i.e., greater percentage reduction in mortality with LDCT) in women than in men, although the interaction of sex and trial arm was not statistically significant. Preliminary results from the NELSON trial also show a greater observed percentage mortality reduction in women, although whether this represents a statistically significant difference is not clear.[11] There were also some observed differences here in lung cancer mortality RRs by age and smoking status, but given the nonsignificant interaction *p* values, it is not clear whether these are real. Note that for age, the *p* value was borderline significant (0.05) for the overall analysis but not close to significant ($p = 0.39$) for the dilution-adjusted analysis, and note also that the analysis of interactions involved multiple comparisons. For sex as well as for age and smoking status, meta-analyses of all LDCT trials may shed some light on whether the effect of LDCT screening is truly differential by these factors. From a public health standpoint, even if the RRs were the same according to, say, smoking status, the higher background lung cancer rate for current versus former smokers indicates that the risk difference (difference in proportions dying of lung cancer across arms) would be greater, and correspondingly, that the NNS would be lower in current versus former smokers.

Additionally, the NLST was not powered for interactions, so modest but potentially clinically significant interactions of the RR with the factors of age, sex, or smoking status could have failed to reach statistical significance. However, within this trial population, all of which was at high risk because of smoking history but was also generally healthy, it is unclear what the biological rationale would be for an interaction with these factors; thus, any true interactions would likely be of small magnitude. With more varied populations potentially undergoing LDCT screening, this might not be the

case, as factors related to ability to undergo curative treatment or to differential lung cancer histologic type might alter the effectiveness of LDCT.

After a median of 11.3 years' follow-up for incidence, or 9.3 years after the last scheduled screen, lung cancer incidence was similar across arms, with an RR of 1.01 (95% CI: 0.95–1.09) for the LDCT arm versus the CXR arm. In contrast, in the original trial period of a median of 6.5 years of follow-up, there was a significantly elevated RR of 1.13.[5] This indicates that so-called "catch-up" likely occurred in the CXR arm, in which the counterparts of those cancers diagnosed early in the LDCT arm were eventually diagnosed in the CXR arm. A mathematical model of lung cancer natural history fit to the original NLST data predicted that 94% of cases, excluding BAC, would become clinically apparent within 10 years of LDCT screen diagnosis.[12] Because the average follow-up of LDCT screen–diagnosed cases is now about 10 years, the estimate of 94% is generally consistent with the current observation of no increase in (non-BAC) lung cancer in the LDCT arm. In contrast, there continued to be a large excess of BAC cases in the LDCT arm (n = 121) versus in the CXR arm (n = 46), with few additional cases identified after the original follow-up period. This is also consistent with the aforementioned model's predictions, which estimated that only around 25% of screen-detected BAC would become clinically apparent within 10 years. Some BAC cases could eventually present clinically after more than 10 years, so the overdiagnosis estimate of 79% for BAC could be an overestimate.

In 2011, a multisociety committee recommended changes to the classification of lung adenocarcinoma, reclassifying BAC into new categories of adenocarcinoma in situ, minimally invasive adenocarcinoma (MIA), invasive lepidic adenocarcinoma, and invasive mucinous adenocarcinoma, and discontinuing use of the term *BAC*.[20,21]. The new categories involve the same *International Classification of Diseases for Oncology* morphology codes as previously used for BAC, with the exception of new codes for MIA. These same codes were used through the entire NLST follow-up period to define BAC (or what was formerly known as BAC); thus, the reclassification should not have affected the overdiagnosis estimate for BAC. Note that MIA tumors were not ascertained in the NLST.

The magnitude of overdiagnosis as estimated from LDCT screening trials depends critically on the length of follow-up after the final screen. For the NLST, the overdiagnosis rate decreased from 18% in the original analysis (a median of 4.5 years' follow-up after the last screen) to 3% with extended follow-up. However, even controlling for follow-up time, there is great variability in overdiagnosis rates across trials. In the Danish trial, after a median 5 years of follow-up after the final screen, the overdiagnosis rate was 67%, whereas in the ITALUNG trial, with a median of 4.5 years of such follow-up, the overdiagnosis rate was zero.[6,22] More research is needed concerning the factors that influence overdiagnosis in LDCT screening.

A limitation of the analysis was that use of LDCT screening after the original trial period was not ascertained. NLST participants were sent a letter in 2010 summarizing trial results, with subjects in the CXR arm told that they might want to discuss LDCT screening with their health care provider and subjects in the LDCT arm told that they might want to discuss continuing screening. However, LDCT screening was not generally covered by private insurance or Medicare until 2015, and survey evidence suggests that use was low in the United States through 2015.[23,24] However, as trial volunteer participants, the NLST subjects may have been more motivated to receive screening than eligible individuals in the general population. In addition, indirect evidence suggests there was little LDCT screening among NLST participants after the screening phase of the trial. As already described, there were few cases of BAC (or in the new terminology but with the same morphology codes, invasive lepidic or mucinous adenocarcinoma), after the screening phase of the trial. After 90 cases of BAC in the LDCT arm during the 3 screening phase years of the trial (T0–T2), including 24 cases after the second incidence (T2) screen, there were an average of only four cases per year (31 in total) for the next 8 years in the LDCT arm and a similar number during that period in the CXR arm. Because BAC cases are generally found only with LDCT screening, such screening was likely low, and similar across trial arms, in the postscreening phase of the trial. Another limitation was that death review was not performed for deaths after the original analysis period. However, an analysis of the agreement between death certificates and death review for the original period showed high levels of agreement and minimal effect on the lung cancer mortality RR.[25]

Conclusion

With further follow-up of NLST subjects, the originally reported reduction in lung cancer deaths in the LDCT arm versus in the CXR arm was sustained; in contrast, the originally reported increase in lung cancer incidence was no longer observed.

Acknowledgments

Cancer incidence data have been provided by the following state cancer registries: Alabama, Arizona, California, Colorado, District of Columbia, Georgia, Hawaii,

Idaho, Indiana, Iowa, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nevada, North Carolina, Ohio, Pennsylvania, Rhode Island, Texas, Utah, Virginia, and Wisconsin. All are supported in part by funds from the Centers for Disease Control and Prevention; National Program for Central Registries; local states; or National Cancer Institute Surveillance, Epidemiology, and End Results Program. The results reported here and the conclusions derived are the sole responsibility of the authors.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at and at https://doi.org/10.1016/j.jtho.2019.05.044.

Appendix

Denise R. Aberle, MD, Department of Radiological Sciences, University of California Los Angeles, Los Angeles, California

William C. Black, MD, Dartmouth School of Medicine, Hanover, New Hampshire

Caroline Chiles, MD, Wake Forest University, Winston-Salem, North Carolina

Timothy R. Church, PhD, University of Minnesota, Minneapolis, Minnesota

Ilana F. Gareen, PhD, Brown University, Providence, Rhode Island

David S. Gierada, MD, Washington University School of Medicine, St. Louis, Missouri

Irene Mahon, MPH, Eastern Cooperative Oncology Group–American College of Radiology Imaging Network, Philadelphia, Pennsylvania

Eric A. Miller, PhD, National Cancer Institute, Bethesda, Maryland

Paul F. Pinsky, PhD, National Cancer Institute, Bethesda, Maryland

JoRean D. Sicks, MS, Brown University, Providence, Rhode Island

References

1. World Health Organization. GLOBOCAN 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. Accessed October 10, 2018.
2. Naidich DP, Marshall CH, Gribbin C, Arams RS, McCauley DI. Low-dose CT of the lungs: preliminary observations. *Radiology*. 1990;175:729–731.
3. Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology*. 1996;201:798–802.
4. Aberle DR, Berg CD, Black WC, et al. The National Lung Screening Trial: overview and study design. *Radiology*. 2011;258:243–253.
5. The National Lung Screening Trial Research Team, Aberle DR, Adams EM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
6. Paci E, Puliti D, Pegna AL, et al. Mortality, survival and incidence rates in the ITALUNG randomized lung cancer screening trial. *Thorax*. 2017;72:825–831.
7. Infante M, Cavuto S, Lutman FR, et al. Long-term follow-up results of the DANTE Trial, a randomized study of lung cancer screening with spiral computed tomography. *Am J Resp Crit Care Med*. 2015;191:1166–1175.
8. Wille MW, Dirksen A, Ashraf H, et al. Results of the randomized Danish Lung Cancer Screening Trial with focus on high-risk profiling. *Am J Resp Crit Care Med*. 2016;193:54–551.
9. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev*. 2012;21:308–315.
10. Doroudi M, Pinsky PF, Marcus PM. Lung cancer mortality in the Lung Screening Study (LSS) feasibility trial. *JNCI Cancer Spectr*. 2018;2:pky042.
11. De Koning H. Mortality results of the NELSON randomized controlled population-based screening trial. Paper presented at: IASLC 19th World Conference on Lung Cancer. September 23–26, 2018; Toronto, Canada.
12. Patz E, Pinsky PF, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med*. 2014;174:269–274.
13. Etzioni R, Self SG. On the catch-up time method for analyzing cancer screening trials. *Biometrics*. 1995;51:31–43.
14. Prorok PC, Aron JL. An analysis of the mortality effect in a breast cancer screening study. *Int J Epidemiol*. 1986;15:36–43.
15. Nelson HD, Fu R, Cantor A, et al. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. Preventive Services Task Force recommendation. *Ann Intern Med*. 2016;164:244–255.
16. Hanley J. Analysis of mortality data from cancer screening studies: looking in the right window. *Epidemiology*. 2005;16:786–790.
17. Pinsky PF, Church TR, Izmirlian G, Kramer BS. The National Lung Screening Trial: results stratified by demographics, smoking history and lung cancer histology. *Cancer*. 2013;119:3976–3983.
18. De Koning H, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;160:311–320.
19. McMahon PM, Meza R, Plevritis SK, et al. Comparing benefits from many possible computed tomography lung cancer screening programs: extrapolating from the National Lung Screening Trial using comparative modeling. *PLoS One*. 2014;9:e99978.

20. Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*. 2011;6:24–248.
21. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors. *J Thorac Oncol*. 2015;10:1243–1260.
22. Heleno BN, Siersma V, Brodersen J. Estimation of overdiagnosis of lung cancer in low-dose computed tomography screening: a secondary analysis of the Danish Lung Cancer Screening Trial. *JAMA Intern Med*. 2018;178:1420–1422.
23. Jemal A, Fedewa SA. Lung cancer screening with low-dose computed tomography in the United States-2010 to 2015. *JAMA Oncol*. 2017;3:1278–1281.
24. Huo J, Shen C, Volk RJ, Shih Y. Use of CT and chest radiography for lung cancer screening before and after publication of screening guidelines: intended and unintended uptake. *JAMA Intern Med*. 2017;177:439–441.
25. Marcus PM, Doria-Rose VP, Gareen IF, et al. Did death certificates and a death review process agree on lung cancer cause of death in The National Lung Screening Trial? *Clin Trials*. 2016;13:434–438.