

# Estimating Long-term Effectiveness of Lung Cancer Screening in the Mayo CT Screening Study<sup>1</sup>

Pamela M. McMahon, PhD  
Chung Yin Kong, PhD  
Bruce E. Johnson, MD  
Milton C. Weinstein, PhD  
Jane C. Weeks, MD, MS  
Karen M. Kuntz, ScD  
Jo-Anne O. Shepard, MD  
Stephen J. Swensen, MD  
G. Scott Gazelle, MD, MPH, PhD

## Purpose:

To use individual-level data provided from the single-arm study of helical computed tomographic (CT) screening at the Mayo Clinic (Rochester, Minn) to estimate the long-term effectiveness of screening in Mayo study participants and to compare estimates from an existing lung cancer simulation model with estimates from a different modeling approach that used the same data.

## Materials and Methods:

The study was approved by institutional review boards and was HIPAA compliant. Deidentified individual-level data from participants (1520 current or former smokers aged 50–85 years) in the Mayo Clinic helical CT screening study were used to populate the Lung Cancer Policy Model, a comprehensive microsimulation model of lung cancer development, screening findings, treatment results, and long-term outcomes. The model predicted diagnosed cases of lung cancer and deaths per simulated study arm (five annual screening examinations vs no screening). Main outcome measures were predicted changes in lung cancer-specific and all-cause mortality as functions of follow-up time after simulated enrollment and randomization.

## Results:

At 6-year follow-up, the screening arm had an estimated 37% relative increase in lung cancer detection, compared with the control arm. At 15-year follow-up, five annual screening examinations yielded a 9% relative increase in lung cancer detection. The relative reduction in cumulative lung cancer-specific mortality from five annual screening examinations was 28% at 6-year follow-up (15% at 15 years). The relative reduction in cumulative all-cause mortality from five annual screening examinations was 4% at 6-year follow-up (2% at 15 years).

## Conclusion:

Screening may reduce lung cancer-specific mortality but may offer a smaller reduction in overall mortality because of increased competing mortality risks associated with smoking.

© RSNA, 2008

<sup>1</sup> From the Institute for Technology Assessment, Massachusetts General Hospital, 101 Merrimac St, 10th Floor, Boston, MA 02114 (P.M.M., C.Y.K., G.S.G.); Departments of Radiology (P.M.M., C.Y.K., J.O.S., G.S.G.) and Medicine (B.E.J., M.C.W., J.C.W.), Harvard Medical School, Boston, Mass; Lowe Center for Thoracic Oncology (B.E.J.) and Department of Medical Oncology/Population Sciences (J.C.W.), Dana-Farber Cancer Institute, Boston, Mass; Department of Health Policy and Management, Harvard School of Public Health, Boston, Mass (M.C.W., K.M.K., G.S.G.); Department of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, Minn (K.M.K.); and Department of Radiology, Mayo Clinic, Rochester, Minn (S.J.S.). Received August 13, 2007; revision requested October 10; revision received November 9; accepted January 2, 2008; final version accepted January 30. Supported in part by the National Cancer Institute (R01 CA97337, G.S.G.; K99 CA126147, P.M.M.). Address correspondence to P.M.M. (e-mail: [pamela@mgh-ita.org](mailto:pamela@mgh-ita.org)).

Lung cancer is the leading cause of cancer death in the United States and is responsible for more deaths annually than breast, prostate, and colon cancers combined (1). A readily identifiable population at increased risk is that of current and former cigarette smokers, as cigarette smoking causes more than 80% of lung cancers (2–4). Imaging with chest radiography or computed tomography (CT) helps detect asymptomatic lung cancers, many of which are amenable to effective treatment; surgical resection of stage I non-small cell lung cancer (NSCLC) offers 5-year survival rates of more than 50% (5–8). Researchers in a recent study reported that patients whose lung cancers were detected by using CT screening and whose cancers were surgically resected have a 10-year survival in excess of 80% (9). In the absence of screening, however, most lung cancers are diagnosed when distant spread has already occurred; only 16% of patients with newly diagnosed lung cancer survive 5 years (1).

Together, all of these factors suggest that lung cancer may be a good candidate for mass screening. Large randomized trials of screening with chest radiography, however, did not demonstrate a reduction in lung cancer mortality (10–13). Recent single-arm studies of lung cancer screening with CT have shown that screening helps detect greater than twice as many early-stage lung cancers than would be expected to be detected without screening (9,13–17). Single-arm study designs that compare a screening population versus an external nonscreening population, however, cannot definitively demonstrate reductions in either lung cancer–specific or all-cause mortality

rates. The possibility of observing higher interval (eg, 5-year) survival after diagnosis in the absence of a mortality reduction is explained by several well-known biases present in screening trial data: lead-time, length-time, and overdiagnosis biases (18–21). Because all three biases can contribute to longer survival of patients with screening-detected cancers, a control arm is critical for parsing out any true effect of screening on mortality.

In conjunction with trials, simulation modeling may be used to integrate available data (22), to evaluate screening programs, and to inform those making screening decisions in the years prior to publication of long-term results from randomized trials. Investigators in two large ongoing randomized trials (23,24) will report short-term results on the effectiveness of lung cancer screening during the next few years, but publication of trial results does not always eliminate uncertainty about the effectiveness of cancer screening (13,25,26). By varying screening protocols, patient populations, or adherence rates, models can be used to interpret and reconcile apparently inconsistent trial results. Modeling also may be used to estimate sample sizes.

The objective of our evaluation was to use individual-level data provided from the single-arm study of helical CT screening at the Mayo Clinic (Rochester, Minn) (16,27) to estimate the long-term effectiveness of screening in Mayo study participants and to compare estimates from an existing lung cancer simulation model with estimates from a different modeling approach that used the same data (28). Individual-level data on smoking histories are essential for modeling the dynamic relationship between smoking behavior and lung cancer risk because summary tables of smoking history and the common pack-year metric collapse critical information about timing and dose.

#### Advances in Knowledge

- The results of our analysis indicate that helical CT screening for lung cancer may offer a moderate reduction in lung cancer mortality in a population of heavy smokers.
- An important caveat is that competing mortality risks from smoking may limit the overall effectiveness of screening.

#### Implication for Patient Care

- On the basis of the results from this simulation model, helical CT screening could play a role in the routine health care of heavy smokers.

## Materials and Methods

### Protected Health Information

The protocol for use of human subject data for this simulation study underwent expedited review (secondary use of medical records) and was approved by the human subjects' institutional review board as compliant with Health Insurance Portability and Accountability Act guidelines. Deidentified records from 1520 current and former smokers aged 50–85 years participating in a Mayo Clinic study of annual CT screening for early detection of lung cancer (16) were provided to Massachusetts General Hospital (Boston, Mass). Participants signed informed consent waivers approved by the institutional review board of the Mayo Clinic before enrollment in the screening study. The transfer of the deidentified data (consisting of demographics, smoking history, and screening results) and the use of the data in this study were approved by the human subjects' review boards of Massachusetts General Hospital and the Mayo Clinic (Rochester, Minn) and were exempt from further informed consent requirements.

### Lung Cancer Policy Model

The Lung Cancer Policy Model (LCPM) is a comprehensive microsimulation model

Published online before print  
10.1148/radiol.2481071446

Radiology 2008; 248:278–287

#### Abbreviations:

CI = confidence interval  
LCPM = Lung Cancer Policy Model  
NLST = National Lung Screening Trial  
NSCLC = non-SCLC  
SCLC = small cell lung cancer  
SEER = Surveillance, Epidemiology, and End Results

#### Author contributions:

Guarantors of integrity of entire study, P.M.M., C.Y.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, P.M.M.; clinical studies, S.J.S.; statistical analysis, P.M.M., C.Y.K., M.C.W., K.M.K.; and manuscript editing, all authors

Authors stated no financial relationship to disclose.

of lung cancer development, disease progression, lung cancer detection, treatment results, and survival (29). The LCPM addresses three important limitations of published models used to evaluate lung cancer screening (30–36). First, the LCPM simulates survival after screening detection as a function of “true” (known in the model) disease characteristics (eg, stage and growth rates). The explicit modeling of survival avoids the problematic assumption—inherent in stage-shift models—that screening-detected cancers behave like non-screening-detected cancers; the screening biases listed previously undermine this assumption (18–21,37,38). Second, the LCPM explicitly models benign nodules. Compared with a range of lung cancer prevalence of 1%–2.7%, 23%–51% of screened smokers have detectable nodules on baseline CT screening scans (14,16,39). High false-positive rates are potentially worrisome because of the burdens that subsequent evaluations place on patients and the health care system and because people with false-positive screening scans may be less likely to participate in subsequent screening examinations (40). Third, the LCPM incorporates the high competing mortality risks faced by cigarette smokers (41); failure to incorporate those risks would yield a biased (inflated) estimate of the effect of screening on mortality.

The LCPM is a state-transition model analyzed as a patient-level Monte Carlo simulation to allow for individual heterogeneity in risk factors (eg, smoking history) and event rates. A 1-month cycle length captures the short survival times of late-stage lung cancer and allows for a variety of event frequencies. Outputs include estimates of incident cancers stratified according to age, cell type, and stage, estimates of survival according to stage at detection, and estimates of non-lung cancer deaths.

Natural history parameters were estimated by populating the LCPM with individuals assigned a smoking history that was representative of a specified age-sex-race-calendar year cohort of the U.S. population (42,43) and by calibrating the model to tumor registry data

from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (8). Data from past cohort studies (44) and other literature sources (4,45–47) that described clinical experience were used as secondary calibration targets.

Data from the Mayo CT screening study (16) were used to estimate age-specific probabilities of benign nodules. To estimate the proportion of adenocarcinoma that was bronchioloalveolar carcinoma and the distribution of the growth parameter for adenocarcinoma, we calibrated the model to the binomial 95% confidence interval (CI) around the prevalence (all cell types combined, excluding sputum-detected and interval cancers) reported from the Mayo CT screening study (16). Remaining end points reported from the study were used as validation targets. Further model details and parameter values may be found in a model profile under “Lung-Massachusetts General Hospital” at the National Cancer Institute Web site (<http://cisnet.cancer.gov/profiles/>).

#### Mayo CT Screening Study Data

The Mayo CT Screening Study was a single-arm evaluation of helical CT screening for lung cancer in current and former smokers (16,27,48). Participants were enrolled during 12 months (January to December 1999). At enrollment, the study participants ( $n = 1520$ ) had a mean age of 59 years (range, 50–85 years); 52% were men and 48% were women; and 99% were white; 1% were African American, Native American, or Hispanic. At enrollment, most (61%) participants were current smokers, and the participants had a smoking history of a median of 45 pack-years (range, 20–230 pack-years). All participants underwent baseline (prevalence) screening and were assigned to three annual CT examinations. Later, the number was changed to four examinations, for a total of five screening examinations.

#### Simulating Outcomes in the Screening Arm

The LCPM was populated with the Mayo CT study participants by drawing

(with replacement) from the individual records. We simulated the published screening protocol (five annual helical CT screening examinations) and assumed perfect adherence for individuals who did not receive a diagnosis of lung cancer. Participants with a confirmed diagnosis of lung cancer were not eligible for screening but could instead undergo surveillance (National Cancer Institute Web site at <http://cisnet.cancer.gov/profiles/>). We modeled participants with screening-detected nodules as undergoing the suggested follow-up protocol, which varies according to the size of the largest nodule, of the study (16). We assumed that helical CT scans could depict nodules as small as 2 mm in diameter (consistent with the 5-mm section thickness and 3.75-mm reconstruction thickness in screening examinations and 1–3-mm section thickness on follow-up thin-section CT scans in the study) and derived estimates of test sensitivity according to size and location from the study (16). We defined an imaging examination with false-positive findings as one in which an ultimately benign nodule of any size was detected. Note that this definition allows us to disregard those few screening examinations for which findings were positive but no nodule existed (per-person specificity of helical CT was assumed to be 98%); follow-up thin-section CT was assumed to be capable of resolving all such screening CT findings (ie, perfect specificity for absence of any pulmonary nodule).

#### Simulating Outcomes in the Control Arm

To simulate a hypothetical control arm for the Mayo CT screening study, we populated the LCPM with the same (de-identified) individual smoking histories from the Mayo CT study and disabled the screening component, although asymptomatic lung cancers could be detected incidentally (National Cancer Institute Web site at <http://cisnet.cancer.gov/profiles/>).

#### Base-Case and Sensitivity Analyses

Base-case outputs from the screening arm (simulated as described previously) were used to validate data against pub-

lished study end points (16,27). Outputs from both simulated study arms were used to estimate the stage shift caused by screening; detection, mortality, and survival rates were stratified according to study arm. Changes in lung cancer-specific and all-cause mortality were predicted. Counts of cases of lung cancer and of deaths according to cause were normalized to a study size of 1520 per study arm for consistency with the Mayo CT study. With sensitivity analyses, we examined the effects of changes in critical but uncertain model inputs. We modeled scenarios of 10 annual screening examinations; reduced operative mortality rates for lobectomy, mediastinoscopy, and excisional (wedge) biopsies (National Cancer Institute Web site at <http://cisnet.cancer.gov/profiles/>); and implemented a protocol with no follow-up for nodules smaller than 4 mm in diameter. Also modeled was a protocol with more sampling biopsies (fine needle, bronchoscopy, mediastinoscopy) and correspondingly fewer excisional biopsies for indeterminate growing nodules (20% vs 50% base-case rate) and large but stable nodules (2% vs 11% base-case rate). An additional sensitivity analysis simulated a 20% annual smoking cessation rate in the screening arm, beginning at the baseline screening and continuing for the life of the cohort.

### Statistical Analysis

We simulated 500 000 individuals per study arm to generate more precise estimates of effectiveness than would be possible with a smaller sample size; accordingly, *P* values for comparisons between study arms are not informative. By comparing outcomes in the simulated control and screening study arms, we were able to simulate the likely results of a two-arm clinical trial of the same or larger size as the Mayo CT study. Estimated changes in 15-year lung cancer-specific mortality were calculated from 10 replicates of 1520 simulations per study arm to reflect the potential variance in effectiveness across multiple trials in which the same number of participants were enrolled. To investigate the effect of sample size, we

increased the number of simulations per study arm in increments of 1000 until a significant ( $P < .05$ , log-rank test) difference in 15-year lung cancer mortality was evident between the study arms in nine of 10 replicates.

## Results

### Base Case

**Screening arm.**—Comparison of LCPM outputs (when populated with the Mayo CT study participants and screening was simulated) and published findings (16,27) are provided in Table 1. In the study, the 1.38% prevalence of lung cancer (binomial 95% CI: 0.80%, 1.97%) corresponds to the 21 cases detected by using CT alone at the baseline screening and diagnosed within a median 2-year follow-up (range, 18–30 months). The LCPM generated a preva-

lence of 1.16%, which was within the 95% CI, that corresponded to 18 simulated cancers ( $n = 1520$ ) detected at the baseline screening and diagnosed as lung cancer within 24 months. The LCPM predicted that 52.0%, compared with a reported rate of 51%, of participants would have positive findings at baseline screening.

The percentage of prevalent cancers that were adenocarcinoma (observed, 71%; predicted, 71.4%), the percentage of prevalent NSCLCs that were stage I (observed, 68%; predicted, 72.6%), and the median diameters of prevalent cancers (observed, 13 mm; predicted, 10 mm) were similar for the model and the study; standard deviations of both observed (13 mm) and predicted (8.5 mm) diameters were large. Among patients with a diagnosis of prevalent lung cancer, the LCPM predicted smoking histories that were con-

**Table 1**

### Comparison of Model Prediction with Observed Study Results: Screening Results with LCPM Populated with Mayo CT Study Participants

End Point	Study Result*	LCPM Prediction
<b>Participants</b>		
With positive findings at baseline CT screening (%) <sup>†</sup>	51 (780/1520)	52
With lung cancer at baseline CT screening (%)	1.38 (21/1520) <sup>‡</sup>	1.16
<b>Prevalent lung cancers (%)</b>		
Adenocarcinoma <sup>§</sup>	71 (15/21)	71.4
SCLC	10 (2/21)	2.7
<b>Prevalent NSCLCs (%)</b>		
Stage I	68 (13/19)	72.6
Stage II	21 (4/19)	8.8
Stage III	11 (2/19)	14.0
Stage IV	0	4.6
Median diameter of prevalent lung cancers (mm) <sup>  </sup>	13 (17 ± 13)	10 (13 ± 8.5)
<b>Characteristics of patients with prevalent lung cancers</b>		
Mean no. of cigarettes smoked per day	27	29
Mean years of smoking	43	48
Proportion of men	0.4	0.5
Proportion of women	0.6	0.5
Participants with lung cancer detected at second screening (%)	0.2 (3/1520)	0.29
Cohort with >0 nodules after 4 y of screening (%)	74 (1118/1520)	69.5

Note.—LCPM-predicted end points were calculated from 500 000 simulated participants. Numbers in parentheses were used to calculate the percentages except where otherwise specified. SCLC = small cell lung cancer.

\* Data are from Swensen et al (16,27) in 2002 and 2005, respectively. Sputum-detected and interval cancers were excluded from prevalence results.

<sup>†</sup> A positive finding at CT screening was defined as the detection of at least one nonsolidly calcified pulmonary nodule.

<sup>‡</sup> The binomial 95% CI of 0.80% to 1.97% was based on data from Swensen et al (16) in 2002.

<sup>§</sup> Includes bronchioloalveolar carcinoma and mixed adenocarcinoma and bronchioloalveolar carcinoma.

<sup>||</sup> Data in parentheses are the mean ± standard deviation.

sistent with the smoking histories of the participants in the study.

At the second screening in the study (ie, first-incidence screening), three participants with lung cancer were identified (observed rate, 0.2%; predicted rate, 0.29%). After 4 years of screening, at least one nodule was identified in a reported 74% of participants, versus a predicted rate of 69.5%. By 2 years of follow-up, 0.46% ( $n = 7$ ), versus the base-case predicted rate of 2%, of study participants had undergone wedge excision of benign pulmonary nodules.

### Comparison with Control Arm

**Lung cancer cases and deaths.**—Predicted outcomes according to years of follow-up for the hypothetical control arm and the simulated screening arm (500 000 simulations each, normalized to the study size of 1520) are presented in Table 2. By 6 years after enrollment, there were an estimated 14 additional diagnosed cases of lung cancer in the screening arm versus the control arm (52 vs 38 cases). The number of addi-

tional lung cancer cases ( $n = 14$ ) corresponds to an absolute increase in lung cancer detection of 0.9%, calculated as  $[(52 - 38)/1520 \cdot 100]$ , or a 37% relative increase, calculated as  $\{[(52/1520) - (38/1520)]/[38/1520]\}$ . Extrapolated over the lifetime of the trial cohort, the two simulated study arms differed by only eight cases of lung cancer (screening-related increase in lung cancer detection, 0.5% [absolute] and 4% [relative]).

At 6 years after enrollment, screening, compared with the hypothetical scenario of no screening, in this cohort prevented an estimated seven lung cancer deaths (reductions in cumulative lung cancer-specific mortality, 0.5% [absolute] and 28% [relative]). At 15 years after enrollment, screening in this cohort prevented an estimated 11 lung cancer deaths (reductions in cumulative lung cancer-specific mortality, 0.7% [absolute] and 15% [relative]). Extrapolated over the lifetime of the cohort, screening, compared with the hypothetical scenario of no screening, yielded an

8% relative reduction in cumulative lung cancer-specific mortality.

One of the 11 hypothetical individuals who would have died from lung cancer in the absence of screening died of other causes within the same 15 years after enrollment. The reductions in 15-year cumulative all-cause mortality were 0.6% (absolute) and 2% (relative).

Iatrogenic deaths (defined as those that resulted from invasive staging or therapy for benign disease) were rare but 37% more likely in the screening arm (30 deaths per 100 000 population) compared with the control arm (22 per 100 000).

**Detection rates, stage shift, and survival.**—NSCLC detection as a function of follow-up time (ie, years since randomization) was higher in the simulated screening arm than in the hypothetical control arm while screening was in place (years 0 through 4 since randomization), but at the end of routine screening, the detection rate in the screening arm decreased below that of the control arm (Fig 1). SCLC detection was the same in both simulated study arms (Fig 1).

For lung cancers diagnosed by using any modality in the first 5 years of follow-up, a stage shift was predicted in the screening arm versus the control arm (Fig 2). Compared with NSCLCs diagnosed in the control arm, NSCLCs diagnosed in the screening arm were 3.6 times more likely to be stage I (68% vs 19%) and 6.5 times less likely to be stage IV (8% vs 52%). Figure 2 shows that the absolute number of stage IV cancers was lower in the screening arm than in the control arm. With screening, more NSCLCs of all cell types were detected; there were nearly six (5.9) times more stage I NSCLCs in the screening arm versus the control arm. Most stage I NSCLCs in both simulated study arms were histologic types other than bronchioloalveolar carcinoma (73% in the screening arm vs 87% in the control arm).

The predicted shift to early-stage cancers was consistent with higher predicted 5-year lung cancer-specific survival rates for cases in the screening

**Table 2**

#### Base-Case Outcomes Predicted by LCPM per 1520 Participants according to Trial Arm and Follow-up

Model-predicted Outcome according to Follow-up	Control Arm	Screening Arm
<b>Patients diagnosed with lung cancer</b>		
1.5-y follow-up	9.2	22.0*
6-y follow-up	37.9	51.9†
10-y follow-up	64.6	74.1
15-y follow-up	97.0	105.5
Lifetime of cohort	171.4	179.0
<b>Lung cancer deaths</b>		
1.5-y follow-up	4.2	4.0
6-y follow-up	26.5	19.1
10-y follow-up	47.8	36.6
15-y follow-up	73.5	62.3
Lifetime of cohort	131.3	120.8
<b>All deaths</b>		
1.5-y follow-up	34.7	34.9
6-y follow-up	162.8	157.0
10-y follow-up	302.3	293.6
15-y follow-up	510.7	501.0
Iatrogenic deaths, 6-y follow-up	0.3	0.5

Note.—Data were from 500 000 simulated participants per study arm, normalized to the study size of 1520 participants. Data for both arms are the numbers of patients.

\* The number of cancers diagnosed with screening was 17.2.

† The number of cancers diagnosed with screening was 34.7.

arm versus cases in the control arm: 59% versus 17%. When the same cases (ie, those diagnosed in the first 5 years of follow-up) were stratified according to mode of detection rather than study arm, higher 5-year survival rates were predicted for lung cancers diagnosed by using screening CT (67%) or an incidental imaging (72%) examination versus lung cancers diagnosed on the basis of symptoms (12%). The increased survival of hypothetical patients with screening-detected cancers (63%) and incidentally detected cancers (67%) versus symptom-detected cancers (11%) persisted through 15 years of follow-up.

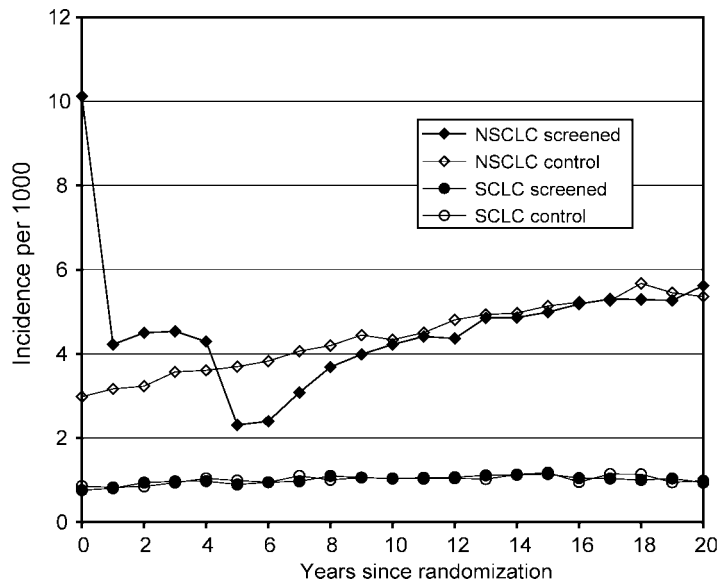
**Evaluating trial size.**—Of 10 replicates of 1520 simulations per trial arm (the size of the study), screening yielded a significant ( $P < .05$ , log-rank test) reduction in 15-year lung cancer-specific mortality in one replicate. A significant ( $P < .05$ ) result was first observed in nine of 10 replicates when the simulated number of patients per trial arm reached 8000.

### Sensitivity Analyses

Extrapolated over the lifetime of the cohort, 10 annual screening examinations, compared with the scenario of no screening, yielded a 14% relative reduction in cumulative lung cancer-specific mortality. The predicted reductions in 15-year cumulative all-cause mortality were 1% (absolute) and 3% (relative).

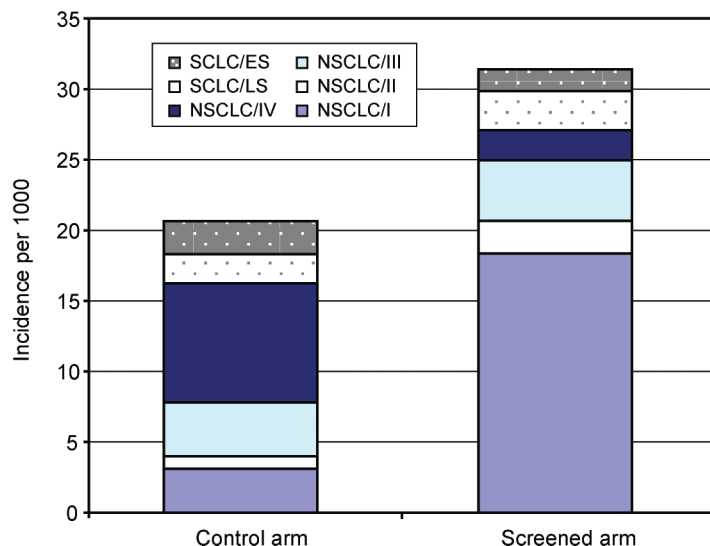
In the analysis for which we assumed that current smokers in the screening arm had a smoking cessation rate of 20% per year (held constant beginning at enrollment), an estimated 75% of current smokers at enrollment had quit smoking by the end of the screening period, with nearly complete cessation by 20 years. The 20% annual cessation rate is similar to the 23% rate reported by researchers in the Early Lung Cancer Action Program (49), versus a lower reported estimate of 14% from the Mayo Clinic (50). Relative to a 3% cessation rate in the simulated control arm, the high cessation rate combined with five annual screening examinations reduced relative 15-year lung cancer-specific mortality by 16% and

**Figure 1**



**Figure 1:** Base-case incidence. Model-predicted incident (first) lung cancers per 1000 participants at risk are shown as a function of years since simulated randomization, in screening arm versus control arm. Baseline screening (year 0) yields a high number of NSCLCs; after five screening examinations, the incidence rate in the screening arm decreases below that of the control arm. Table 1 provides further details about baseline screening.

**Figure 2**



**Figure 2:** Base-case stage shift. Model-predicted lung cancer cases (per 1000 participants) detected in the first 5 years after simulated randomization, stratified according to trial arm. Compared with nonscreening cohort, screening yields a higher number of NSCLCs, a greater proportion of which are stage I. Also predicted is a stage shift from stage IV NSCLC to earlier stages. No comparable increase in cases of SCLCs was predicted, and the distribution of limited stage (LS) versus extensive stage (ES) was not substantially changed with screening.

relative 15-year all-cause mortality by 11%.

Reducing the operative mortality rates from lobectomy, mediastinoscopy, and video-assisted thoracoscopic surgery reduced the rate of iatrogenic deaths at 6 years to 7.6 predicted deaths per 100 000 participants in the control arm and 12.4 predicted deaths per 100 000 in the screening arm. Absolute death rates decreased by approximately two-thirds compared with the base-case rate, but the higher rate of follow-up in the screening arm relative to the control arm yielded a relative increase in the risk of death of 63%.

In the scenario of no follow-up for lesions smaller than 4 mm in diameter, fewer iatrogenic deaths occurred in the control arm (12 deaths per 100 000), but the number of deaths in the screening arm was unchanged from the number in the base-case analysis. The percentage of screening participants who underwent excision of benign nodules within 2 years was also unchanged from the base-case rate of 2%. In the scenario of fewer excisional biopsies, 1.4% of participants in the screening arm, versus 0.6% of participants in the control arm, underwent excision of benign nodules within 2 years of simulated randomization.

## Discussion

We populated the LCPM with individuals that were representative of a recent single-arm screening trial (16) and simulated screening and nonscreening scenarios to estimate the effectiveness of screening. In nine of 10 simulated trials of 1520 asymptomatic current and former smokers in each of two study arms (control arm and screening arm with five annual helical CT examinations), screening increased lung cancer detection and survival but yielded nonsignificant ( $P > .05$ ) decreases in lung cancer mortality and overall mortality. By simulating larger trial sizes, however, the LCPM predicted that five annual helical CT screening examinations will reduce lifetime lung cancer mortality by an estimated 8% in this cohort. The overall mortality reduction from lung cancer screening with helical

CT will probably be small (2% reduction in 15-year cumulative all-cause mortality), however, because of high competing mortality risks from smoking (eg, heart disease and other cancers). SCLC represents a substantial percentage of lung cancer cases (15%), but because of its short doubling times and aggressive disease progression, the benefit from detecting SCLC with screening is generally believed to be minimal. We predicted no difference in detection of SCLC in the two trial arms.

Our overall conclusion—that CT screening may offer a moderate lung cancer mortality reduction—lies between the conclusions of two recent high-profile studies, one of which concluded that screening offered a large benefit (9) and the other of which concluded that it offered no benefit (28). In the next paragraphs, we place our findings in the context of these two contradictory claims.

The International Early Lung Cancer Action Program published results from a single-arm collaborative screening study of more than 31 000 individuals (9). The reported 88% 10-year survival of participants with stage I lung cancers is consistent with the LCPM-predicted stage shift and 10-year survival rate of 87% for participants with screening-detected stage I NSCLCs. Although survival is prolonged, the LCPM provides no corroboration for a reduction as large as 80%, as estimated in the discussion section of the article (9), in lung cancer deaths with CT screening. For this cohort, the LCPM predicted that the relative reduction in lifetime lung cancer mortality would increase from 8% with five screening examinations to 14% with 10 screening examinations. With single-arm study designs, changes in mortality cannot be quantified (51).

By using individual data from three single-arm screening studies, the largest of which was the Mayo CT study (16), Bach et al (28) used an existing prediction model to estimate the participants' hypothetical outcomes in the absence of screening. Their main finding was that CT screening in the combined cohorts of a total of 3210 participants offered no significant reduction in lung cancer mor-

tality (relative risk, 1.0; 95% CI: 0.7, 1.3;  $P = .9$ ) (28), but the conclusion of Bach et al that there is no benefit was based on exclusion of all lung cancer deaths in the 1st year of the hypothetical control arm from the mortality calculation. For the Mayo study in particular, if the 26.68 1st-year deaths were not excluded (table 2 in Bach et al [28]), the Bach model would have predicted a 29% relative reduction in lung cancer-specific mortality at 6 years of follow-up, versus 28% from the LCPM (which predicted 26.4 1st-year deaths). The justifications cited by Bach et al (28) for excluding the 1st-year deaths are (a) the requirement that study individuals be asymptomatic at enrollment and (b) the implausibility of an immediate decrease in lung cancer mortality from screening. We did not exclude lung cancer deaths from the 1st year of the hypothetical control arm follow-up, because the LCPM explicitly models symptom detection and therefore the two simulated trial arms would be identical at enrollment, as in a real trial. Similarly, in a microsimulation model such as the LCPM, there is no need to make the rather strong assumption that screening could *never* detect a fast-growing cancer and prevent a lung cancer death within the next year. No lung cancer deaths were observed in the 1st year of the Mayo CT study, but the 95% CI would be wide, given the number of study participants. Bach et al (28) excluded participants older than 80 years of age and lighter smokers (more than 5% of the Mayo CT study participants), so the higher cumulative mortality rate (1.85% vs 1.74% in the LCPM) is not unexpected.

The LCPM predicted a lower relative risk for lung cancer diagnosis in the screening arm versus the control arm (1.3 at 1.5 years of follow-up) than did the model in the study of Bach et al (28) (1.99 at 2 years), but our predicted increase in iatrogenic deaths mirrored the predicted increase in resections that resulted from screening with the model of Bach et al. Unlike the model of Bach et al, the LCPM predicted a decrease in advanced (stage IV) NSCLC because of screening, but the methods for estimat-

ing stage distributions are different. To estimate the proportion of advanced cancers, the prediction model of Bach et al was adjusted by using observed stage distributions in the SEER registries (matched for age and sex but not smoking status, as smoking status is not recorded in the SEER registries). Nonsmokers are known to have less aggressive histologic types of cancer (eg, adenocarcinoma) and therefore may be more likely than smokers to receive a diagnosis of early-stage cancers. As part of the development of the LCPM, the model was populated with specific age, sex, race, and calendar-year cohorts that are representative of the U.S. population in terms of smoking history (ie, including nonsmokers) and was calibrated to stage distributions in corresponding cohorts in the SEER registries.

It is logical that screening will allow diagnosis of some cancers that would not have been apparent during the individual's lifetime: Without screening, the individual dies of unrelated causes, unaware of the asymptomatic cancer. Thus, the question of how much overdiagnosis is too much is difficult to answer on the basis of numbers of cancers in each study arm. Policy makers will benefit from analyses that enumerate the trade-offs between the lung cancer deaths that are prevented and the iatrogenic deaths that are caused by invasive staging and treatment that arise from screening. Cost-effectiveness analyses will be necessary to evaluate whether imaging-based screening is a good use of resources, relative to other interventions, including effective smoking cessation programs or improved treatments.

Volunteer populations are highly selected and probably differ from the general population in unidentified ways. In addition to exclusion of persons who were symptomatic for lung cancer and the possibilities of healthy volunteer bias (28) and endemic histoplasmosis, participants in the Mayo CT screening study were recruited by using television news coverage and 99% were white. Therefore, the results of our analysis may not be predictive of the effect of screening in a broader population of current and

former U.S. smokers. Compared with the Mayo CT study, the National Lung Screening Trial (NLST) (24) enrolled more participants ( $n = 50\,000$ ) with stricter requirements for age (55–74 years) and smoking history ( $\geq 30$  pack-years). Also, in the NLST, CT was compared with chest radiography rather than screening with CT versus no screening, as in our simulation. Because of the differences in participants and trial design, our findings from this study should not be interpreted as evidence that the NLST trial will be underpowered to detect a significant result. However, the NLST is 90% powered to detect a 21% mortality reduction at 6 years of follow-up between screening with CT and screening with chest radiography (52), versus our estimate of a 28% mortality reduction at 6 years of follow-up between screening with CT and no screening. In the event that a significant result is not observed from the NLST, simulation modeling may help in interpretation of those findings.

The analysis in this study had several limitations in addition to those common to all studies with modeling (53). Randomized controlled screening trials have not yet demonstrated a reduction in lung cancer–specific mortality and cannot therefore inform model inputs. The LCPM was calibrated (in the absence of screening) to lung cancer incidence rates in the nine core SEER registries, in which much of the South is excluded and thus may not be generalizable to the entire U.S. population. Information about doubling times and aggressiveness of lung cancers other than the four main histologic types of lung cancer, which together comprise approximately 90% of lung cancers (54), is limited. The LCPM does not generate the identical prevalence as observed in the Mayo CT study; rather, accepting a prevalence within the 95% CI avoids overfitting to the small numbers of study participants and cancers. We predicted a higher rate of excision of benign nodules than was observed in the study, because of our modeling assumption that individuals undergo

guideline staging and treatment. However, with the screening arm, a 2.3-fold increased risk of excisional biopsy was observed for benign disease, even in the scenario with fewer excisional biopsies. A critical output, the increased risk of iatrogenic death caused by screening, was variable and dependent on assumptions about details of the follow-up protocol.

Our assumption that all lung cancer patients (in both simulated study arms) received guideline treatment may not be generalizable to the U.S. population and may lead to overestimation of the survival gains that are possible from screening. In the Mayo study, 14% of participants had incidental findings other than lung cancer at screening examinations. Including any benefit and harm accruing from such incidental findings would affect the predicted overall mortality reduction, although these effects would probably be minimal (55). The LCPM explicitly simulates only lung cancer, so we cannot address questions about whether participation in regular lung cancer screening would influence an individual's overall cancer screening behavior or health care management for other diseases. We omitted possible effects of screening CT on future lung cancer risk (56) and screening behavior (57). The results of our analyses suggest that the rates of smoking cessation achievable in combination with a screening program may greatly influence the overall effectiveness of screening.

Modeling offers a way to use available data—including that from observational or single-arm studies—to inform those making current screening decisions, while they await long-term mortality data from randomized trials. Our results suggest that adding a control arm to the Mayo CT study would have provided some, but not significant ( $P > .05$ ), evidence of a moderate reduction in lung cancer–specific mortality caused by screening.

**Acknowledgments:** We thank members of the Cancer Intervention and Surveillance Modeling Network Lung Group of the National Cancer Institute for invaluable advice and contributions and Aaron Bungum, Mayo Clinic, Rochester,



Minn, for assistance with the transfer of deidentified data and interpretation of coding.

## References

- American Cancer Society. Cancer facts and figures 2007. <http://www.cancer.org/downloads/STT/CAFF2007PWSecured.pdf>. Accessed February 13, 2007.
- American Cancer Society. Prevention & early detection: cigarette smoking. [http://www.cancer.org/docroot/PED/content/PED\\_10\\_2X\\_Cigarette\\_Smoking.asp](http://www.cancer.org/docroot/PED/content/PED_10_2X_Cigarette_Smoking.asp). Accessed February 15, 2007.
- Barbone F, Bovenzi M, Cavallieri F, Stanta G. Cigarette smoking and histologic type of lung cancer in men. *Chest* 1997;112:1474-1479.
- Damber LA, Larsson LG. Smoking and lung cancer with special regard to type of smoking and type of cancer: a case-control study in north Sweden. *Br J Cancer* 1986;53:673-681.
- Naruke T, Goya T, Tsuchiya R, Suemasu K. The importance of surgery to non-small cell carcinoma of lung with mediastinal lymph node metastasis. *Ann Thorac Surg* 1988;46:603-610.
- Winton T, Livingston R, Johnson D, et al. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. *N Engl J Med* 2005;352:2589-2597.
- Mountain CF. Revisions in the International System for Staging Lung Cancer. *Chest* 1997;111:1710-1717.
- Surveillance, Epidemiology, and End Results (SEER) Program ([www.seer.cancer.gov](http://www.seer.cancer.gov)) SEER\*Stat version 5.3.0, database: incidence—SEER 9 registries public use, November 2002 submission (1973-2000). Bethesda, Md: National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2003, based on November 2002 submission. Accessed September 15, 2004.
- International Early Lung Cancer Action Program investigators, Henschke CI, Yankelevitz DF, et al. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;355:1763-1771.
- Marcus PM, Bergstralh EJ, Fagerstrom RM, et al. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst* 2000;92:1308-1316.
- Marcus PM, Bergstralh EJ, Zweig MH, Harris A, Offord KP, Fontana RS. Extended lung cancer incidence follow-up in the Mayo Lung Project and overdiagnosis. *J Natl Cancer Inst* 2006;98:748-756.
- Kubik AK, Parkin DM, Zatloukal P. Czech Study on Lung Cancer Screening: post-trial follow-up of lung cancer deaths up to year 15 since enrollment. *Cancer* 2000;89:2363-2368.
- Humphrey LL, Teutsch S, Johnson M; U.S. Preventive Services Task Force. Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: an update for the U.S. Preventive Services Task Force. *Ann Intern Med* 2004;140:740-753.
- Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999;354:99-105.
- Henschke CI, Naidich DP, Yankelevitz DF, et al. Early Lung Cancer Action Project: initial findings on repeat screening. *Cancer* 2001;92:153-159.
- Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002;165:508-513.
- Sone S, Li F, Yang ZG, et al. Results of 3-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* 2001;84:25-32.
- Black WC. Overdiagnosis: an underrecognized cause of confusion and harm in cancer screening [editorial]. *J Natl Cancer Inst* 2000;92:1280-1282.
- Patz EF Jr, Goodman PC, Bepler G. Screening for lung cancer. *N Engl J Med* 2000;343:1627-1633.
- Reich JM. Improved survival and higher mortality: the conundrum of lung cancer screening. *Chest* 2002;122:329-337.
- Kramer BS, Gohagan JK, Prorok PC, eds. Cancer screening: theory and practice. New York, NY: Marcel Dekker, 1999.
- Ramsey SD, McIntosh M, Etzioni R, Urban N. Simulation modeling of outcomes and cost effectiveness. *Hematol Oncol Clin North Am* 2000;14:925-938.
- Prorok PC, Andriole GL, Bresalier RS, et al. Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control Clin Trials* 2000; 21(6 suppl):273S-309S.
- National Lung Screening Trial. National Cancer Institute Web site. <http://www.cancer.gov/nlst>. Accessed May 4, 2007.
- Collins MM, Barry MJ. Controversies in prostate cancer screening: analogies to the early lung cancer screening debate. *JAMA* 1996;276:1976-1979.
- Feig SA. Screening mammography controversies: resolved, partly resolved, and unresolved. *Breast J* 2005;11(suppl 1):S3-S6.
- Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: 5-year prospective experience. *Radiology* 2005;235:259-265.
- Bach PB, Jett JR, Pastorino U, Tockman MS, Swensen SJ, Begg CB. Computed tomography screening and lung cancer outcomes. *JAMA* 2007;297:953-961.
- McMahon PM. Policy assessment of medical imaging utilization: methods and applications [doctoral thesis]. Cambridge, Mass: Harvard University, 2005.
- Caro JJ, Klittich WS, Strauss G. Could chest x-ray screening for lung cancer be cost-effective? *Cancer* 2000;89:2502-2505.
- Baba Y, Takahashi M, Tominguchi S, Kiyota S. Cost-effectiveness decision analysis of mass screening for lung cancer. *Acad Radiol* 1998;5(suppl 2):S344-S346.
- Chirikos TN, Hazelton T, Tockman M, Clark R. Screening for lung cancer with CT: a preliminary cost-effectiveness analysis. *Chest* 2002;121:1507-1514.
- Mahadevia PJ, Fleisher LA, Frick KD, Eng J, Goodman SN, Powe NR. Lung cancer screening with helical computed tomography in older adult smokers. *JAMA* 2003;289:313-322.
- Marshall D, Simpson K, Earle C, Chu C. Potential cost-effectiveness of one-time screening for lung cancer (LC) in a high risk cohort. *Lung Cancer* 2001;32:227-236.
- Marshall D, Simpson K, Earle C, Chu CW. Economic decision analysis model of screening for lung cancer. *Eur J Cancer* 2001;37:1759-1767.
- Wisnivesky JP, Mushlin AI, Sichertman N, Henschke C. The cost-effectiveness of low-dose CT screening for lung cancer. *Chest* 2003;124:614-621.
- Knudsen AB, McMahon PM, Gazelle GS. Use of modeling to evaluate the cost-effectiveness of cancer screening programs. *J Clin Oncol* 2007;25:203-208.
- Eddy DM. Screening for lung cancer. *Ann Intern Med* 1989;111:232-237.
- Diederich S, Wormanns D, Semik M, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. *Radiology* 2002;222:773-781.
- Taylor KL, Shelby R, Gelmann E, McGuire C. Quality of life and trial adherence among participants in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *J Natl Cancer Inst* 2004;96:1083-1094.

41. McMahon PM, Zaslavsky AM, Weinstein MC, Kuntz KM, Weeks JC, Gazelle GS. Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis. *Med Decis Making* 2006;26:497-511.
42. National health interview survey. National Center for Health Statistics Web site. <http://www.cdc.gov/nchs/nhis.htm>. Accessed April 10, 2008.
43. U.S. Department of Health and Human Services, National Center for Health Statistics, Centers for Disease Control and Prevention. Third National Health and Nutrition Examination Survey (NHANES III), 1988;-94. Public use data file series 11. Hyattsville, Md: Centers for Disease Control and Prevention, 1997.
44. Thun MJ, Myers DG, Day-Lally C, et al. Age and the exposure-response relationships between cigarette smoking and premature death in cancer prevention study II, chap 5. In: Changes in cigarette-related disease risks and their implication for prevention and control, National Cancer Institute Smoking and Tobacco Control monograph 8. NIH publication no. 97-4312. Washington, DC: National Cancer Institute, National Institutes of Health, 1997; 383-475.
45. Beadsmoore CJ, Screaton NJ. Classification, staging and prognosis of lung cancer. *Eur J Radiol* 2003;45:8-17.
46. Capewell S, Sankaran R, Lamb D, McIntyre M, Sudlow M. Lung cancer in lifelong non-smokers. *Thorax* 1991;46:565-568.
47. Beckles MA, Spiro SG, Colice GL, Rudd RM. Initial evaluation of the patient with lung cancer: symptoms, signs, laboratory tests, and paraneoplastic syndromes. *Chest* 2003; 123(1 suppl):97S-104S.
48. Swensen SJ, Jett JR, Hartman TE, et al. Lung cancer screening with CT: Mayo Clinic experience. *Radiology* 2003;226:756-761.
49. Ostroff JS, Buckshee N, Mancuso CA, Yankelevitz DF, Henschke C. Smoking cessation following CT screening for early detection of lung cancer. *Prev Med* 2001;33: 613-621.
50. Cox LS, Clark MM, Jett JR, et al. Change in smoking status after spiral chest computed tomography scan screening. *Cancer* 2003; 98:2495-2501.
51. McMahon PM, Christiani DC. Computed tomography screening for lung cancer [editorial]. *BMJ* 2007;334:271.
52. Marcus P. Lung screening feasibility study. Presented at the Sixth Annual National Veterans Administration Oncology Conference, 2002. <http://www.cancer.gov/prevention/lss/vaslides.html>. Accessed February 7, 2007.
53. Oreskes N, Shrader-Frechette K, Belitz K. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 1994;263:641-646.
54. Rivera M, Detterbeck F, Loomis D. Epidemiology and classification of lung cancer. In: Detterbeck F, ed. *Diagnosis and treatment of lung cancer: an evidence-based guide for the practicing clinician*. Philadelphia, Pa: Saunders, 2001.
55. Beinfeld MT, Wittenberg E, Gazelle GS. Cost-effectiveness of whole-body CT screening. *Radiology* 2005;234:415-422.
56. Brenner DJ. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. *Radiology* 2004; 231:440-445.
57. Ford ME, Havstad SL, Flickinger L, Johnson CC. Examining the effects of false positive lung cancer screening results on subsequent lung cancer screening adherence. *Cancer Epidemiol Biomarkers Prev* 2003;12:28-33.