

Comparative Analysis of 5 Lung Cancer Natural History and Screening Models That Reproduce Outcomes of the NLST and PLCO Trials

Rafael Meza, PhD¹; Kevin ten Haaf, MSc²; Chung Yin Kong, PhD^{3,4}; Ayca Erdogan, PhD⁵; William C. Black, MD⁶; Martin C. Tammemagi, MD⁷; Sung Eun Choi, SM^{3,4}; Jihyouon Jeon, PhD⁸; Summer S. Han, PhD⁵; Vidit Munshi, MS^{3,4}; Joost van Rosmalen, PhD^{2,9}; Paul Pinsky, PhD¹⁰; Pamela M. McMahon, PhD^{3,4}; Harry J. de Koning, MD²; Eric J. Feuer, PhD¹¹; William D. Hazelton, PhD⁸; and Sylvia K. Plevritis, PhD⁵

BACKGROUND: The National Lung Screening Trial (NLST) demonstrated that low-dose computed tomography screening is an effective way of reducing lung cancer (LC) mortality. However, optimal screening strategies have not been determined to date and it is uncertain whether lighter smokers than those examined in the NLST may also benefit from screening. To address these questions, it is necessary to first develop LC natural history models that can reproduce NLST outcomes and simulate screening programs at the population level. **METHODS:** Five independent LC screening models were developed using common inputs and calibration targets derived from the NLST and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO). Imputation of missing information regarding smoking, histology, and stage of disease for a small percentage of individuals and diagnosed LCs in both trials was performed. Models were calibrated to LC incidence, mortality, or both outcomes simultaneously. **RESULTS:** Initially, all models were calibrated to the NLST and validated against PLCO. Models were found to validate well against individuals in PLCO who would have been eligible for the NLST. However, all models required further calibration to PLCO to adequately capture LC outcomes in PLCO never-smokers and light smokers. Final versions of all models produced incidence and mortality outcomes in the presence and absence of screening that were consistent with both trials. **CONCLUSIONS:** The authors developed 5 distinct LC screening simulation models based on the evidence in the NLST and PLCO. The results of their analyses demonstrated that the NLST and PLCO have produced consistent results. The resulting models can be important tools to generate additional evidence to determine the effectiveness of lung cancer screening strategies using low-dose computed tomography. *Cancer* 2014;120:1713–24. © 2014 American Cancer Society.

KEYWORDS: Cancer Intervention and Surveillance Modeling Network (CISNET), comparative modeling analyses, low-dose CT screening, lung cancer screening, cancer natural history models, smoking and lung cancer, simulation model.

INTRODUCTION

The National Lung Screening Trial (NLST) found a significant lung cancer (LC) mortality reduction in its low-dose computed tomography (CT) screening arm in comparison with its chest radiography (CXR) screening arm,¹ suggesting that screening heavy smokers with low-dose CT can be effective in the early detection of LC. Meanwhile, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) found no statistical difference in LC mortality when comparing a no-screen control arm versus a CXR screening arm.² Consequently, several health policy groups have made recommendations endorsing low-dose CT for LC screening based on the NLST entry criteria and LC screening programs are being established across the United States.³ However, there is still uncertainty regarding the optimal screening strategies because the NLST only evaluated the impact of 3 consecutive annual screens among current and former smokers aged 55 years to 74 years at the time of enrollment with an exposure of at least 30 pack-years and with ≤ 15 years since quitting. It is unknown whether current and former smokers with lower levels of exposure would also benefit from screening. Furthermore, screening effectiveness may vary by sex, number of screens, and periodicity. In the absence of results from other randomized control trials (RCTs) evaluating these questions, mathematical modeling of the natural

Corresponding author: Rafael Meza, PhD, Department of Epidemiology, University of Michigan, 1415 Washington Heights, Ann Arbor, MI; Fax: (734) 936-2084; rmeza@umich.edu

¹Department of Epidemiology, University of Michigan, Ann Arbor, Michigan; ²Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands; ³Institute for Technology Assessment, Massachusetts General Hospital, Boston, Massachusetts; ⁴Department of Radiology, Harvard Medical School, Boston, Massachusetts; ⁵Department of Radiology, Stanford University, Stanford, California; ⁶Department of Radiology, Geisel School of Medicine at Dartmouth, Dartmouth College, Hanover, New Hampshire; ⁷Department of Community Health Sciences, Brock University, St. Catharines, Ontario, Canada; ⁸Program in Biostatistics and Biomathematics, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; ⁹Department of Biostatistics, Erasmus University Medical Center, Rotterdam, the Netherlands; ¹⁰Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland; ¹¹Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland

Additional supporting information may be found in the online version of this article.

DOI: 10.1002/cncr.28623, **Received:** October 4, 2013; **Revised:** December 3, 2013; **Accepted:** December 5, 2013, **Published online** February 27, 2014 in Wiley Online Library (wileyonlinelibrary.com)

history of LC may be the only approach to integrate available evidence and estimate the effectiveness and cost-effectiveness of different LC screening strategies in the general population.^{3,4}

Mathematical models of cancer natural history have been shown to be valuable in assessing and determining optimal cancer prevention and control strategies. Recent examples include analyses of the impact of tobacco control on LC mortality rates,⁵ comparative studies assessing the effects of different screening modalities in patients with colorectal cancer,⁶ cost-effectiveness analyses of breast cancer screening strategies,⁷ and studies evaluating the impact of prostate-specific antigen screening in reducing prostate cancer rates.^{8,9} All these examples used a comparative modeling framework by which researchers across institutions can directly compare and contrast results from distinct models.¹⁰⁻¹² The conclusions arising from comparative modeling analyses are more robust and reliable than single-model studies and this approach has been cited as an example of good modeling practices.¹³

To estimate the potential impact of LC screening at the US population level, a consortium of National Cancer Institute (NCI)-sponsored investigators, the Cancer Intervention and Surveillance Modeling Network (CISNET; cisnet.cancer.gov), developed 5 independent natural history models of LC and screening. In the current study, we describe the models' development and calibration approach to the NLST and PLCO, the common shared inputs and calibration targets, and the differences and similarities between models. We compared model predictions versus observed trial outcomes and highlighted the advantages and challenges of developing natural history models based on large-scale RCTs.

MATERIALS AND METHODS

Data

Deidentified data from all NLST and PLCO participants were provided to CISNET after obtaining Institutional Review Board approvals from each institution. These data included smoking history variables such as the age at the start of smoking, the average number of cigarettes smoked per day (CPDs), and the age at quitting for former smokers. Screening variables included the individual's age at entry into the study and, for screened individuals, age at each screen, outcomes of each screen, and the follow-up procedures for positive screens. For each individual, the age at death or censoring and (if applicable) the cause of death were available. For individuals diagnosed with LC, the age at diagnosis, LC

histology, and LC stage (according the 6th edition of the American Joint Committee on Cancer) were provided, as well as information regarding the screen associated with the LC diagnosis for screen-detected cancers.

NLST

The NLST was a RCT that compared the impact of low-dose CT versus CXR screening on LC mortality. From August 2002 through April 2004, a total of 53,454 individuals aged 55 years to 74 years were recruited; follow-up occurred through December 31, 2009. Entry criteria included a minimum exposure of 30 pack-years and ≤ 15 years since quitting for former smokers. Individuals in both screening arms received up to 3 annual screens. The trial found a 20% LC mortality reduction in the low-dose CT versus the CXR arm.¹

A small percentage of LC cases (50 cases; 2.4%) had missing histology and/or stage information. To complete the missing data, a multistep imputation procedure based on observed histology and stage distributions, tumor sizes, and expert opinion was conducted. More details are provided in the supplementary material. Final analyses included data from 53,342 individuals, due to the exclusion of 112 subjects who died or were diagnosed with LC before the first screen (110 patients) or those with missing smoking information (age at start and/or time since quitting).

PLCO

The PLCO was a RCT that compared the impact of CXR screening (intervention arm) versus usual care (no-screening control arm) on LC mortality. The trial recruited 154,901 individuals aged 55 years to 74 years between November 1993 and July 2001. Participants were followed through December 31, 2009 or for 13 years from the time of enrollment, whichever came first. No minimum smoking exposure was required to enroll. Individuals in the intervention arm received up to 4 annual CXR screens. The study found no difference in LC mortality between the intervention and control arms.² Contamination (CXR screening) in the control arm was limited (11% contamination rate³).

Additional smoking variables came from a supplemental questionnaire implemented toward the middle of the trial. Missing baseline data regarding the age at the start of smoking or CPDs for ever-smokers were imputed according to the corresponding US distributions by birth cohort and age. Final analyses included data from 148,025 individuals, after the exclusion of individuals with missing baseline smoking status or (if applicable) age at time of quitting. For more details please see supplementary material.

Models

Models were developed by investigators at 5 institutions: Erasmus Medical Center (model E), Fred Hutchinson Cancer Research Center (model F), Massachusetts General Hospital (model M), University of Michigan (model U), and Stanford University (model S). The models were developed independently but the groups collaborated to develop common inputs and define standardized analyses. Below we provide a description of the five models. Additional details are provided in the Supplementary Material.

Smoking dose-response module

All models simulate individual LC natural history and include a dose-response module that translates personal cigarette exposure to LC risk. This smoking dose-response module can be used to simulate age-specific LC outcomes given an individual's smoking history.⁵ Model M uses as its dose-response module a probabilistic LC risk model previously calibrated to Surveillance, Epidemiology, and End Results (SEER) and US LC data^{14,15} and recalibrated to the NLST and PLCO, whereas all other groups use multistage carcinogenesis models.¹⁶⁻¹⁸ Both multistage^{5,16,17,19} and probabilistic models have been used extensively to investigate the effects of smoking on LC risk.^{12,20,21} Model E uses a multistage model based on the Nurses' Health Study (NHS) and Health Professionals Follow-Up Study (HPFS).¹⁶ Model S uses a modified version of this model. Model U uses a LC multistage model by histology, also calibrated to the NHS/HPFS. Model F uses a multistage model calibrated to the NLST and PLCO. Three models (models F, M, and U) use histology-specific smoking dose-response modules, and 3 models (models E, F, and M) recalibrated their smoking dose-response to the NLST and PLCO. More details are given in Table 1 and in the supplementary material. All models are capable of accommodating detailed individual-level smoking histories, including temporal factors such as age at start, age at cessation, and age-specific changes in CPDs. The variability across dose-response modules reflects the modelers' judgment regarding the best available data and approaches to capture the complex relationship between smoking and LC. The NHS and HPFS are arguably the best prospective cohorts with which to investigate smoking-related LC. They have > 30 years and > 20 years of follow-up, respectively, and collect smoking information every 2 years. However, their LC histology information is much less comprehensive than that of the NLST and PLCO, and staging information was not available. The NLST and PLCO are excellent data sources with thorough information available regarding LC histology and staging, but have more limited follow-up and less extensive smoking data than

NHS/HPFS. In addition, the NLST includes only ever-smokers and individuals in both arms were screened for LC. Approximately one-half of individuals in PLCO were also screened.

Histology distribution

Three models (models F, M, and U) have smoking dose-response modules that are histology-specific. In these models, the LC histology distribution is a model outcome that depends on the dose-response module and the participants' smoking histories. Two models (models E and S) have smoking dose-response modules that are not histology-specific, and therefore they calibrated their histology to the NLST and PLCO. Histology categories varied by model (Table 1). Differences in histology categorization across models are due partly to differences in dose-response modules, which are based on different data sets that vary in their LC histology classifications (NHS/HPFS, NLST/PLCO, and SEER). However, they are also due to variations in model structure, and the modelers' judgment regarding the histology detail needed to characterize screening efficacy.

Stage progression

All models assume that stage progression rates vary by sex and histology. Models E and U use Markov state transition processes to model stage progression.²² Model U further assumes that the progression rate at each stage is dependent on tumor size (cell number). Models F, M, and S model stage as a function of tumor size and the presence or absence of metastasis. Variability in stage categorization (Table 1) is due to the underlying data inputs, model structure, and the modelers' criteria regarding the stage detail needed to capture the effects of screening on LC mortality.

LC survival

All models assume that LC survival varies by histology and stage. Models F, M, S, and U also assume that survival varies by sex. Model U further assumes that survival varies by age at diagnosis.

Models E, M, and U use LC survival modules calibrated to the SEER 17 (2004-2008) survival. Survival in model S was calibrated to SEER 17 (1988-2003) survival. LC survival in model F was calibrated to the NLST and PLCO.

Other-cause mortality

Model E uses an other-cause mortality (OCM) module based on the NCI's smoking history generator, which produces OCM rates consistent with the US population.^{23,24} All other models use OCM based on the NLST and PLCO (Table 1).

TABLE 1. Model Comparison

	Model E	Model F	Model U	Model M	Model S
Central smoking dose-response model	TSCE	Longitudinal multistage observation by histology	Multistage clonal expansion model by histology	Probabilistic by histology	TSCE
Central dose-response parameter calibration	NHS/HPFS,SEER,NLST,PLCO	NLST, PLCO; PLUSS CT, CARET	NHS/HPFS	SEER, NLST, PLCO	NHS/HPFS
Histological types	Adenocarcinoma + BAC + Large Cell, squamous, SCLC, ONSCLC IA, IB, II, IIIA, IIIB, IV	Adenocarcinoma, large cell, squamous, BAC, ONSCLC, SCLC IA1, IA2, IB, II, IIIA, IIIB,IV	Adenocarcinoma+BAC, ONSCLC, SCLC IA1, IA2, IB, II, IIIA, IIIB, IV	Adenocarcinoma, BAC, large cell, squamous, SCLC, and other IA1, IA2, IB, II, IIIA, IIIB, IV	Adenocarcinoma, large cell, squamous, SCLC
LC stages	Markov state-transition by histology	Based on tumor size and presence of metastasis	Markov state-transition by histology and sex; rates proportional to tumor size	Based on tumor volume and meta-static burden	Early (I-II), advanced (III-IV)
Stage progression model					Based on tumor volume and meta-static burden
LC survival	Based on SEER 17 2004–2008 survival	Based on NLST and PLCO	By sex, histology, stage, and age at diagnosis. Based on SEER 17 2004–2008 survival	Calibrated to SEER 17 2004–2008 survival	Based on SEER 17 1988–2003 survival
OC mortality	US rates (NCI Smoking History Generator)	As observed in NLST and PLCO	Gompertz model of OC mortality calibrated to each trial	Cox-model of OC mortality calibrated to each trial	Gompertz model of OC mortality based on NLST
Calibration method	Neider-Mead optimization of likelihood-based deviance criterion	Maximum likelihood approach	MCMC and Neider-Mead simplex	Simulated annealing based on weighted-sum total deviance	Neider-Mead simplex for natural history model calibration to SEER, and multidimensional grid search for calibration to trials
Data sources used for calibration	NLST, PLCO, SEER 17 (2004–2008) incidence by age, stage, histology; NHS/HPFS	NLST, PLCO; originally fitted to PLUSS CT, CARET	NLST, NHS/HPFS LC incidence by histology; SEER LC survival by sex, age, histology, and stage	NLST, SEER 1990–2000 incidence by age, stage, histology; survival by stage; Mayo CT; LSS	NLST, PLCO, NHS/HPFS LC incidence, SEER 1988–2003 survival by histology and sex
No. of parameters estimated by calibration	110	90	50	53	13 in natural history, 8 for calibration
Screening sensitivity model by modality (CT vs CXR)	By stage and histology	By size (no. of cells), histology, and sex	By size (no. of cells), histology, and sex	By size (mm) and location in lung (central/peripheral)	By size (mm) and histology
Screening effectiveness mechanism	Cure model	Combination cure model and stage-shift model	Stage-shift model, with adjustments for age	Not stage-shift model	Not stage-shift model
Positive nodule follow-up algorithm	Implicit	Implicit based on NLST follow-up rates	Implicit based on NLST follow-up rates	Explicit; based on size at diagnosis and smoking history. LCs diagnosed on follow-up are categorized as “non-screen-detected”	Explicit

Abbreviations: BAC, bronchioloalveolar carcinoma; CARET, Carotene and Retinol Efficacy Trial; CT, computed tomography; HPFS, Health Professionals' Follow-up Study; LC, lung cancer; LSS, Lung Screening Study; MCMC, Markov chain Monte Carlo; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model S, Stanford University; model U, University of Michigan; NCI, National Cancer Institute; NHS, Nurses' Health Study; NLST, National Lung Screening Trial; OC, other cause; ONSCLC, other non-small cell lung cancer; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PLUS, Pittsburgh Lung Screening Study; SCLC, small cell lung cancer; SEER, Surveillance, Epidemiology, and End Results; TSCE, Two-stage clonal expansion model.

Screening and follow-up

Screening sensitivities vary by model. In model E, screen sensitivity varies by modality, stage, and histology. Models F and U have screen sensitivities that also vary by tumor size (cell number). Sensitivities in models M and S depend on screening modality, tumor size (in mm), and lung nodule location (central vs peripheral). Model S also considers histology. The variability in assumption is primarily due to differences in model structure (eg, models that do not model tumor size explicitly cannot have size-dependent sensitivities). Follow-up examinations are defined as those received after a positive screen but before diagnosis, if it occurred. Algorithms for follow-up of a positive screen are simulated with varying detail; models M and S include detailed algorithms based on nodule size thresholds and risk factors (explicit), whereas models E, F, and U incorporate a global probability of receiving several follow-up examinations (implicit) based on the observed frequency of imaging examinations per positive screen in the NLST. Because the NLST and PLCO did not specify a follow-up regimen, models M and S specify less aggressive protocols than the Fleischner Society guidelines,²⁵ to approximate the observed follow-up rate in the NLST.

Trial simulations

Four models (models E, M, S, and U) generate individual LC outcomes using microsimulations.²⁶ The simulation depends on individual smoking history, sex, age at enrollment, and screening arm. The specific simulation approach depends on the model's structure. Three models (models E, M, and S) simulate age at onset of lung tumors via their smoking dose-response module and then simulate each tumor's natural history, including malignant conversion, stage progression (models E, M, and S), tumor growth (models M and S), and clinical and screen detection (models E, M, and S). Model U simulates the initiation of tumors via mutations of normal cells, and then the premalignant and malignant tumor cell dynamics (cell division, death, stage progression, and clinical and screen detection). Model F uses a likelihood-based approach to estimate LC outcomes and death via a longitudinal, multistage, observation model.¹⁸ All models simulate all trial participants and then compare their aggregate modeled outcomes with those of the trials (LC incidence and mortality and OCM by screening arm, sex, histology, and stage).

Screening effectiveness and mortality reduction

All models evaluate screening effectiveness, but based on different assumptions that depend on model structure.

Model M assumes that patients with early-stage non-small cell LC (NSCLC) would undergo resection (lobectomy, consistent with practice guidelines), which removes the primary tumor. In model M, therefore, for patients without undetected distant metastases or additional primary LCs in another lobe, resection is curative for LC. In model U, the benefit of screening is due to the early detection of LC, leading to improved cure probabilities and survival times, which depend on histology, stage, sex, and age at diagnosis, but not on detection mode. Model F assumes that screen-detected cancers are treated according to clinical practice guidelines with cure rates that vary by tumor stage and histology. In model E, screen-detected patients experience a reduced risk of LC mortality versus clinically detected cases. This improved prognosis is represented as a cure fraction (dependent on stage and screening modality for stages IA, IB, and II) calibrated to the trials. Model S estimates probabilities of lethal metastases as function of tumor size, histology, and sex. All advanced stage LCs are, by definition, detected after the onset of lethal metastases. Some early-stage cancers may have occult lethal metastases at the time of detection. For patients with early-stage and late-stage tumors detected after the onset of lethal metastases, LC survival is not affected by screening. However, with screening, patients are more likely to be detected at early stages before the onset of lethal metastases, and therefore are cured of their disease after standard care.

Model Calibration and Validation Approach

Models were first calibrated to the NLST LC incidence and mortality by screening arm, sex, histology, stage, and detection mode. Models were then validated against PLCO by first comparing model predictions and observed LC incidence and mortality by sex and screening arm in the subset of individuals in PLCO who would have been eligible for the NLST (PLCO-NLST-eligible). Model predictions were consistent with the observed outcomes in the PLCO-NLST-eligible group, demonstrating the consistency between the 2 trials. However, model outcomes did not consistently match against observed outcomes among PLCO participants who were not eligible for the NLST (never-smokers and light smokers). As a result, models were further calibrated to fit the whole PLCO data set to ensure that they could be used with confidence to extrapolate the effects of CT screening to smokers with lower exposure (< 30 pack-years). Calibration methods (targets, measures of goodness of fit, and optimization algorithms) varied by

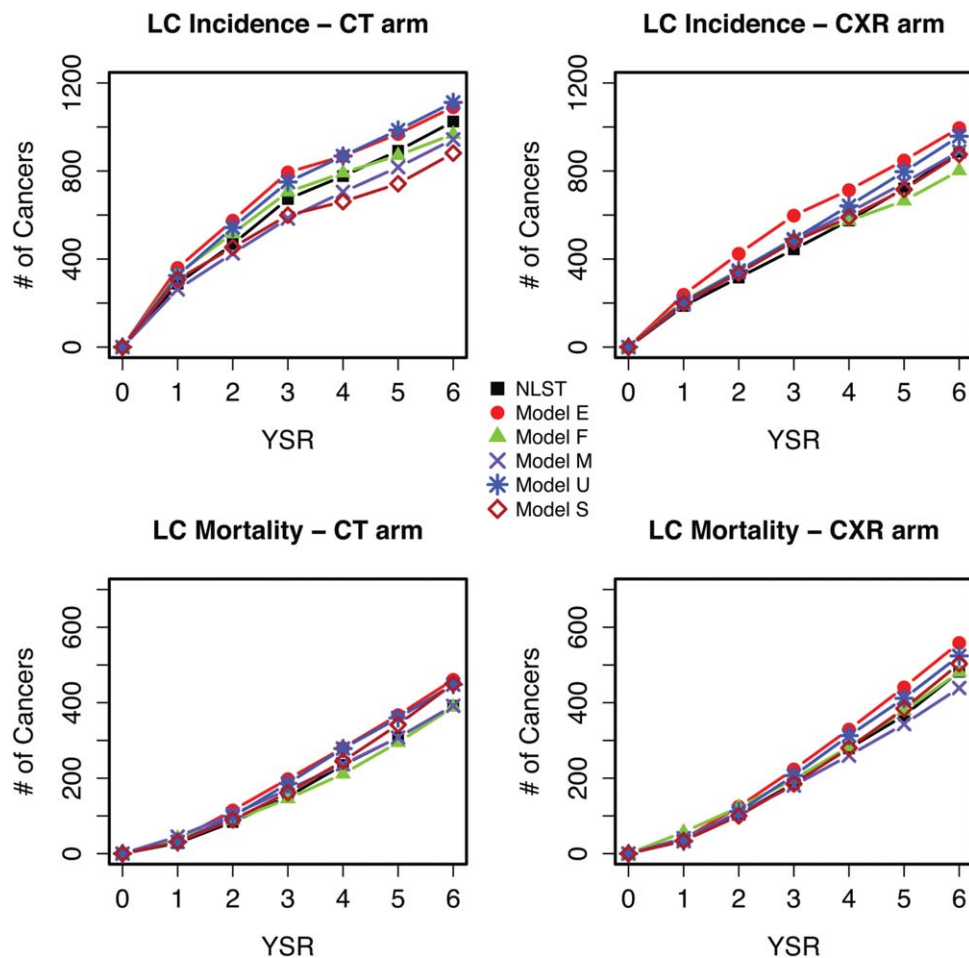


Figure 1. National Lung Screening Trial (NLST) observed and modeled incidence and mortality are shown by screening arm and years since randomization (YSR). LC indicates lung cancer; CT, computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.

model and are described in Table 1 and in the supplementary material.

RESULTS

After final calibration, all models produced LC outcomes consistent with both trials (within the confidence intervals of the data). We demonstrated several measures of LC incidence and mortality in the NLST and PLCO for both sexes combined and compared observed and model outcomes. Calibration targets varied by model, and therefore the modeling results shown in each figure include combinations of calibrated outcomes and model predictions/extrapolations. Modeled outcomes were computed using the “final” version of each model.

Figure 1 shows NLST observed and modeled incidence and mortality by screening arm and years since randomization (YSR). The figure shows that as previously

reported,¹ the observed cumulative LC incidence was higher in the CT screening arm, whereas the cumulative mortality was higher in the CXR screening arm. Figures 2 and 3 display observed versus modeled LC cases and deaths in the NLST by detection modality (screen-detected vs non-screen-detected), screening arm, and YSR. The figures show the contrasting pattern between screen-detected and non-screen-detected cancers, with an early increase and peaking by YSR for screen-detected cancers in both screening arms, in contrast with the slow progressive rise for non-screen-detected cancers. The figures indicate that the models reproduce the general patterns of incidence and mortality by screening arm, detection modality, and YSR.

Figure 4 shows observed versus model-predicted LCs in the NLST by histology. Because models have varying LC histology categories, we grouped them here as small cell LC (SCLC) and NSCLC. The figure shows that

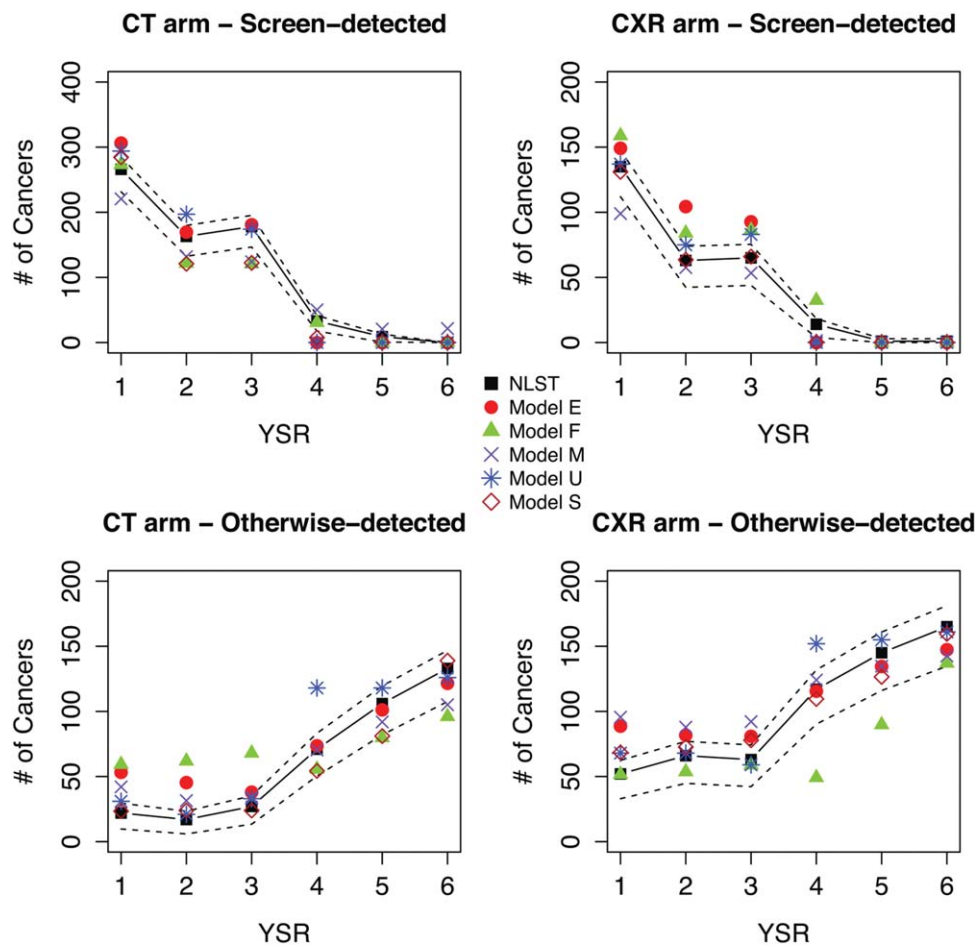


Figure 2. Observed versus modeled lung cancer cases in the National Lung Screening Trial (NLST) are shown by detection modality (screen vs non-screen-detected), arm and years since randomization (YSR). CT indicates computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University. Dashed lines represent 95% binomial CIs for the observed values. Observed screen-detected cancers after year 3 are due to delay in diagnosis after the last screen.

the observed NSCLC incidence was higher in the CT arm, whereas the SCLC incidence was approximately similar in both screening arms. Modeled histology distributions matched well with the observed distributions. Figure 5 shows the NLST observed versus predicted NSCLC incidence by clinical stage and screening arm. The figure demonstrates the shift toward earlier stages in NSCLC incidence in the CT versus CXR arm.

Figures 6 and 7 show full PLCO and PLCO-NLST-eligible observed and modeled deaths by screening arm, detection mode (CXR arm), and YSR. The figures display the early increase and peaking of screen-detected cancers in the CXR arm by YSR, and the slower increase of otherwise detected cancers in the CXR arm and for all cancers in the control arm. The figures demonstrate a decrease in the non-screen-detected cancers in the CXR

and control arms toward the end of the trial, most likely due to the weeding out and loss to follow-up of high-risk individuals. All models reproduce the general patterns of incidence and mortality in PLCO.

DISCUSSION

Main Findings

We derived 5 independent LC and screening natural history models calibrated to the 2 largest screening trials to date, the NLST and PLCO. The 5 models are diverse in structure, assumptions, and additional data inputs. All models produce outcomes that are generally consistent with the trial results. We found that models calibrated only to the NLST validated well against the PLCO-NLST-eligible population, thereby demonstrating the consistency between the 2 trials. However calibrating only

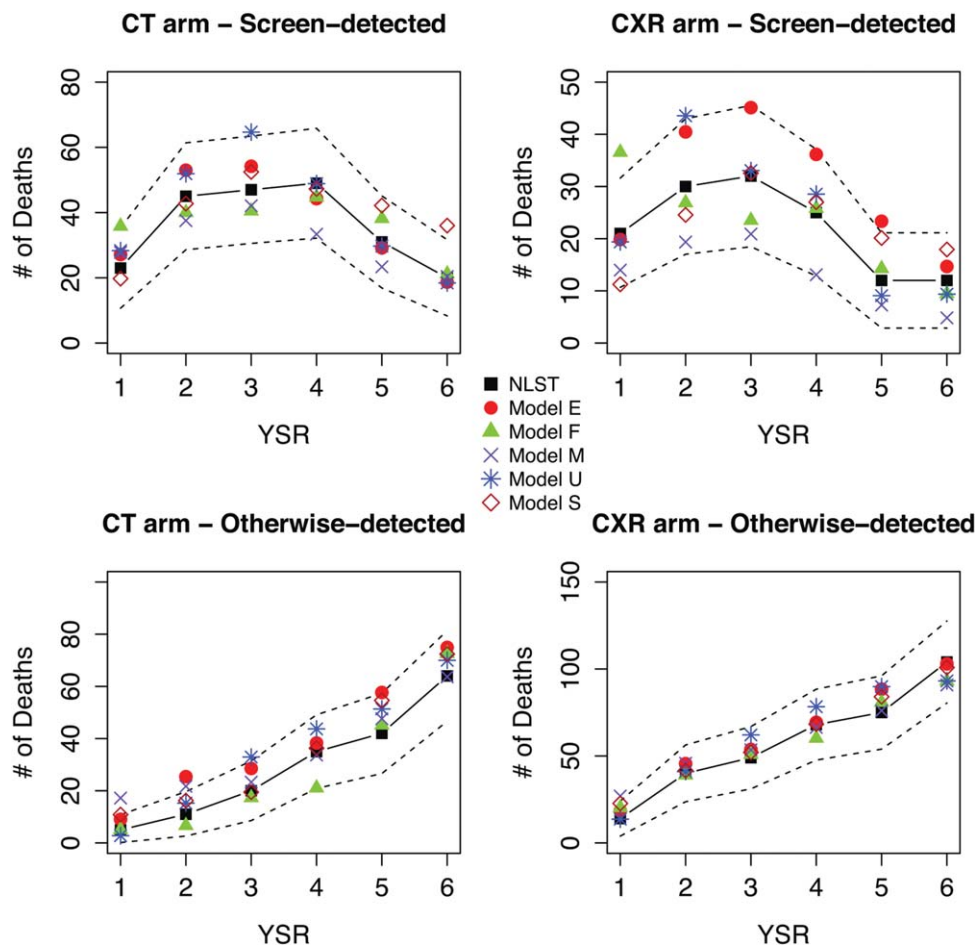


Figure 3. Observed versus modeled lung cancer deaths in the National Lung Screening Trial (NLST) are shown by detection modality (screen vs non-screen-detected), arm and years since randomization (YSR). CT indicates computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University. Dashed lines represent 95% binomial CIs for the observed values.

to the NLST may be insufficient for the purposes of evaluating screening protocols, allowing for lower smoking exposures, and making projections for the US population. This is particularly true for models that base their smoking dose-response fully on the NLST and also for models with histology distributions based on observed trial data, because the NLST only includes information regarding current and former heavy smokers and it is well-documented that the LC risk from smoking varies greatly by histology.^{27,28} To derive models that could be used with confidence to extrapolate the impact of low-dose CT screening to smokers with lower exposures (< 30 pack-years) and to the US population, it is essential to calibrate such models to data sets with information on LC risk for light and never-smokers, such as NHS/HPFS or PLCO.

Study Limitations and Strengths

The current study has some limitations. First, as in any mathematical modeling approach, our models are simplifications of the biological complexity of lung carcinogenesis and neglect the influence of various endogenous and exogenous LC risk factors such as family history, chronic obstructive pulmonary disease, residential radon, occupational exposures, race, and socioeconomic status. However, it is well known that smoking still accounts for the large majority of LC deaths ($\geq 90\%$ ²⁹) and our models do capture the complex relation between smoking and LC via their smoking dose-response module. Furthermore, in contrast with the majority of LC risk models in the literature, several of our models do account explicitly for the differential impact of smoking on LC risk by histology. The diversity in model structure, assumptions,

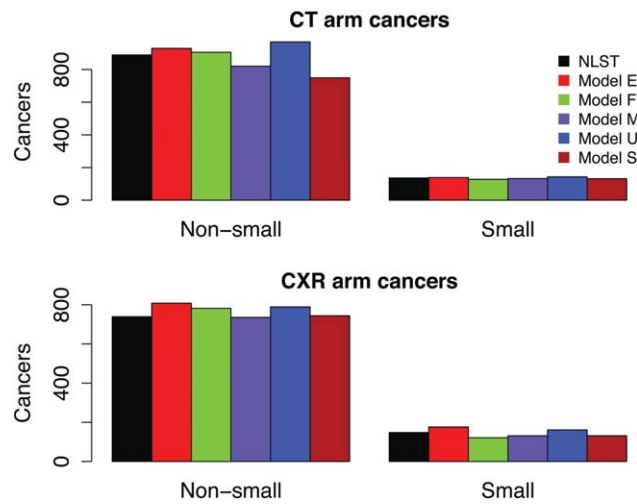


Figure 4. Observed versus model-predicted lung cancers in the National Lung Screening Trial (NLST) are shown by histology. CT indicates computed tomography; CXR, chest radiography; model E, Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University.

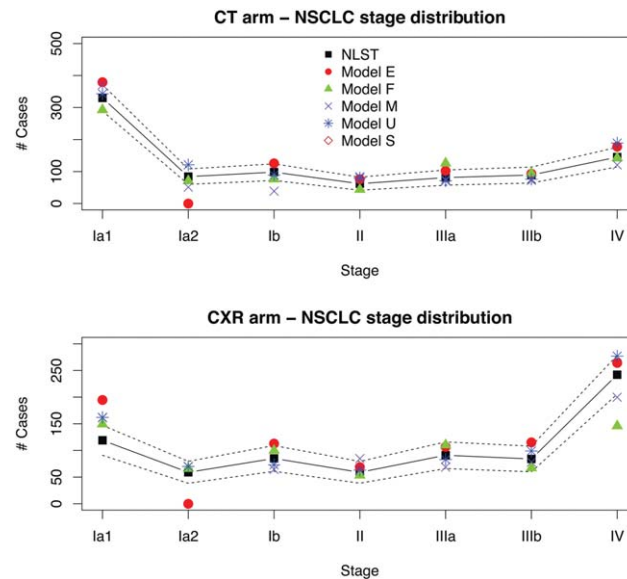


Figure 5. National Lung Screening Trial (NLST) observed versus predicted non-small cell lung cancer (NSCLC) incidence is shown by clinical stage and screening arm. The figure demonstrates the shift toward earlier stages in NSCLC incidence in the computed tomography (CT) versus chest radiography (CXR) arm. Model E indicates Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University. Dashed lines represent 95% multinomial CIs for the observed values. Model E does not model separately Ia1 and Ia2 cancers, so their Ia1 value represents all IA cancers. Model S models early versus late stage cancers.

and data sources provides additional strength (and an assessment of model uncertainty) to the conclusions of our comparative modeling analysis, as does the long history of collaboration between the CISNET groups.

Another potential limitation of the current study is that the screening mortality reductions predicted by each

model are largely dependent on the findings of the NLST and PLCO. To our knowledge the NLST and PLCO are currently the best existing studies of LC screening reporting on the main outcome of LC mortality (reduction), and therefore calibrating models to these trials is the best available option. Some other studies, particularly in

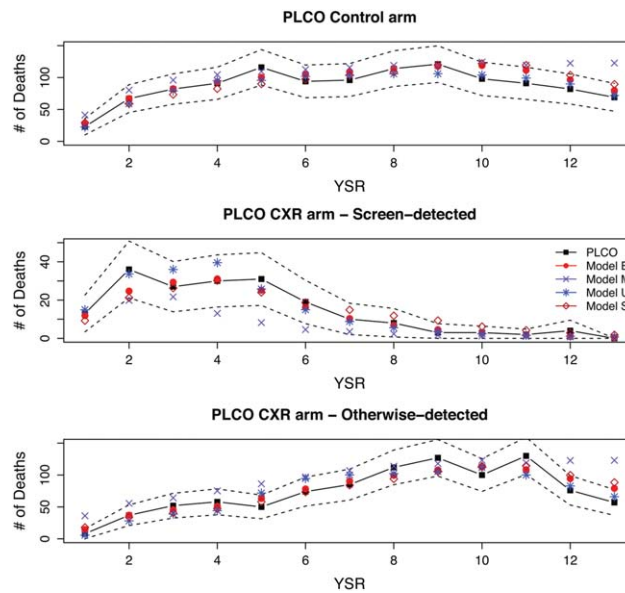


Figure 6. Full Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) observed and modeled deaths are shown by screening arm, detection mode (chest radiography [CXR] arm), and years since randomization (YSR). Model E indicates Erasmus Medical Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University. Dashed lines represent 95% binomial CIs for the observed values.

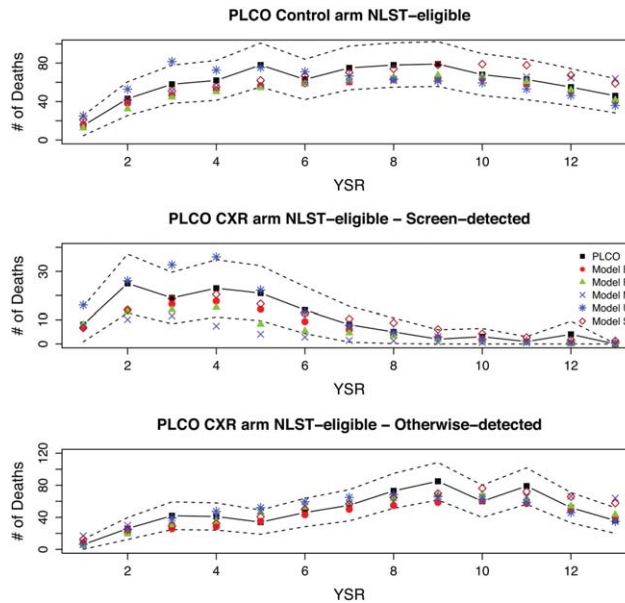


Figure 7. Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)-National Lung Screening Trial (NLST)-eligible observed and modeled deaths are shown by screening arm, detection mode (chest radiography [CXR] arm), and years since randomization (YSR). Model E indicates Erasmus Medical Center; model F, Fred Hutchinson Cancer Research Center; model M, Massachusetts General Hospital; model U, University of Michigan; model S, Stanford University. Dashed lines represent 95% binomial CIs for the observed values.

Europe, have been underpowered to demonstrate the benefits of low-dose CT screening whereas others are still ongoing.³⁰ Once data from other trials become available,

which is not expected for a few years, the models could be validated against new trials and, if deemed necessary, calibrated further, particularly if applied to non-US

populations. In any case, the models will be helpful to compare trial results and, if needed, to investigate the reasons behind any potential discrepancies.

Finally, the current study highlights the benefits of modeling as a way to synthesize information coming from diverse and complex data sources. The models developed use individual data from RCTs (the NLST and PLCO), prospective cohort studies (eg, NHS/HPFS), and cancer registry data (NCI-SEER). These data sources are extremely valuable on their own, and provide information regarding different aspects of LC. However, it is only through modeling that they can be integrated and jointly inform the biology and epidemiology of LC, as well as the potential benefits of LC screening at the population level.

Implications and Future Research

The results of the current analyses demonstrate that the NLST and PLCO produced consistent results, and suggest that it is critical to use data covering a wide range of smoking histories (never-smokers, light smokers, and heavy smokers) to develop models that can extrapolate the effects of screening to the general population. The 5 models presented herein are currently being used to evaluate the impact of alternative low-dose CT screening protocols on LC mortality in the United States. Specifically, we are assessing the effectiveness of screening programs with varying age eligibility, exposure criteria, and screening frequency.³¹ In the near future, we will use the models to predict the potential levels of overdiagnosis due to LC screening and determine optimal screening strategies at both the national and state levels. Using models calibrated to the NLST and PLCO will enhance the validity of effectiveness and cost-effectiveness analyses of LC screening.

FUNDING SUPPORT

This report is based on research conducted by the National Cancer Institutes (NCI)'s Cancer Intervention and Surveillance Modeling Network (CISNET) through support from an interagency agreement with the Agency for Healthcare Research and Quality (AHRQ) (administrative supplement to NCI grant U01-CA152956).

CONFLICT OF INTEREST DISCLOSURES

Dr. Meza received an NCI grant for work related to the current study. Mr. ten Haaf and Dr. de Koning received an NCI/National Institutes of Health (NIH) grant for work conducted by CISNET through support from an interagency agreement with the AHRQ and an AHRQ administrative supplement to grant U01-CA152956. In addition, Mr. ten Haaf and Dr. de Koning have received a grant from Sunnybrook Health Sciences (Toronto, Ontario, Canada) (project name: Health Technology Assessment for CT Lung Screening Including MISCAN Modeling of Outcomes

and Cost-Effectiveness) and a grant from the NELSON-Netherlands-Leuven Lung Cancer Screening Trial supported by Zorg Onderzoek Nederland-Medische Wetenschappen (ZonMw), KWF Kankerbestrijding, and Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen (RVVZ) for a trial conducted outside of the current study. For this same trial, Roche Diagnostics provided a grant for the performance of proteomics research and Siemens Germany provided 4 digital workstations and LungCARE for the performance of 3-dimensional measurements. The Roche Diagnostics Medical Advisory Board provided the Department of Public Health of Erasmus University Medical Center with €1500 for a Medical Advisory Board meeting. Dr. Black has received an NIH consulting grant for work related to the current study. Dr. McMahon was supported by NCI grant U01CA152956 to Massachusetts General Hospital and an AHRQ supplement to this grant for work related to the current study.

REFERENCES

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* 2011;365:395-409.
2. Oken MM, Hocking WG, Kvale PA, et al;PLCO Project Team. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial. *JAMA.* 2011;306:1865-1873.
3. Bach PB, Mirkin JN, Oliver TK, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA.* 2012;307:2418-2429.
4. Bach PB, Gould MK. When the average applies to no one: personalized decision making about potential benefits of lung cancer screening. *Ann Intern Med.* 2012;157:571-573.
5. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975-2000. *J Natl Cancer Inst.* 2012;104:541-548.
6. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med.* 2012;366:687-696.
7. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med.* 2009;151:738-747.
8. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. *Cancer.* 2012;118:5955-5963.
9. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med.* 2012;367:595-605.
10. Feuer EJ, Etzioni R, Cronin KA, Mariotto A. The use of modeling to understand the impact of screening on U.S. mortality: examples from mammography and PSA testing. *Stat Methods Med Res.* 2004;13:421-442.
11. Kuntz KM, Lansdorp-Vogelaar I, Rutter CM, et al. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. *Med Decis Making.* 2011;31:530-539.
12. McMahon PM, Hazelton WD, Kimmel M, Clarke LD. Chapter 13: CISNET lung models: comparison of model assumptions and model structures. *Risk Anal.* 2012;32(suppl 1):S166-S178.
13. Weinstein MC, O'Brien B, Hornberger J, et al;ISPOR Task Force on Good Research Practices-Modeling Studies. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices-Modeling Studies. *Value Health.* 2003;6:9-17.
14. McMahon PM, Kong CY, Johnson BE, et al. Estimating long-term effectiveness of lung cancer screening in the Mayo CT screening study. *Radiology.* 2008;248:278-287.

15. McMahon PM, Kong CY, Johnson BE, et al. Chapter 9: the MGH-HMS lung cancer policy model: tobacco control versus screening. *Risk Anal.* 2012;32(suppl 1):S117-S124.
16. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model. *Cancer Causes Control.* 2008;19:317-328.
17. Hazelton WD, Clements MS, Moolgavkar SH. Multistage carcinogenesis and lung cancer mortality in 3 cohorts. *Cancer Epidemiol Biomarkers Prev.* 2005;14:1171-1181.
18. Hazelton WD, Goodman G, Rom WN, et al. Longitudinal multistage model for lung cancer incidence, mortality, and CT detected indolent and aggressive cancers. *Math Biosci.* 2012;240:20-34.
19. Schollnberger H, Manuguerra M, Bijwaard H, et al. Analysis of epidemiological cohort data on smoking effects and lung cancer with a multi-stage cancer model. *Carcinogenesis.* 2006;27:1432-1444.
20. Tammemagi MC, Lam SC, McWilliams AM, Sin DD. Incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction. *Cancer Prev Res (Phila).* 2011;4:552-561.
21. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst.* 2007;99:715-726.
22. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Med Decis Making.* 2012;32:690-700.
23. Jeon J, Meza R, Krapcho M, Clarke LD, Byrne J, Levy DT. Chapter 5: actual and counterfactual smoking prevalence rates in the U.S. population via microsimulation. *Risk Anal.* 2012;32(suppl 1):S51-S68.
24. Rosenberg MA, Feuer EJ, Yu B, et al. Chapter 3: cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Anal.* 2012;32(suppl 1):S25-S38.
25. MacMahon H, Austin JH, Gamsu G, et al; Fleischner Society. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology.* 2005;237:395-400.
26. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic microsimulation models for health outcomes: a review. *Med Decis Making.* 2011;31:10-18.
27. Haiman CA, Stram DO, Wilkens LR, et al. Ethnic and racial differences in the smoking-related risk of lung cancer. *N Engl J Med.* 2006;354:333-342.
28. Kenfield SA, Wei EK, Stampfer MJ, Rosner BA, Colditz GA. Comparison of aspects of smoking among the four histological types of lung cancer. *Tob Control.* 2008;17:198-204.
29. Centers for Disease Control and Prevention (CDC). Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000-2004. *MMWR Morb Mortal Wkly Rep.* 2008;57:1226-1228.
30. Humphrey LL, Deffenbach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive Services Task Force recommendation. *Ann Intern Med.* 2013;159:411-420.
31. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and Harms of Computed Tomography Lung Cancer Screening Programs for High-Risk Populations. AHRQ Pub. No. 13-05196-EF-2. Rockville, MD: Agency for Healthcare Research and Quality; 2013.