Comparing Benefits from Many Possible Computed Tomography Lung Cancer Screening Programs: Extrapolating from the National Lung Screening Trial Using Comparative Modeling

Pamela M. McMahon1,2*, Rafael Meza3, Sylvia K. Plevritis4, William C. Black5, C. Martin Tammemagi6, Ayca Erdogan4, Kevin ten Haaf7, William Hazelton8, Theodore R. Holford9, Jihyoun Jeon10, Lauren Clarke11, Chung Yin Kong1,2, Sung Eun Choi1, Vidit N. Munshi1, Summer S. Han4, Joost van Rosmalen7*, Paul F. Pinsky12, Suresh Moolgavkar13, Harry J. de Koning7*, Eric J. Feuer14

1 Institute for Technology Assessment, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 2 Department of Radiology, Harvard Medical School, Boston, Massachusetts, United States of America, 3 Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, United States of America, 4 Department of Radiology, Stanford University, Stanford, California, United States of America, 5 Department of Radiology, Dartmouth Medical School, Hanover, New Hampshire, United States of America, 6 Department of Community Health Sciences, Brock University, Ontario, Canada, 7 Department of Public Health, Erasmus MC, Rotterdam, Netherlands, 8 Program of Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, 9 Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, United States of America, 10 Department of Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, 11 Cornerstone Systems Northwest, Inc., Lynden, Washington, United States of America, 12 Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland, United States of America, 13 Department of Epidemiology, School of Public Health University of Washington, Seattle, Washington, United States of America, and Department of Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, 14 Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, United States of America

Abstract

Background: The National Lung Screening Trial (NLST) demonstrated that in current and former smokers aged 55 to 74 years, with at least 30 pack-years of cigarette smoking history and who had quit smoking no more than 15 years ago, 3 annual computed tomography (CT) screens reduced lung cancer-specific mortality by 20% relative to 3 annual chest X-ray screens. We compared the benefits achievable with 576 lung cancer screening programs that varied CT screen number and frequency, ages of screening, and eligibility based on smoking.

Methods and Findings: We used five independent microsimulation models with lung cancer natural history parameters previously calibrated to the NLST to simulate life histories of the US cohort born in 1950 under all 576 programs. ‘Efficient’ (within model) programs prevented the greatest number of lung cancer deaths, compared to no screening, for a given number of CT screens. Among 120 ‘consensus efficient’ (identified as efficient across models) programs, the average starting age was 55 years, the stopping age was 80 or 85 years, the average minimum pack-years was 27, and the maximum years since quitting was 20. Among consensus efficient programs, 11% to 40% of the cohort was screened, and 153 to 846 lung cancer deaths were averted per 100,000 people. In all models, annual screening based on age and smoking eligibility in NLST was not efficient; continuing screening to age 80 or 85 years was more efficient.

Conclusions: Consensus results from five models identified a set of efficient screening programs that include annual CT lung cancer screening using criteria like NLST eligibility but extended to older ages. Guidelines for screening should also consider harms of screening and individual patient characteristics.


Editor: Juan P. de Torres, Clinica Universidad de Navarra, Spain

Received February 25, 2014; Accepted May 21, 2014; Published June 30, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This report is based on research conducted by the Cancer Intervention and Surveillance Modeling Network under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Administrative Supplement to U01 CA152956). The National Cancer Institute supported the infrastructure for the CISNET models. The Agency for Healthcare Research and Quality funded this work and provided review. The authors worked with USPSTF members to specify the overall questions. The findings and conclusions in this document are those of the authors, who are responsible for its contents, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: One of the authors (Lauren Clarke) is the owner of Cornerstone Systems Northwest, Inc., which performed services to the National Cancer Institute in support of the performance of the research reported in the manuscript. Another author (Pamela McMahon) began employment with a scientific consulting firm, Exponent, in Feb., 2014, after completion of the project reported in this manuscript. (She simultaneously holds the academic appointments as listed in the manuscript.) Neither of these affiliations interferes with, or could reasonably be perceived as interfering with, the full and objective presentation, peer review, editorial decision-making, or publication of the research. Therefore, these affiliations do not alter the authors’ adherence to all PLOS ONE policies on sharing data and materials.
Introduction

In the National Lung Screening Trial (NLST) [1], participants aged 55–74 years randomized to three annual CT examinations experienced a 20% reduction in lung cancer mortality at 6.5 years of follow up (16% at 7.5 years) [2], compared to participants randomized to receive three annual chest radiographs. The NLST was designed to determine the efficacy of CT screening, but the eligibility criteria and the number of screens offered were not meant to represent a population screening strategy. Multiple clinical guidelines, however, recommend lung cancer screening for individuals meeting the NLST eligibility criteria [3,4]. Other guidelines expanded recommendations for screening to individuals who would have been ineligible for the NLST [5–7].

The NLST provided no direct evidence of further reductions in lung cancer mortality from additional screens, or of potential benefits of screening individuals with lighter smoking histories (fewer than 30 pack-years of cigarette smoking or former smokers who had quit more than 15 years prior) or individuals younger than 55 or older than 74 years at the beginning of screening.

We extrapolated the findings of the NLST and compared various screening programs if adopted in the US population. Five modeling groups used independent approaches to combine multiple sources of data to simulate the underlying natural history of lung cancer and to estimate the benefit of alternative screening programs. In a single cohort of people born in 1950, each model estimated the benefits from 576 screening programs that varied eligibility criteria and frequency of screens, and two reference scenarios. We sought to rank programs according to a measure of efficiency, to reduce the number of programs that would require closer evaluation. The 1950 birth cohort was selected because they reach age 63 (about mid-range of participants in the NLST) in 2013. When independent models reach consensus on the characteristics of efficient screening programs, as reported here, the results can better inform screening guidelines. As in prior comparative modeling studies of important public health questions [8,9] independent modeling groups collaborated, sharing inputs and standardizing analyses to remove uncertainty due to incongruent modeled populations, endpoints and metrics.

Methods

Models

The microsimulation models used were developed independently by investigators at five institutions funded by the National Cancer Institute’s Cancer Intervention and Surveillance Modeling Network (CISNET, www.cisnet.cancer.gov) consortium through a peer-reviewed, cooperative award (2010–2015) from the National Institutes of Health: Erasmus MC in the Netherlands (Model E), Fred Hutchinson Cancer Research Center (Model F), Massachusetts General Hospital (Model M), Stanford University (Model S) and the University of Michigan (Model U). Additional investigators (see also Acknowledgments) collaborated to develop common inputs and standardize analyses. The analyses and results described in this report were part of a project to inform recommendations for lung cancer screening issued by the US Preventive Services Task Force [10].

Each of the five models simulated the underlying natural history of lung cancer, including dose-response modules that relate an individual’s detailed, dynamic cigarette smoking history to lung cancer risk (by histology and sex), and estimated (as an output) the effect of early detection with CT screening on lung cancer survival (Table 1, Part A in File S1, and Table S1 in File S1). Algorithms for following up a positive screening test (defined in our analysis as suspicious for lung cancer) were simulated with varying detail (Table 1). Prior to this analysis, all models were populated with de-identified trial participant histories and adjusted to match the trial design (e.g., numbers of screens and screening modality). All models were calibrated to reproduce multiple endpoints consistent with NLST and the Prostate, Lung, Colorectal and Ovarian (PLCO) [11] cancer screening trial [12]. Because the models simulate the natural history of disease, they can predict outcomes in years after the last year of observed follow up and in what-if scenarios with hypothetical screening programs and participants.

Common Model Inputs

Publicly available data were used for this analysis. All models simulated US men and women (all races) born in 1950. Detailed smoking histories (including non-smokers) and non-lung-cancer mortality risks were created as described below and in Part C in File S1, and Figures S1 and S2 in File S1, and used by all models as common inputs. Smoking histories and quit rates that were previously estimated through 2000 [13] were updated to calendar year 2009 for this analysis [14] and years past 2009 were projected; similarly, tables of non-lung-cancer mortality rates specific to smoking history (i.e., categories of current smokers had increased risks relative to never smokers, with former smoker mortality interpolated as a function of years since quitting) [15] were updated to 2009 and projected past 2009. The proportion of the 1950 cohort that had accumulated the specified number of pack-years by a given age is shown in Figure S4 in File S1.) In the NLST and the PLCO trial, individuals had substantially lower non-lung cancer mortality than the general population even after adjusting for their smoking status. Our use of US population other-cause mortality rates rather than the lower rates observed in the NLST or PLCO was based on an assumption that the “healthy volunteer” effect in the trials would not persist if screening for lung cancer disseminated widely.

Standardized analyses

Each model was used to simulate men and women who were born in 1950 from age 45 (calendar year 1995) to death or age 90, under 576 programs and 2 reference scenarios (a no screening scenario and a scenario with a maximum of 3 screens; Table 2). Screening programs varied according to five criteria: age to start screening (45, 50, 55, 60); age to stop screening (75, 80, 85); screen frequency (every 1, 2, or 3 years); minimum number of pack-years of cigarette exposure (10, 20, 30, 40); and (for former smokers) maximum years since quitting (10, 15, 20, 25). We refer to programs using shorthand for Periodicity (A, annual, B, biennial, or T, triennial), Start Age - Stop Age - Minimum Pack-Years - Maximum Years Since Quit. For example A55-75-30-15 represents starting screening at age 55 years and ending screening at age 75, for individuals with a minimum smoking history of 30 pack-years, and a maximum years since quitting of 15 years. This program, which we refer to as ‘NLST eligibility’ is similar to the...
<table>
<thead>
<tr>
<th>Model Features</th>
<th>Erasmus MC</th>
<th>Fred Hutchinson Cancer Research Center</th>
<th>University of Michigan</th>
<th>Massachusetts General Hospital</th>
<th>Stanford University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central dose-response model</td>
<td>Two-stage clonal expansion (TSCE) [40]</td>
<td>Longitudinal multistage observation</td>
<td>Multistage clonal expansion</td>
<td>Probabilistic [41]</td>
<td>TSCE [40] with modifications</td>
</tr>
<tr>
<td>Diagnostic follow-up algorithm</td>
<td>Implicit. Stochastic chance (separately for patients with lung cancer diagnoses versus false positives) of receipt of a set number of follow-up exams, based on the observed frequency of exams per positive exam in the NLST CT arm.</td>
<td>Implicit (see model E).</td>
<td>Implicit (see model E).</td>
<td>Explicit. Detailed algorithms based on size thresholds and risk factors. Simulated less-aggressive algorithms than the Fleischner guidelines [42] to approximate the observed frequency in the NLST, which did not specify an algorithm.</td>
<td>Explicit (see Model M).</td>
</tr>
<tr>
<td>Screening effectiveness mechanism</td>
<td>Cure model. Screen-detected cases experience a reduced risk of dying from lung cancer (compared to the stage-specific survival had the same tumor been diagnosed clinically). The improved prognosis is represented as a cure fraction (specific to stage, estimated via calibration to screening trial results). If curative treatment fails, the patient survives as long as if the tumor had been diagnosed clinically, corrected for lead-time.</td>
<td>Combination cure model and stage shift Model F assumes that screen-detected cancers were treated according to clinical practice guidelines with estimated cure rates that depend on both tumor stage and histology.</td>
<td>Stage shift model, with adjustments for age. Time to death from lung cancer detection is based on survival models that define cure by histology, stage, gender, and age at diagnosis with better outcomes associated with younger age at detection. Screening can lead to improved survival due to detection at earlier stages.</td>
<td>Cure model with possibility of recurrence. Patients with early-stage non-small cell lung cancer undergo resection (lobectomy, consistent with consensus practice guidelines) which removes the primary cancer. For patients with neither undetected distant (lethal) metastases nor undetected primary lung cancers in another lobe of the lung, resection is curative for lung cancer.</td>
<td>Cure model. The probability of lethal metastases is estimated as a function of tumor size, histology and sex. With screening, patients are more likely to be detected at early stages and before the onset of lethal metastases, and cured following standard of care; patients are not cured if detected in early stages but after the onset of lethal metastases or in advanced stages.</td>
</tr>
</tbody>
</table>

Supplementary Model Descriptions and Table S1 in File S1 provide additional details, including data used to develop and verify models. doi:10.1371/journal.pone.0099978.t001
Benefits of Lung Cancer Screening Programs

Table 2. Screening programs evaluated.

<table>
<thead>
<tr>
<th>Program Characteristic</th>
<th>Values</th>
<th># of Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of screening</td>
<td>Annual, every 2 years, every 3 years</td>
<td>3</td>
</tr>
<tr>
<td>Age to begin screening</td>
<td>45, 50, 55, 60</td>
<td>4</td>
</tr>
<tr>
<td>Age to end screening</td>
<td>75, 80, 85</td>
<td>3</td>
</tr>
<tr>
<td>Minimum PY for screening</td>
<td>10, 20, 30, 40</td>
<td>4</td>
</tr>
<tr>
<td>Maximum YSQ for screening</td>
<td>10, 15, 20, 25</td>
<td>4</td>
</tr>
<tr>
<td>Total (including 2 reference programs)</td>
<td></td>
<td>578</td>
</tr>
</tbody>
</table>

NLST design except that screening was not limited to 3 screenings (a maximum of 21 screens are possible from ages 55 to 75).

As individuals age, their accumulated pack-years or years since quitting may change. In this analysis, the models assessed eligibility annually; to be screened at a specific age within the qualifying age range, an individual also had to meet both the pack-years and the years-since-quitting criteria. Thus lighter smokers may not begin screening at the start age and former smokers may cease screening prior to the stop age.

All simulations were performed assuming idealized, perfect screening adherence for eligible individuals and smoking cessation was assumed to be unaffected by screening results.

For the biennial and triennial programs, the frequency of screening exams was changed while retaining each model’s natural history parameters, which simulate the underlying progression of disease.

Model M generated a second set of results that added operative candidacy (i.e., healthy enough for curative surgery) as an eligibility criteria for screening and reduced rates of operative candidacy in older patients (Part A in File S1) [16].

Outcome Metrics

For each program, each model generated counts of screening exams and lung cancer deaths avoided relative to no screening, separately for males and females. All events are ‘per person in the population’ rather than ‘per person screened’ because programs defining eligibility based on smoking history may screen similar proportions of the population but screen dissimilar people, even for identical starting and stopping ages. Counts of screening exams excluded follow-up and incidental CT exams. Counts of deaths avoided per screening scenario were expressed as the proportion of the (within-model) maximum possible deaths avoided from any of the screening programs evaluated.

In this analysis, we sought to formally represent the tradeoffs between maximizing the benefits (here, lung cancer deaths avoided accruing to a specific screening program while simultaneously minimizing the harms (here, the numbers of screening exams required to avoid the lung cancer deaths). One way to compare alternative programs that represent different tradeoffs is to generate an “efficiency frontier”. Each model generated efficiency frontiers for each sex that connected the screening programs that prevented the most deaths for each possible value of the number of CT screens. (Note that our definition of efficiency is not equivalent to identifying the lowest ratio of screens per death avoided. As screening intensity increases, the number of screens per death avoided will increase, but among programs with similar numbers of screens, some [the most efficient] will prevent more deaths.) For each model’s results, we generated a rank score (decile of distance [17] from the model’s frontier) for each program not on the frontier (Part B in File S1). Programs on or closest to the frontier (first three deciles) as predicted by at least 3 models were identified for males and females separately. Programs that were in both male and female lists were defined as consensus programs.

For each consensus program, we combined counts per 100,000 persons from males and females and calculated the mean predicted counts of lung cancer cases, lung cancer deaths, life years, and screening CT exams performed. We calculated the percent of the cohort receiving at least one screening exam and the number of persons ever screened per lung cancer death avoided (number needed to screen, NNS).

A secondary set of consensus programs for which the benefit (i.e., the y axis) was measured as life years saved (with the x axis remaining counts of CT screens) was also identified, using the identical steps as above.

Results

Using eligibility criteria like those in NLST, neither 3 annual screens (A62-64-30-15) nor 21 annual screens (A55-75-30-15) appears on the frontier for any model (Figure 1 and Figure S7 in File S1). There was variability among the models with respect to the effects of the smoking criteria on distance from the frontier, but consensus was clear regarding age: compared with A55-75-30-15, all models placed A55-85-30-15 closer to (or on) the frontier, indicating that continuing screening to older ages was more efficient than stopping at age 75. Conversely, initiating screening at younger ages (A45-75-30-15) was farther from the frontier (less efficient). Less-frequent (B55-75-30-15) screens provided fewer benefits, as did increasing the pack-year maximum (A55-75-40-15). The most intensive annual program (A45-85-10-25) was the upper right of the frontier for all models.

We identified 120 consensus programs. Of these, 119 had a stopping age of 80 or 85 (Figure 2, Table S2 in File S1, and Figure S8 in File S1). Across the 120 consensus programs, the average start age (54.8 y) and the average minimum pack-years (27.1) were close to the NLST criteria but the average maximum years since quit was higher (19.9 y). For all models (Figure 3), the 120 consensus programs are close to the model’s own frontier.

Results from a selected subset of 41 (every third, sorted by percent ever screened) consensus programs are provided in Table 3.
Between 11% and 40% of the cohort was screened, requiring between 43,000 to over 920,000 CT screens per 100,000 persons (Table 3). The models predicted an average of 3,719 lung cancer deaths per 100,000 in the no screening scenario (SD 820.43; Figure S6 in File S1). Per 100,000 persons, the 41 consensus programs would avoid between 153 and 846 lung cancer deaths and save between 1,883 and 9,851 years of life, relative to no screening, and the mean predicted NNS varied from 34.5 to 94.2.

Based on results from one model (M), reducing the proportions of older individuals screened (due to ineligibility for surgical resection) resulted in fewer CT screens and fewer lung cancer deaths avoided (13.3% and 14.8%, respectively, across the consensus programs), but programs that extended screening to ages 80 and 85 remained on the efficiency frontier (Figure S9 in File S1).

When the benefit of screening was measured as life years saved rather than lung cancer deaths avoided, the second set of consensus efficient programs had younger average start and stop ages (49.5 y and 80.9 y, respectively) but similar average minimum pack-years and maximum years since quit (Table S3 in File S1).

**Discussion**

Five independent models ranked 576 lung cancer screening programs by weighing one metric of their potential benefits (lung cancer deaths avoided) against one measure of harms or resource use (counts of CT screening exams) in the US cohort born in 1950. The models had been previously calibrated to multiple endpoints in NLST, but heterogeneity in the underlying model structures and assumptions yielded heterogeneous predictions for absolute numbers of lung cancer deaths avoided when extrapolating (mean and SD of results from the five models). Between 11% and 40% of the cohort was screened, requiring between 43,000 to over 920,000 CT screens per 100,000 persons (Table 3). The models predicted an average of 3,719 lung cancer deaths per 100,000 in the no screening scenario (SD 820.43; Figure S6 in File S1). Per 100,000 persons, the 41 consensus programs would avoid between 153 and 846 lung cancer deaths and save between 1,883 and 9,851 years of life, relative to no screening, and the mean predicted NNS varied from 34.5 to 94.2.

Based on results from one model (M), reducing the proportions of older individuals screened (due to ineligibility for surgical resection) resulted in fewer CT screens and fewer lung cancer deaths avoided (13.3% and 14.8%, respectively, across the consensus programs), but programs that extended screening to ages 80 and 85 remained on the efficiency frontier (Figure S9 in File S1).

When the benefit of screening was measured as life years saved rather than lung cancer deaths avoided, the second set of consensus efficient programs had younger average start and stop ages (49.5 y and 80.9 y, respectively) but similar average minimum pack-years and maximum years since quit (Table S3 in File S1).
Figure 3. Normalized plots from all models showing consensus programs. Shown are efficiency frontiers for all 5 models, with the 120 consensus programs marked. All vertical axes are normalized to within-model predictions, as in Figures 1 and 2.
doi:10.1371/journal.pone.0099978.g003
<table>
<thead>
<tr>
<th>Program</th>
<th>Start-Stop-PY-YSQ</th>
<th>Mean % cohort ever screened</th>
<th>SD</th>
<th>Mean % cohort ever screened</th>
<th>SD</th>
<th>Mean number of CT screens</th>
<th>SD</th>
<th>Mean number of life-years saved</th>
<th>SD</th>
<th>Mean number of lung cancer deaths avoided</th>
<th>SD</th>
<th>Mean NNS</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T60-75-40-10</td>
<td>11.1</td>
<td>42.8%</td>
<td>2.517</td>
<td>133</td>
<td>45.0%</td>
<td>3.123</td>
<td>233</td>
<td>2.786</td>
<td>1083</td>
<td>1201</td>
<td>7746</td>
<td>1993</td>
<td></td>
</tr>
<tr>
<td>T60-80-40-10</td>
<td>11.2</td>
<td>46.8%</td>
<td>2.57</td>
<td>135</td>
<td>48.0%</td>
<td>3.223</td>
<td>235</td>
<td>2.837</td>
<td>1086</td>
<td>1204</td>
<td>7847</td>
<td>1995</td>
<td></td>
</tr>
<tr>
<td>B60-80-40-10</td>
<td>11.3</td>
<td>49.8%</td>
<td>2.73</td>
<td>137</td>
<td>50.0%</td>
<td>3.286</td>
<td>237</td>
<td>2.951</td>
<td>1088</td>
<td>1206</td>
<td>7948</td>
<td>1997</td>
<td></td>
</tr>
<tr>
<td>T60-80-40-15</td>
<td>12.0</td>
<td>53.8%</td>
<td>2.917</td>
<td>139</td>
<td>55.0%</td>
<td>3.396</td>
<td>239</td>
<td>3.065</td>
<td>1100</td>
<td>1209</td>
<td>8049</td>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>A55-80-40-10</td>
<td>13.9</td>
<td>63.8%</td>
<td>3.13</td>
<td>141</td>
<td>65.0%</td>
<td>3.507</td>
<td>241</td>
<td>3.173</td>
<td>1112</td>
<td>1212</td>
<td>8150</td>
<td>2002</td>
<td></td>
</tr>
<tr>
<td>A55-80-40-15</td>
<td>15.9</td>
<td>68.8%</td>
<td>3.33</td>
<td>143</td>
<td>70.0%</td>
<td>3.617</td>
<td>243</td>
<td>3.281</td>
<td>1114</td>
<td>1214</td>
<td>8251</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>A55-80-40-20</td>
<td>16.9</td>
<td>73.8%</td>
<td>3.53</td>
<td>145</td>
<td>75.0%</td>
<td>3.727</td>
<td>245</td>
<td>3.389</td>
<td>1116</td>
<td>1216</td>
<td>8352</td>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>A55-80-40-25</td>
<td>17.9</td>
<td>78.8%</td>
<td>3.73</td>
<td>147</td>
<td>80.0%</td>
<td>3.837</td>
<td>247</td>
<td>3.497</td>
<td>1118</td>
<td>1218</td>
<td>8453</td>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>A55-80-40-30</td>
<td>18.9</td>
<td>83.8%</td>
<td>3.93</td>
<td>149</td>
<td>85.0%</td>
<td>3.947</td>
<td>249</td>
<td>3.605</td>
<td>1120</td>
<td>1220</td>
<td>8554</td>
<td>2010</td>
<td></td>
</tr>
<tr>
<td>A55-80-40-35</td>
<td>19.9</td>
<td>88.8%</td>
<td>4.13</td>
<td>151</td>
<td>90.0%</td>
<td>4.057</td>
<td>251</td>
<td>3.713</td>
<td>1122</td>
<td>1222</td>
<td>8655</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>A55-80-40-40</td>
<td>20.9</td>
<td>93.8%</td>
<td>4.33</td>
<td>153</td>
<td>95.0%</td>
<td>4.167</td>
<td>253</td>
<td>3.821</td>
<td>1124</td>
<td>1224</td>
<td>8756</td>
<td>2014</td>
<td></td>
</tr>
</tbody>
</table>

*Mean (SD) predicted benefits from 5 models for 41 selected (of 120) consensus programs (both sexes combined).
Table 3. Cont.

<table>
<thead>
<tr>
<th>Program characteristics:</th>
<th>% cohort ever screened</th>
<th>% cohort ever screened</th>
<th>Number of CT screens</th>
<th>Number of CT screens</th>
<th>Lung cancer deaths avoided</th>
<th>Lung cancer deaths avoided</th>
<th>NNS (mean)</th>
<th>NNS (mean)</th>
<th>Life-years saved</th>
<th>Life-years saved</th>
</tr>
</thead>
<tbody>
<tr>
<td>FreqStart-Stop-PY-YSQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A45-80-20-25</td>
<td>29.9</td>
<td>1.1</td>
<td>721,956</td>
<td>19,536</td>
<td>780</td>
<td>453</td>
<td>49</td>
<td>25</td>
<td>9206</td>
<td>4531</td>
</tr>
<tr>
<td>A55-85-10-15</td>
<td>29.9</td>
<td>2.3</td>
<td>448,193</td>
<td>26,722</td>
<td>651</td>
<td>332</td>
<td>59</td>
<td>34</td>
<td>6876</td>
<td>3909</td>
</tr>
<tr>
<td>A60-85-10-25</td>
<td>31.1</td>
<td>2.1</td>
<td>427,669</td>
<td>21,334</td>
<td>660</td>
<td>322</td>
<td>59</td>
<td>32</td>
<td>6474</td>
<td>3951</td>
</tr>
<tr>
<td>A50-80-10-15</td>
<td>34.6</td>
<td>2.3</td>
<td>583,756</td>
<td>35,681</td>
<td>700</td>
<td>388</td>
<td>63</td>
<td>34</td>
<td>8036</td>
<td>4143</td>
</tr>
<tr>
<td>A55-85-10-25</td>
<td>36.0</td>
<td>2.0</td>
<td>590,101</td>
<td>31,172</td>
<td>768</td>
<td>397</td>
<td>59</td>
<td>31</td>
<td>8109</td>
<td>4454</td>
</tr>
<tr>
<td>A50-85-10-20</td>
<td>37.5</td>
<td>2.0</td>
<td>685,484</td>
<td>39,445</td>
<td>795</td>
<td>422</td>
<td>59</td>
<td>31</td>
<td>8772</td>
<td>4509</td>
</tr>
<tr>
<td>A50-85-10-25</td>
<td>38.9</td>
<td>1.9</td>
<td>767,313</td>
<td>40,320</td>
<td>851</td>
<td>443</td>
<td>57</td>
<td>28</td>
<td>9151</td>
<td>4735</td>
</tr>
<tr>
<td>A45-80-10-25</td>
<td>40.3</td>
<td>1.9</td>
<td>920,505</td>
<td>45,739</td>
<td>846</td>
<td>479</td>
<td>60</td>
<td>29</td>
<td>9851</td>
<td>4737</td>
</tr>
<tr>
<td>Average CV</td>
<td>0.06</td>
<td>0.04</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.58</td>
<td></td>
</tr>
</tbody>
</table>

Percentage of cohort screened, numbers of CT screens, lung cancer deaths avoided, and life years saved are all normalized to cumulative counts per 100,000 people in the cohort at age 45 (including non-smokers and persons not screened), followed to age 90. See Table S2 in File S1 for complete list of 120 consensus programs identified from the 576 programs evaluated.

Frequency, A = annual, B = biennial (every 2 years), T = triennial (every 3 years); Start Age, Stop Age, PY = minimum pack-year, YSQ = maximum years since quit.

NNS, Number (people) needed to screen (ever) to prevent one lung cancer death.

1 Percent of cohort that received at least one screen; eligible individuals varied across programs.

2 **Numbers of lung cancer deaths avoided and life years saved were first calculated per model, comparing each model to its own results for lung cancer deaths in the no-screening arm. Shown are averages across models. The average (across models) number of lung cancer deaths in the no screening scenario was 3719 (SD 820).

Average Coefficient of Variation (CV) calculated as the average of (SD/mean) for each program in the table. Lower values indicate less dispersion of estimates from the models for that endpoint, across the selected consensus programs. doi:10.1371/journal.pone.0099978.t003
beyond the trial data. A key finding of our analysis was that despite differences in absolute benefits across the models, the ranking of programs was consistent; while accounting for the heterogeneity in model predictions, we were able to identify a set of consensus efficient programs. Annual screening with eligibility based on NLST criteria (beginning at age 55, continuing to age 75 for current and former smokers with a minimum of 30 pack-years and less than 15 years since quitting) was not among the programs on the efficient frontier of any of the five models. Results from all models showed that programs that extended the screening age beyond 75 prevented more lung cancer deaths for relatively few additional screens. Note that in our modeling, the stopping age for a program was the last screen for any individuals who still met the smoking cutoffs, and not the last year to be invited to begin a screening program. In the NLST which had an upper eligibility age of 74 years, individuals were as old as (77 or, rarely, 78) at the third screen. Our finding that programs that screened eligible individuals past age 75 years were efficient was unchanged when more older patients were ineligible for screening due to comorbidities that categorized them as non-operative candidates (based on results from one model) or when life years saved was substituted for the measure of benefit. While in other cancers (e.g., breast and colorectal) screening is not generally recommended beyond age 75 and not generally recommended every year, in lung cancer annual screening to older ages can be beneficial because: (1) the age-specific incidence curve for lung cancer is quite steep, and (2) the high lethality of the disease makes early detection worthwhile, even among individuals with a somewhat modest life expectancy. It is also important to note that had we defined life years saved (instead of lung cancer deaths avoided) as the measure of benefit, one could logically predict that strategies with younger stopping ages would be more likely to emerge as ‘consensus efficient’.

Our predicted NNS for A55-80-30-15 varied across models, ranging from 19.8 (Model F) to 100.5 (Model M), but all were below published estimates of NNS for only 3 screens of (256) [18] and closer to published NNS for mammography (95) or FOBT (roughly 130) for healthy 50 year-olds [19].

For consensus programs with screening until age 80, between 11% (for the least frequent programs with strictest eligibility, e.g., T60-75-40-10) and 40% (for the annual programs with more inclusive eligibility, e.g., A45-80-10-25) of the cohort born in 1950 would be screened at least once after age 45. Although not directly comparable to earlier estimates that 6% (8.7 million people) of US adults over 40 would meet the NLST eligibility cutoffs for lung cancer screening each year [20,21], our estimate of 11% of individuals seems reasonable.

We identified a set of consensus efficient programs rather than a single optimal strategy, because the efficiency frontiers did not identify a consensus inflexion point at which additional screens provided diminishing benefits. The least intensive programs at the lower left of the frontiers (Figure 2) may be less attractive, however, since annual screening consistently prevented more lung cancer deaths than did triennial or biennial programs. The most-intensive screening programs, on the other hand, will lead to more accumulated harms (radiation exposure from additional imaging examinations, overdagnosis, invasive biopsies) and costs.

Screening programs cannot be evaluated in isolation from the follow-up algorithm. In the NLST, an average of 24% of individuals in a given round of screening (CT arm) had results requiring some follow-up, but the trial did not specify a follow-up regimen, leaving open the question of the optimal regimen for individuals with positive screens, most of whom are healthy [4,22]. In models (E, F, U) that used implicit follow-up algorithms based on the experience of participants in the NLST, extrapolating the rate of follow-up to less frequent screening programs was dependent on the assumption that the rates of follow up exams and early detection of lung cancers (defined in the NLST and models E, F, and U as ‘screen-detected’ even if first seen on a follow-up exam) would not change. In the models (M, S) that explicitly modeled follow-up programs based on size, follow-up exams could change the timing of detection of a lung cancer, but the assumptions used here for frequency of follow-up imaging may not be representative of eventual practice patterns.

Several limitations of our analysis are important to note. The models do not simulate non-lung cancer incidental findings (e.g., coronary artery calcification, AAA, or other malignancies), so our results do not include potential benefits (or harms) due to their detection and treatment. There are few data to predict adherence patterns for lung cancer screening [20,23], and many possibilities to model. We conducted an idealized analysis with the goal of informing guidelines and did not consider that individuals will self-select for participation in screening based on their comorbidities, specific smoking history, or family history, as observed in screening trials [24,25]. It will be important to monitor how lung cancer screening is implemented in community settings (including recruitment, participation, positive screen evaluations, diagnosis, referral for treatment), and modeling can suggest the most important leverage points to optimize the process. Definitive evidence on the relationship between smoking cessation and NLST screening results was not available in time for our analyses. Based on limited data with non-standardized definitions of ‘quit’ [26–29] and the PLCO Trial, which found no correlation between CXR screening result and smoking behavior [30], we assumed screening did not affect background smoking patterns.

Efficient screening programs might differ in populations with different smoking patterns or other-cause mortality risks than the cohort we simulated. To simplify the comparison of hundreds of programs, we performed our analyses in a single birth cohort and did not estimate total lung cancer deaths avoided in the US [31]. Our requirement that individuals meet all eligibility criteria (including years since quitting) was transparent and is a step towards risk-based screening criteria (our models account for decreasing risks of death from lung cancer and other causes after quitting), but may not reflect guidelines, which typically define eligibility to begin screening. Future analyses to examine programs that define eligibility based on risk models will require that the models and population input files include additional characteristics (e.g., BMI, education) that go beyond age and smoking exposure [32–36]. We did not incorporate increases in operative mortality rates by age, or special clinical considerations individual to a particular patient.

Although the rankings of programs were consistent across models, uncertainty in absolute numbers of lung cancer deaths avoided (and life years saved) remained, due to variation in the underlying assumptions regarding unobserved disease processes [37]. Underlying the differences across models in predicted absolute benefits is a variation in the predicted future number of lung cancer cases in the absence of screening (Figure S3 in File S1). Essentially, our consortium of 5 models served as a sensitivity analysis on model structure and demonstrated that even when model heterogeneity was specifically taken into account, the models identified similar efficient programs (i.e., the consensus set). Our results highlight tradeoffs between preventing greater numbers of lung cancer deaths and the additional screening exams required. Guidelines for screening also consider tradeoffs in gains in life expectancy and important harms, including invasive biopsies for benign disease, overdagnosis, and lung cancers.
related to radiation from diagnostic imaging examinations [10]. Difficulties with estimating population effects of screening include the potential for concurrent smoking cessation programs to augment the benefits from screening, and the heterogeneity of the radiation dose attributable to a given CT exam, which could vary as much as 10-fold depending on the size of the patient, the generation of scanner, and the protocol in use at the clinical setting [38]. All smokers, whether undergoing screening or not, should receive cessation assistance and be encouraged to quit [39].

Supporting Information

File S1 Supporting figures and tables. Figure S1, Prevalence of smoking by age in 1950 birth cohort. Summary of shared input data (used by all 5 models) on smoking patterns for the US cohort born in 1950. Prevalence shown is estimated in the absence of lung cancer mortality. Version 1.0 of the Smoking History Generator (SHG) refers to published data through 2000 (Anderson, et al.), and version 1.5 supplies the 1950 birth cohort used for this analysis with data through 2009 and projections past 2009. Figure S2, Other-cause mortality, by smoking quintile, in 1950 birth cohort. These curves show the other-cause (non-lung cancer) mortality for never smokers and for current smokers by smoking quintile (Q, of cigarettes per day) for the male birth cohort of 1950, out to age 99. Former smokers are intermediate to current and never smokers. There is a similar plot for females. These were shared inputs used by all the models. Note that the rates of non-lung cancer mortality represent the US population, not trial (NLST or PLCO) participants. Figure S3, Prevalence of smoking by age in 1950 birth cohort. Output from one model showing smoking prevalence by age (calendar year), in a no screening scenario. Proportions of current/former/never smokers are in the presence of lung cancer mortality as well as all-cause mortality. Figure S4, Prevalence of smoking by age and pack-years in 1950 birth cohort. Output from one model showing smoking prevalence by category of pack-year and age. The proportion of the cohort by age that has accumulated the specified number of pack-years in the presence of lung cancer mortality and other-cause mortality. Figure S5, Incidence, no screening scenario, output from all models. For predictions past observed SEER data (over age 60) there are no observed data, but we used an age-period-cohort model to project past observed years (‘Projected’ red double line in plots below), which shows that the models are most divergent after age 85, when SEER data become most sparse. We cannot strictly compare incidence to that in prior birth cohorts since smoking patterns are dissimilar, and incidence varies by cohort. Figure S6, Mortality, no screening scenario, output from all models. The vertical line at age 90 indicates age at which all event counts (deaths, deaths and deaths averted, and life years gained) were truncated for the analyses reported here. Although the models ranked programs similarly, there was variability in the total numbers of predicted lung cancer cases, deaths, and therefore lung cancer deaths prevented. The differences in rates in the no screening scenario in large part explains the predicted differences between models. The four models (E, F, S, and U) which use two-stage or multi-stage clonal expansion models have more similarly shaped curves than the fifth model (M), which does not use a clonal expansion component (see Table S1 in File S1). Figure S7, Results from all models analogous to Figure 1 in article. Figure S8, Results from all models analogous to Figure 2 in article. Figure S9, Secondary results with reduced operative candidacy with age. The dashed line denotes the efficiency frontier in the main analysis. Table S1, Additional Detail on Models. Table S2, Complete List of 120 Consensus Efficient Scenarios. Table S3, Comparison of Consensus Efficient Scenarios Identified Using Life-years Saved or Lung Cancer Deaths Avoided as Measure of Benefit. (DOCX)

Acknowledgments

In addition to the authors, a team of investigators from multiple institutions (NLST and PLCO investigators, MGH, Stanford, Yale, Cornerstone Systems NW, and IMS) contributed to the analysis.

Author Contributions


References


