Overdiagnosis in Low-Dose Computed Tomography Screening for Lung Cancer

Edward F. Patz Jr, MD; Paul Pinsky, PhD; Constantine Gatsonis, PhD; JoRean D. Sicks, MS; Barnett S. Kramer, MD, MPH; Martin C. Tammemägi, PhD; Caroline Chiles, MD; William C. Black, MD; Denise R. Aberle, MD; for the NLST Overdiagnosis Manuscript Writing Team

**IMPORTANCE** Screening for lung cancer has the potential to reduce mortality, but in addition to detecting aggressive tumors, screening will also detect indolent tumors that otherwise may not cause clinical symptoms. These overdiagnosis cases represent an important potential harm of screening because they incur additional cost, anxiety, and morbidity associated with cancer treatment.

**OBJECTIVE** To estimate overdiagnosis in the National Lung Screening Trial (NLST).

**DESIGN, SETTING, AND PARTICIPANTS** We used data from the NLST, a randomized trial comparing screening using low-dose computed tomography (LDCT) vs chest radiography (CXR) among 53,452 persons at high risk for lung cancer observed for 6.4 years, to estimate the excess number of lung cancers in the LDCT arm of the NLST compared with the CXR arm.

**MAIN OUTCOMES AND MEASURES** We calculated 2 measures of overdiagnosis: the probability that a lung cancer detected by screening with LDCT is an overdiagnosis (P_o), defined as the excess lung cancers detected by LDCT divided by all lung cancers detected by screening in the LDCT arm; and the number of cases that were considered overdiagnosis relative to the number of persons needed to screen to prevent 1 death from lung cancer.

**RESULTS** During follow-up, 1089 lung cancers were reported in the LDCT arm and 969 in the CXR arm of the NLST. The probability is 18.5% (95% CI, 5.4%-30.6%) that any lung cancer detected by screening with LDCT was an overdiagnosis, 22.5% (95% CI, 9.7%-34.3%) that a non–small cell lung cancer detected by LDCT was an overdiagnosis, and 78.9% (95% CI, 62.2%-93.5%) that a bronchioalveolar lung cancer detected by LDCT was an overdiagnosis.

The number of cases of overdiagnosis found among the 320 participants who would need to be screened in the NLST to prevent 1 death from lung cancer was 1.38.

**CONCLUSIONS AND RELEVANCE** More than 18% of all lung cancers detected by LDCT in the NLST seem to be indolent, and overdiagnosis should be considered when describing the risks of LDCT screening for lung cancer.
Screening for lung cancer has been proposed for decades. It is fundamentally based on the principle that tumors will be detected at a smaller size and earlier stage, when treatment is more effective, resulting in a reduction in lung cancer mortality.1,2 An ideal screening program targets individuals at the highest risk of lung cancer, uses a cost-effective test to detect tumors at an early stage, and efficiently excludes patients with clinically insignificant abnormalities.

Unfortunately, there is currently no ideal screening test for lung cancer, and a clear understanding of the risks and benefits should be considered in the design of a population-based screening program. Low-dose computed tomography (LDCT) has been suggested as a screening tool for lung cancer, and recent results from the National Lung Screening Trial (NLST) demonstrated an encouraging 20% relative reduction in lung cancer-specific mortality compared with screening using chest radiography (CXR).3 Whereas the decrease in mortality highlights the primary benefit of screening, the trial also found more cases of lung cancer in the LDCT group compared with the CXR group; this is a limitation of screening because some of these tumors may be indolent and clinically insignificant.

In previous CXR lung cancer screening trials, more lung cancers were detected in the screened arm than in the observational group.4–6 The excess number of early-stage lung cancers, even after extended follow-up, is usually attributed to overdiagnosis.7,8 Overdiagnosis is defined as the detection, usually by screening, of a cancer that would not otherwise have become clinically apparent; overdiagnosis is often an intrinsic feature of screening, which by definition seeks to detect occult disease in asymptomatic individuals. Overdiagnosis is one of the limitations of screening because it incurs unnecessary treatment, morbidity (and mortality in rare cases), follow-up, cost, and anxiety and labels a patient with a disease that otherwise would never have been detected.9

Estimating the true level of overdiagnosis in a screening trial such as the NLST requires a sufficiently long period of post-screening follow-up because there is typically a “catch-up” period in the nonscreened arm, during which more cases of cancer are diagnosed and potentially catch up to the greater number of cases detected earlier in the screened arm. The length of the potential catch-up period is related to the lead time associated with the screening modality, where the lead time is the difference between the time when diagnosis would have been made without screening and the time that the diagnosis was actually made as a result of early detection by screening.

In the NLST, there were 4 to 5 years of follow-up after screening, which may not be sufficient to detect all cancers in the control CXR arm; with longer follow-up, additional catch-up may occur. Therefore, the present study calculated rates of excess cancers in the LDCT vs CXR arm, for all lung cancer and for various histologic subtypes. These excess cancer rates provide an upper bound on the true overdiagnosis rate associated with LDCT screening relative to CXR screening.

To complement the descriptive analysis, we also developed a standard convolution-type model that when fit to the NLST data can be used to estimate overdiagnosis and excess cancer rates over various screening scenarios different from that used in the NLST.

Methods

Design and Overview of the NLST

A detailed description of the NLST design, methods, and initial results has been previously reported; the present study used extended follow-up data through December 31, 2009.3 From August 2002 through April 2004, 53,452 individuals at high risk for lung cancer, with at least a 30-pack-year history of cigarette smoking (former smokers who had quit within the past 15 years), between the ages of 55 and 74 years were enrolled at 33 US medical centers into a prospective screening trial. The study protocol was approved by the institutional review board at each of the 33 screening centers, and written informed consent was obtained from each participant before randomization. All participants were randomly assigned to receive either 3 annual LDCT studies (26,722 participants) or 3 annual single-view posterior-anterior CXRs (26,730 participants) and then observed for up to an additional 5 years. The primary trial objective was to determine the effect of LDCT screening vs CXR screening on lung cancer mortality.

Lung cancer diagnoses were ascertained primarily through standardized forms administered to study participants at 6-month or 1-year intervals, which inquired about any recent cancer diagnoses. Trained abstractors confirmed reported cases of lung cancer using medical records and pathology reports. A screen-detected cancer was defined as a cancer diagnosed within 1 year of a positive screening result or diagnosed after a longer period on the basis of diagnostic procedures prompted by the screen.

Excess Cancer Cases

Here we use 2 definitions of an excess cancer rate, 1 emphasizing the clinical perspective (denoted \(P_s\)) and 1 emphasizing a more public health perspective (denoted \(P_a\)). For each, the numerator of the rate is the same, namely, the number of excess lung cancer cases in the LDCT arm as compared with the CXR arm, ie, the difference in the total count of lung cancers between the LDCT and CXR arms. For \(P_s\), the denominator is the total number of screen-detected lung cancer cases in the LDCT arm. The quantity \(P_s\) is a measure of the probability that a participant’s LDCT screen-detected cancer would not have become clinically apparent during the given screening phase if LDCT screening had not been performed. For \(P_a\), the denominator is the total number of lung cancers diagnosed in the LDCT arm. Thus, \(P_a\) is the fraction of all lung cancer cases diagnosed during a given period in a cohort who underwent LDCT screening that would not have been diagnosed during that period absent the LDCT screening. We estimated these 2 indices using the observed lung cancer counts from the NLST; confidence intervals were obtained using bootstrapping. We also estimated the number of cases that were considered overdiagnosed, relative to the number of participants needed to screen to prevent 1 death from lung cancer.

Values of \(P_s\) and \(P_a\) were estimated for all lung cancer, all non–small cell lung cancer (NSCLC), bronchioalveolar carcinoma (BAC), and NSCLC excluding BAC. As described in the Introduction, these excess cancer rate estimates represent an
upper bound to true overdiagnosis rates because the post-screening follow-up period in the NLST may not have been long enough to totally differentiate overdiagnosis from the effects of lead time.

We estimated the number of overdiagnoses relative to the number needed to screen to prevent 1 lung cancer death as the number of excess lung cancer cases in the LDCT arm of the NLST divided by the difference in lung cancer deaths in the LDCT and CXR arms of the NLST.

**Convolution Model Based on NLST Data**

In addition to the aforementioned descriptive analysis of excess cancers from the NLST, we also fit a standard convolution model to the NLST data to estimate excess cancers relative to no screening and excess cancers expected if follow-up continued lifelong. The convolution model postulates a pre-clinical phase of disease and a mean sojourn time in that phase before clinical diagnosis; the model also estimates the sensitivity of screening (here with LDCT or CXR).10,11 See eAppendix (in Supplement) for more details.

**Results**

**Empirical Estimates of Overdiagnosis**

The total numbers of lung cancer cases by year and by study arm are shown in **Table 1**. The mean follow-up in the LDCT arm was 6.41 years, and the mean follow-up in the CXR arm was 6.37 years. **Table 2** shows the number of screen-detected and non–screen-detected cases of lung cancer in each arm according to general histological categories.

At the end of the entire trial, there were 1089 total lung cancer cases in the LDCT arm (649 detected by LDCT screening)
and 969 cases in the CXR arm, for an excess of 120 cases. This gives excess cancer rates \( P_2 \) and \( P_4 \) of 18.5% (95% CI, 5.4%-30.6%) and 11.0% (95% CI, 3.2%-18.2%), respectively (Table 3).

With respect to excess cancer rates for NSCLC, there were a total of 926 cases in the LDCT arm and 793 cases in the CXR arm, for a difference of 133 cases. Excess cancer rates were \( P_2 = 22.5\% \) (95% CI, 9.7%-34.3%) and \( P_4 = 14.4\% \) (95% CI, 6.1%-21.8%). There were 111 cases of BAC in the LDCT arm and 36 in the CXR arm, giving excess cancer rates of \( P_2 = 78.9\% \) (95% CI, 62.2%-93.5%) and \( P_4 = 67.6\% \) (95% CI, 53.5%-78.5%).

From the original NLST report, the number needed to screen to prevent 1 lung cancer death was 320. There were 443 and 356 lung cancer deaths in the CXR and LDCT arms, respectively, giving a difference of 87. As mentioned, there was an excess of 120 lung cancer cases in the LDCT compared with the CXR arm. Therefore, the number of cases of overdiagnosis found in the 320 participants needed to screen to prevent 1 death from lung cancer is 120/87 = 1.38.

Model-Based Estimates

In general, the model fits the NLST data relatively well. The fit of the model, demonstrated by observed vs expected (modeled) case counts by mode of detection, time, and arm (LDCT vs CXR), is provided in eTable 1 (in Supplement).

The parameter estimates for NSCLC and BAC using the convolution model are shown in Table 4. The mean sojourn time for non-BAC NSCLC was 3.6 (95% CI, 3.0-4.3) years. This implies that 24.1% (42.4%) of cases would become clinically apparent within 1 year (2 years) of entering the preclinical phase (assuming no other-cause mortality). Estimated sensitivity was 83% (95% CI, 72%-94%) for LDCT and 33% (95% CI, 26%-42%) for CXR. For BAC, estimated sensitivity was 38% (95% CI, 7%-62%) for LDCT and 4% (95% CI, 1%-9%) for CXR. Mean sojourn time for BAC was 32.1 (95% CI, 17.3-270.7) years, with 14.4% (26.8%) of cases becoming clinically apparent in 5 (10) years.

Estimates of excess cancer rates \( P_2 \) with lifetime follow-up under various screening scenarios are shown in Table 5. For the NLST scenario of 3 screens and roughly 7 years of total follow-up, the excess cancer rates for all NSCLCs were 31% vs no screening and 19% vs CXR. For BAC under this scenario, rates were 85% (vs no screening) and 71% (vs CXR), whereas for non-BAC NSCLCs, rates were 21% (vs no screening) and 9% (vs CXR). For NLST screening (ie, 3 annual screens) but with lifetime follow-up, excess cancer rates decreased substantially, to 11% (vs no screening) and 9% (vs CXR) for all NSCLCs and 2.6% (vs no screening) and 1.2% (vs CXR) for non-BAC NSCLCs; rates for BAC were 49% and 41%, respectively. These latter rates, with lifetime follow-up, are thus estimates of actual overdiagnosis rates. Under a scenario of 5 annual screens and 5 years of total follow-up, excess cancer rates for non-BAC NSCLC were relatively high, 45% compared with no screening and 16% compared with CXR. However, with follow-up extended through participants’ lifetimes, excess cancer rates were similar to those for 3 annual screens.

On the basis of the estimate for non-BAC NSCLC of a mean sojourn time (and mean lead time) of 3.6 years, approximately 25% of non-BAC NSCLC tumors would have lead times of longer than 5 years. This helps explain why the non-BAC NSCLC excess cancer rate decreased substantially when the follow-up was changed from 7 years total (5 years past the last screen) to lifetime follow-up.

Discussion

Screening for lung cancer with LDCT in the NLST showed a 20% relative reduction in mortality, and 320 participants were needed to screen to prevent 1 lung cancer death. These findings were met with enthusiasm, but before a widespread public health screening program is implemented, risks of screening also need to be considered. One of the limitations and potential harms is overdiagnosis because it is not clear that all early-stage lesions detected in asymptomatic individuals will progress to cause symptoms and affect long-term outcome. These patients may undergo an invasive diagnostic procedure, have surgical resection, be given a diagnosis of lung cancer, and require multiple sequential follow-up studies when some tumors are potentially clinically insignificant. These cases of overdiagnosis are treated as any other lung cancer because it is generally not possible to distinguish indolent lesions from more aggressive tumors.

The true extent of overdiagnosis in lung cancer is difficult to determine because most of what we know about this...
Overdiagnosis in Screening for Lung Cancer

In the new International Association for the Study of Lung Cancer histologic classification of adenocarcinomas, many of these tumors would be designated as minimally invasive adenocarcinomas, suggesting an indolent behavior and good long-term outcome. These data raise the question as to the necessity and type of therapy required if a diagnosis of minimally invasive adenocarcinoma is established and challenge the diagnostic community to develop a classification scheme that could accurately phenotype all lung tumors.

In addition to the NLST data, the present study used a convolution model to explore the effect of other screening scenarios on overdiagnosis. The modeling generally provides a good fit to the NLST data and is useful for several reasons. First, it makes it possible to estimate excess cancers relative to no screening (and not just to CXR screening). Second, with the model, one can extrapolate to different screening and follow-up scenarios, not just what was observed in the NLST. Third, by extrapolating beyond the NLST follow-up period, the model can provide an estimate of actual overdiagnosis rates and not just an upper bound. Finally, the model may also be used to determine how overdiagnosis varies by lung cancer risk.

Using the model, we found that the excess cancer rate associated with 3 annual LDCT screens decreased substantially in changing from the NLST follow-up (approximately 5 years after screening) to lifetime follow-up, with the latter estimates representing true overdiagnosis. There was a relatively low rate of true overdiagnosis for non-BAC NSCLC of approximately 3% (relative to no screening), although BAC still had high rates of overdiagnosis (approximately 50%).

The model also found that overdiagnosis rates with LDCT were greater relative to no screening than relative to screening with CXR. These data are consistent with prior screening trials using CXR and sputum cytologic analysis as compared with no screening in the standard of care group because the CXR group had an excess number of cases of lung cancer.

As with any model, one should be cautious in extrapolating much beyond the data on which the model was based, which in this case are 3 annual screens and a total of approximately 7 years of follow-up. Therefore, the estimates of true overdiagnosis, based on the lifetime follow-up scenarios, must be treated cautiously. Furthermore, the model does not always reflect true clinical practice because patients are not as

Table 5. Overdiagnosis Rates ($P_d$) Associated With Low-Dose Computed Tomography (LDCT) Screening Under Various Scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>NSCLC Excluding BAC</th>
<th>BAC</th>
<th>All NSCLC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_d$ (%) (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>vs No Screening vs CXR</td>
<td>vs No Screening vs CXR</td>
<td>vs No Screening vs CXR</td>
</tr>
<tr>
<td>3 Annual screens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With 7 y follow-up</td>
<td>21 (16-25) 9 (6-12)</td>
<td>85 (69-93) 71 (52-83)</td>
<td>31 (27-34) 19 (16-23)</td>
</tr>
<tr>
<td>With lifetime follow-up</td>
<td>2.6 (2.0-3.3) 1.2 (0.7-1.7)</td>
<td>49 (34-71) 41 (28-62)</td>
<td>11 (7-15) 9 (5-13)</td>
</tr>
<tr>
<td>5 Annual screens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With 5 y follow-up</td>
<td>45 (39-49) 16 (12-21)</td>
<td>91 (84-96) 73 (56-85)</td>
<td>53 (48-56) 25 (21-29)</td>
</tr>
<tr>
<td>With lifetime follow-up</td>
<td>3.2 (2.6-3.8) 1.3 (0.8-1.9)</td>
<td>51 (33-71) 42 (27-60)</td>
<td>12 (7-15) 9 (5-13)</td>
</tr>
</tbody>
</table>

Abbreviations: BAC, bronchioloalveolar cell carcinoma; CXR, chest radiography; LDCT, low-dose computed tomography; NSCLC, non–small cell lung cancer.

* Overdiagnosis rate is total cancers in LDCT cohort minus total cancers in CXR or no screening cohort divided by total number of screen-detected cancers in LDCT cohort.
predictable and reliable as the model system would suggest, particularly if different risk categories are explored. However, the model does provide a framework to explore a variety of screening parameters so that potential limitations of various screening conditions can be understood.

In summary, screening for lung cancer with LDCT has the potential to detect indolent tumors, resulting in overdiagnosis. Whereas the NLST demonstrated a relative mortality reduction with LDCT, the limitations of the screening process, including the magnitude of overdiagnosis, should be considered when guidelines for mass screening programs are constructed. In the future, once there are better biomarkers and imaging techniques to predict which individuals with a diagnosis of lung cancer will have more or less aggressive disease, treatment options can be optimized, and a mass screening program can become more valuable.

Role of the Sponsor: The National Institutes of Health had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: The members of the NLST Overdiagnosis Writing Team are Deni Aberle, MD, University of California, Los Angeles; Judith Amorosa, MD, Robert Wood Johnson University Hospital, East Brunswick, New Jersey; Christine Berg, MD, National Cancer Institute, Bethesda, Maryland; William C. Black, MD, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire; Caroline Chiles, MD, Wake Forest University Health Sciences Center, Winston-Salem, North Carolina (Chiles); Department of Radiology, Dartmouth Medical School, Hanover, New Hampshire (Black); Department of Community and Family Medicine, Dartmouth, Tammmägli, Hanover, New Hampshire (Black); Department of Radiology, University of California, Los Angeles (Aberle).

Author Contributions: Dr Patz had full access to all of the data in the study and takes full responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Patz, Gatsenis, Sicks, Kramer, Black, Aberle. Acquisition of data: Gatsenis, Sicks, Tammmägli, Chiles. Analysis and interpretation of data: Patz, Pinsky, Gatsenis, Sicks, Kramer, Tammmägli, Black. Drafting of the manuscript: Patz, Pinsky, Gatsenis, Black. Critical revision of the manuscript for important intellectual content: Patz, Pinsky, Gatsenis, Sicks, Kramer, Tammmägli, Black. Statistical analysis: Patz, Pinsky, Gatsenis, Black. Obtained funding: Gatsenis. Administrative, technical, or material support: Gatsenis, Sicks. Study supervision: Patz, Gatsenis, Kramer, Tammmägli.

Conflict of Interest Disclosures: Dr Patz conducted a research project on serum protein biomarkers and indeterminate pulmonary nodules funded by Laboratory Corporation of America through a sponsored research agreement with Duke University. Dr Gatsenis is a consultant/member of the scientific advisory board for Wilox AG and a board member for Frontier Science & Technology Research Foundation and has served as a consultant to Endocyte Inc and Genentech.

Funding/Support: This research was supported by the National Institutes of Health (grants U01 CA079778 and U01 CA080098 and contracts NO1-CN-25511, NO1-CN-25512, NO1-CN-25513, NO1-CN-25514, NO1-CN-25515, NO1-CN-25516, NO1-CN-25518, NO1-CN-25522, NO1-CN-25524, NO1-CN-75022, NO1-CN-25476, and NO2-CN-63300).

REFERENCES