Inconsistencies in Findings From the Early Lung Cancer Action Project Studies of Lung Cancer Screening

Peter B. Bach

Manuscript received December 21, 2010; revised May 6, 2011; accepted May 10, 2011.

Correspondence to: Peter B. Bach, MD, MAPP, Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 307 E 63rd St, 2nd Floor, New York, NY 10021.

Long-standing guidelines against screening high-risk individuals for lung cancer may change following the publication of the randomized National Lung Screening Trial (NLST), which shows a benefit of computed tomography compared with chest x-ray screening. Guideline panels will likely also seek additional information from nonrandomized studies of computed tomography screening, such as the Early Lung Cancer Action Project (ELCAP). However, for the ELCAP findings to be incorporated into new guidelines, some inconsistencies in the published data should first be resolved. Specifically, some of the reports from ELCAP appear to contradict others in terms of important endpoints, and several findings from ELCAP appear to be statistically improbable or outliers when compared with analyses and studies by other research groups. Clarification of both internal and external inconsistencies is a prerequisite for evaluation of the body of work published by ELCAP investigators.

J Natl Cancer Inst 2011;103:1002–1006

The National Lung Screening Trial (NLST) recently reported that computed tomography (CT) lung cancer screening, when applied to high-risk individuals (aged 55–74 years and with at least 30 pack-years of smoking and if quit, did so within 15 years), is associated with a 20% reduction in lung cancer mortality compared with chest x-ray screening (http://www.cancer.gov/images/DSMB-NLST.pdf). Bodies that publish cancer screening recommendations will soon be contemplating this landmark finding and will be comparing it with several other CT screening studies. Proper consideration of the findings published by the Early Lung Cancer Action Project (ELCAP), which also publishes under the names New York-ELCAP (NY-ELCAP) and International-ELCAP (I-ELCAP), will require reconciling what in some cases appear to be contradictory, statistically unlikely, or outlier results that are substantially different from those generated by other investigators examining the same questions.

Apparently Contradictory Results

Mismatch in the Number of Lung Cancer Deaths Across Analyses of ELCAP Data

A key goal of lung cancer screening is to reduce the mortality rate from lung cancer, which makes the ascertainment of deaths from lung cancer among screened individuals central to the evaluation of a screening approach. In their influential study published in the New England Journal of Medicine in 2006 (1), ELCAP reported that among the 31,567 subjects who were followed from their enrollment to their death or to an end date of May 31, 2006, there were 75 lung cancer deaths (p. 1767). However, in other reports (2–4), two small subgroups of this study were reported to have more lung cancer deaths by an earlier follow-up date. Specifically, ELCAP researchers reported (2,3) that among the 7994 subjects in the New York State subsample, 64 lung cancer deaths were observed by December 31, 2005 [5 months before the end date of the study above (1)]. Although I was not part of the ELCAP research group, I was part of a research team that analyzed ELCAP data for subjects who were enrolled at the Moffitt Cancer Center in Florida. Our analysis (4) found that 13 lung cancer deaths occurred among the 1151 participants in that study by December 31, 2004 [17 months before the end date of the study cited above (1)]. Together, these two subsets of the I-ELCAP cohort (2–4) report two more deaths (n = 77) from lung cancer than the total number of lung cancer deaths (n = 75) reported in the larger study (1), which included an additional 22,422 non-New York and non-Moffitt participants. These two numbers cannot both be correct, because it is impossible that more lung cancer deaths occurred in a subset of a cohort by an earlier follow-up date than occurred in the larger cohort with a later follow-up date.

One might speculate that there were some subtle shifts in assignment of cause of death between the two analyses, but such an event is unlikely to explain the discrepancy. The total number of screened subjects in ELCAP (n = 31,567) exceeds by 22,422, the number of subjects enrolled in the New York State and Moffitt Florida subsets combined (n = 9145). If one assumes that the same proportion of lung cancer deaths occurred in the non-New York and non-Moffitt portion as occurred in the New York and Moffitt parts of the study, then the number of lung cancer deaths reported in (1) would have been approximately 264 instead of the 75 that were reported.

The deaths reported in the New York and Moffitt subsets may be more accurate than the deaths reported in the entire study population (1). In those two subsets, the vital status of all study...
subjects was evaluated through national death registries with information on cause of death gleaned from death certificates, which is the reference standard for how such a follow-up should be conducted. In the 2006 article (1), there is no mention of the use of public death records to assess vital status.

Mismatch in the Entry Times and Follow-up of Subjects With Untreated Stage I Lung Cancer

ELCAP survival statistics also appear to be inconsistent between their 2006 (1) and 2008 (5) publications. I explained this discrepancy in a letter published in The Oncologist (6). Specifically, in 2006 (1), ELCAP researchers reported that they had complete follow-up, from diagnosis until death, for eight subjects in their study who were diagnosed with early stage lung cancer but received no treatment. They reported that all eight subjects were observed to die within 5 years, meaning that at the time of their reporting they had no such untreated patients who were still alive and in active follow-up (1). In 2008 (5), ELCAP researchers reported that they had 13 such untreated patients, meaning five additional subjects, all followed from diagnosis until their death. In this later report (5), ELCAP researchers similarly noted that all such subjects had been observed to die within 5 years, meaning that none of these subjects was in active follow-up and thus had been censored. The authors included a Kaplan–Meier curve for these 13 subjects, which provided the timing of the deaths relative to the time of their diagnosis (5).

As I noted in my letter in The Oncologist (6), it is not possible for both reports to be correct because it was not possible for all eight untreated patients to have been followed to an outcome of death and then approximately a year later for five more patients to be both diagnosed and observed to die of lung cancer in the same study. Even if one assumes that the additional five individuals had the shortest five survival times of the 13 subjects in the censored group, and that the maximum plausible amount of time had elapsed between the end date of the previous study and the publication date of the latter study, not enough months had elapsed to make this possible. The fifth shortest time to death occurred at 20 months after diagnosis but only 18 months had passed between the last follow-up time of the first study (1) and the final acceptance of the second article for publication (5).

Statistically Unlikely Results

Unlikely Censoring Pattern

In the same letter (6) in which I noted the concern about the number of new subjects being diagnosed with lung cancer and being followed to their death from lung cancer in a period briefer than was possible, I also noted that it was statistically unlikely that all of the untreated subjects would be followed to an outcome of death, instead of at least a few being censored at the point of last study follow-up. I based my argument on the censoring distribution for the cohort as a whole and pointed out that censoring in studies such as ELCAP is mostly due to rolling entry times, and thus it cannot be avoided through more diligent follow-up. I provided a calculation of the probability that there would be no censoring for the untreated group based on the reported event timing and the censoring distribution for the study as a whole. I estimated that one or more of the subjects would be censored 98.4% of the time under these conditions, whereas no censoring of untreated subjects, as was reported by the ELCAP researchers, would occur only 1.6% of the time (6).

Consistency of the Estimate of the Proportion of Lung Cancers That Are “Early” Stage Among Three Analyses

The first major finding published by ELCAP appeared in The Lancet in 1999 (7). In that study, which involved 1000 patients and 27 case patients with lung cancer, ELCAP researchers reported that the proportion of lung cancers diagnosed in the cohort that were stage I, rather than a more advanced stage, was 85%. In 2006, ELCAP published two more articles (1,8), which presented data on the proportion of lung cancers that were of stage I. In both of these articles, involving 269 and 484 patients with lung cancer, respectively, the fraction who had stage I cancer was also 85% (ie, between 84.5% and 85.49%). Even though the two smaller cohorts (7,8) can be considered subsets of the larger cohort [reported in (1)], it is unlikely that they would all have the same proportion of stage I lung cancers. Based on the binomial distribution and the joint probability of each cohort coming out at exactly 85%, I estimate that 98% of the time one or both of the smaller cohorts (7,8) would have had a proportion other than 85%, whereas roughly 2% of the time all three would have had exactly the same result.

Outlier Results

ELCAP Stage of Lung Cancer at Screen Detection Compared With Other Analyses

The 85% estimate for the proportion of all lung cancers that are stage I in screened cohorts also appears to be an outlier when compared with other studies (9–16) with analyses of the same end-point. In those other studies, which include single-arm studies in which all study subjects were screened with CT and preliminary reports from ongoing randomized trials in which only some subjects were screened with CT, the percentage of stage I lung cancers among subjects screened with CT is substantially lower than that reported by ELCAP (Figure 1). Combining the non-ELCAP estimates using random effects meta-analysis, the pooled proportion of stage I lung cancers, 56% (95% confidence interval [CI] = 50% to 62%), is less than the estimated 85% reported by ELCAP. In a funnel plot analysis, which simultaneously displays the estimate (x-axis) and the SE (y-axis) for the various studies, all three ELCAP estimates lie outside the funnel, which suggests that they are probably outliers compared with the other studies (Figure 2).

The reason that ELCAP estimates were higher than those of other researchers may be related to how stage at diagnosis was captured and reported (1,7,8). In one case, patients with advanced disease appear to have been excluded from ELCAP analyses, whereas in others (17–19), patients with advanced lung cancer may have been reclassified as having early stage disease. In the report published in 1999 (7), not all the patients found to have lung cancer through CT screening were included in the study. Instead, the authors describe excluding two patients who were found to have lung cancer in the airway and another two patients who were found to have lung cancer in the mediastinum. Although stage information on these subjects was not provided, it is reasonable to assume
that the two patients with cancer in the mediastinum almost certainly had advanced lung cancer. In an article published in 2007 (17), ELCAP researchers report that they categorized those patients who had more than one lung cancer at the time of diagnosis as having stage I disease, even though the TNM system would classify such patients as having either stage IIIA or stage IV lung cancer, but the number of patients whose stage was reclassified in this manner was not reported.

There is also some concern that for a few patients the stage of diagnosis may have been inaccurately captured. In two separate clarifications (18,19), ELCAP researchers reported that 10 of the original 13 untreated subjects who were originally reported as having stage IIIA or stage IV lung cancer, but the number of patients whose stage was reclassified in this manner was not reported.

ELCAP “Lung Cancer–Specific Survival After Screen Detection” Compared With Other Analyses

In the 2006 article (1), ELCAP reported a 5-year lung cancer–specific survival rate of 85% for all patients who were diagnosed with lung cancer through screening (Figure 2 of that article). In a later analysis with a longer follow-up (5), ELCAP reported a similar 5-year lung cancer–specific survival rate (Figure 8 of that article). However, other analyses (9,20–23) of this endpoint have produced a different, substantially lower, estimate. In an analysis in which I was involved of three single-arm studies of CT screening (4), the lung cancer–specific survival rate by the 2-year time point among individuals diagnosed through CT screening (77%, 95% CI = 68% to 84%) was less than the ELCAP 5-year estimate of 85%. Given the 2-year survival estimate from our analyses (4), it is unlikely that a survival time distribution at 5-years would be much greater than 55%. For example, an exponential distribution of survival times would lead to a 5-year lung cancer–specific survival rate of 52% in our analysis (4).
A possible explanation for why lung cancer–specific survival appears to be much higher in the ELCAP research than in other studies may be related to an undercount of the number of lung cancer deaths, as discussed above. The “event” in an analysis of lung cancer–specific survival is death from lung cancer. If such deaths are incompletely captured, as may have happened in the ELCAP studies, the lung cancer–specific survival estimate will be biased upward.

The NLST researchers have also examined the 5-year survival endpoint in the CT screening arm of their study (5-year lung cancer–specific survival after diagnosis, 54%, 95% CI = 51% to 58%; Dr Christine Berg, personal communication). The consistency between the NLST and our study (5) lends weight to the estimate of a 5-year lung cancer–specific survival rate for individuals with screen-detected lung cancer being closer to 50%–55%, than the 85% reported by ELCAP. The design and conduct of the NLST has several strengths that enhance its reliability, including data oversight, safety committees, and a systematic approach for ascertaining numbers and cause of death.

Conclusions
Before ELCAP results are incorporated into evidence-based clinical guidelines, there should be greater clarity regarding their reported research. The lingering questions involve discrepancies, statistically improbable observations, and apparent outlier findings on the key endpoints of the ELCAP research program. These include how frequently individuals die of lung cancer after screening, what proportion of individuals are diagnosed at stage I rather than a more advanced stage, and the likelihood an individual will survive after being diagnosed with lung cancer through screening.

Did only 75 subjects die of lung cancer in the study published in 2006 in *New England Journal of Medicine* (1), or is the number closer to 264, as suggested by the estimates from two ELCAP subset analyses that used death record searches to ascertain these events? Is the proportion of lung cancers that are found at stage I in screening programs around 85%, as all three of ELCAP’s studies report (1,7,8), or is it closer to 56%, as suggested by the other analyses of this endpoint (9–16)? Were all patients with early stage cancer who were untreated followed to the endpoint of death, or were there some individuals who were censored, which would cause the survival estimates for this group to possibly be higher than the 0% that ELCAP has reported (1,5)? Is the lung cancer–specific survival of patients after screen detection extremely high, on the order of 85% at 5 years, as ELCAP has reported (1,5), or is it more on the order of 50%–55% at 5 years, as the NLST has found (Dr Christine Berg, personal communication) and our study seemed to suggest (4)?

The Institute of Medicine recently released a set of standards for the incorporation of evidence into systematic reviews that form the basis of clinical practice guidelines (24). Several of their standards focus on the reliability of published data and emphasize the need to clarify outstanding issues before incorporating such data into an evidence review. I believe that in the case of ELCAP data, the findings cannot be properly interpreted without understanding the basis of some of the conflicting, improbable, and outlier findings.

**References**


**Funding**
This work was supported by an award from The Canary Foundation (San Jose, CA).

**Notes**
I thank Joshua N. Mirkin, A.B., Health Outcomes Research Group, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, for his input into the Commentary. The funders did not have any involvement in the design of the study; the collection, analysis, and interpretation of the data; the writing of the article; or the decision to submit the article for publication.

**Affiliation of author:** Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY.