Improved Survival and Higher Mortality*

The Conundrum of Lung Cancer Screening

Jerome M. Reich, MD, FCCP

(CHEST 2002; 122:329–337)

Key words: lung neoplasm; mass chest radiograph; screening

Abbreviations: CG = control group; CR = chest radiograph; ELCAP = Early Lung Cancer Detection Project; FN = false negative; FP = false positive; HRCT = high-resolution CT; IG = intervention group; LC = lung cancer; LCM = lung cancer mortality; LDCT = screening helical low-dose CT; MLP = Mayo Lung Project; mSv = millisieverts; NSCLC = non-small cell lung cancer; SCR = screening chest radiograph; TN = true negative; TP = true positive; TVDT = tumor volume doubling time

Improvement in cancer survival rates constitutes such intuitively forceful evidence of progress in cancer treatment that its corollary, improvement in mortality rate, appears self-evident. However, Welch et al.,1 employing data from the nationally compiled Surveillance, Epidemiology, and End Results program, found no overall correlation between increased survival and mortality rates for 20 of the most common types of solid tumors. Survival (5-year) is the percentage of individuals alive 5 years following a cancer diagnosis. Incidence and (cause-specific) mortality rates are age-adjusted, population-based rates. Incidence is the number of new cases per 100,000 population per year, and the (cause-specific) mortality rate is the number of deaths (from the disease) per 100,000 population per year. Increased survival, with no change in mortality rate, is most often attributable to lead-time bias, in which improvements in ascertainment permit diagnosis at an earlier point in time without affecting longevity (outcome). Thus, earlier diagnosis of a highly lethal and untreatable neoplastic disorder would improve survival without affecting mortality rate. Less obviously, if an increase in longevity resulting from intervention in some individuals was offset by a delayed decrease in longevity in others, survival would improve and the mortality would remain unchanged. Survival and mortality are not complementary measures.

Increased incidence may reflect a genuine secular trend, but it will be influenced by diagnostic improvements as well, particularly screening. In cancers with known effective treatments (eg, breast and testicular cancer), increased survival is, by definition, accompanied by decreased mortality, irrespective of an increased incidence attributable to improved diagnostics. Thus, compared with the period from 1950 to 1954, in the period from 1989 to 1995 the 5-year breast cancer survival rate increased from 60 to 86%, incidence increased by 55%, and the mortality rate decreased by 8%. For the same periods, the 5-year survival rate for testicular cancer increased from 57 to 96%, incidence increased by 106%, and the mortality rate decreased by 73%. In contrast, although the 5-year survival rate for lung cancer (LC) more than doubled (from 6 to 14%) in the same periods, the increase in mortality rate (259%) exceeded the increase in incidence (249%).1 An increase in LC incidence and survival are the expected consequences of the widespread employment of screening chest radiographs (SCRs) augmented by diagnostic and therapeutic advances. If this combination were effective, the mortality rate would necessarily decrease. If the observed increased LC mortality (LCM) rate reflected a secular trend in tumor aggressiveness, one would expect survival to decrease. Because increased survival is credited to therapeutic effectiveness, increased mortality and increased survival appear to be mutually exclusive outcomes. Two large, prospective, randomized, controlled, LC screening trials exhibited the same apparent paradox.

*Dr. Reich is in private practice in Portland, OR.
Manuscript received June 8, 2001; revision accepted January 4, 2002.
Correspondence to: Jerome M. Reich, MD, FCCP, 5051 SW Barnes Rd, Portland, OR 97221-1517; e-mail: Reichje@duanmail.com

www.chestjournal.org

For editorial comment see pages 1, 3
LC Screening

Ten studies addressing the efficacy of LC screening, each demonstrating no mortality rate benefit in the screened group, have been vetted by Richert-Boe and Humphrey,² Strauss et al.,³ Frame,⁴ Patz et al.,⁵ and Porter and Spiro,⁶ and, more succinctly and acerbically, by Parkin and Pisano.⁷ ⁸ The latter observed that failure to achieve a mortality rate reduction despite a survival rate improvement can be plausibly ascribed to the following three effects of screening: lead-time bias; length-biased sampling; and overdiagnosis.⁷ ⁸ Lead-time bias refers to making an earlier diagnosis without affecting mortality. Length-biased sampling refers to the observation that screening preferentially identifies phenotypically indolent LC. These lethal cancers have a lengthier preclinical phase than aggressive LCs and are, therefore, much more likely to be identified by screening than aggressive LCs, which more often will become clinically manifest or will run their course between screenings. This bias is evident principally in prevalence screenings. Overdiagnosis bias refers to the observation that some preclinical cancers that are capable of being identified by screening either are unaggressive or are associated with lethal comorbidities such that, in either case, they do not progress to clinical disease during the persons lifetime (ie, they are not clinically relevant). (Prostate cancer, a frequent incidental finding at postmortem examination, is a familiar example.) Patz et al⁵ have provided a detailed discussion and a graphic analysis of these biases. Screening proponents⁵⁻⁷ point out that the survival rate of patients with LC that has been diagnosed in stage IA (ie, solitary pulmonary nodules; and peripheral malignancies ≤ 3 cm, free of hilar or mediastinal nodal metastases or pleural invasion), who are precisely those patients who are likely to be identified by frequent screening, is so favorable (60% at 5 years¹⁰) compared to cancers that become clinically evident that screening should be encouraged in high-risk groups (principally, elderly smokers with COPD). However, mortality is considered to be the definitive screening outcome variable because it incorporates the consequences of intervention and it is unaffected by the three biases. For LC screening to achieve a reduction in mortality rate, it is necessary, though not sufficient, that it result in a stage shift. That is, in comparison with the control group (CG), the expected increase in LCs that are identified at a low stage in the screened group must be accompanied by an absolute decrease in the number of persons identified with advanced (unresectable) stages of the disease.

The Mayo Lung Project (MLP),¹¹ completed in 1984 and one of two large-scale, randomized, controlled, prospective studies, has provoked the greatest discussion both because of its design and its unanticipated findings. Following a prevalence screening of 10,933 male smokers ≥ 45 years of age, the investigators randomized into two groups. 9,211 individuals who were free of detectable LC, had a life expectancy of at least 5 years, and had sufficient respiratory reserve to tolerate a lobectomy. The intervention group (IG) of 4,618 persons was asked to have chest radiographs (CRs) and sputum cytologic examinations every 4 months for 6 years, and the CG of 4,593 persons received the following standard MLP advice: obtain an annual CR and sputum cytologic examination. At completion of the study, more cases of LC were found in the IG (206) than in the CG (160), and the LC survival was substantially better in the former group. There were two unexpected and disturbing findings as follows: (1) LC deaths were slightly higher in the IG (122) than in the CG (115) despite nearly identical surgical mortality rates; and (2) while the number of resectable LCs in the IG (99) was nearly twice that in the CG (51), the numbers of patients with advanced and unresectable LCs were nearly identical (IG, 107 patients; CG, 109 patients) [ie, the expected stage shift did not occur]. The investigators did not provide details of the all-cause mortality rate, and the post-1983 incidence of LC is unknown for either group because they were not followed-up after completion of the study.

Eddy¹² posited overdiagnosis in the IG to account for the 46 “missing cases” in the CG. Strauss et al.³ contended that the overdiagnosis of LC is implausible, and that “population heterogeneity” (ie, an imbalance in risk factors, due to flawed randomization, resulting in a higher proportion of individuals in the IG than the CG who were destined to develop LC) accounted for the surplus. As Parkin and Pisano⁸ noted, the probability of imbalance of risk factors in a randomized sample of this size is < 1%. Moreover, an analysis of the long-term mortality rate¹³ showed that both known LC risk factors and causes of death were nearly identical in both arms of the study, suggesting that randomization had achieved its expected goal.

Critics of the MLP have pointed out its deficiencies as follows: there was contamination of the CG (ie, about half reported obtaining an interval CR); it provided insufficient follow-up; and a beneficial effect could have been missed because of insufficient power. If contamination had materially weakened the study, the substantially improved survival in the IG would not have been observed. The brevity of the follow-up (mean, 3 years) has been suggested as an explanation for the 46 “missing cases” in the CG.
Clinically relevant but unrecognized LC would be expected to surface symptomatically with continued observation, resulting in a long-term CG LCM rate equaling or exceeding (if screening had been helpful) the IG LCM rate. Marcus et al\textsuperscript{13} have updated this study, with a median follow-up of 20.5 years as of 1996, and have found this not to be the case. Of 6,523 participants in the IG and CG who were known to be alive in 1983, the LCM rate was higher in the IG (4.4 deaths per 1,000 person-years) than in the CG (3.9 deaths per 1,000 person-years) \([p = 0.09]\). (The discrepancy in LCM rate may be attributable, in part, to “sticking diagnosis,” \(ie\), an incorrect attribution of death to a known preexisting disease.) LC survival was higher in the IG than in the CG from the time of randomization, suggesting that lead-time bias had little effect. The authors concluded that the much higher incidence and survival of LC in the IG excluded the possibility of significant weakening from contamination, that overdiagnosis was a plausible explanation for the higher incidence of LC in the IG, and that radiographic and cytologic screening provided no demonstrable reduction in mortality rate.

Kubík and Polá\v{k}\textsuperscript{14} undertook a similar study of Czechoslovakian male smokers who were 40 to 64 years of age. After an initial prevalence screening, which eliminated individuals with identifiable LC, the remaining participants were stratified by known LC risk factors and were randomized to an IG (3,172 persons) to receive an SCR every 6 months for 3 years and a CG (3,174 persons) that would receive only the final screening at 3 years. At 3 years, twice as many LCs (37) were discovered in the IG than in the CG (19). Although LC was identified at an earlier stage and the 5-year survival was better in the IG (as in the MLP), the mortality was higher in the IG (LCM, 45 deaths; all-cause mortality, 126) than in the CG (LCM, 40; all-cause mortality, 109). This study differed from the MLP by providing an SCR to both groups at the end of the initial 3-year period and an annual SCR for the following 3 years. At 6 years, 108 LCs were found in the IG and 82 LCs were found in the CG, a difference largely accounted for by the number of screening-detected cases in the initial 3-year period. After an additional 3 years of follow-up, which was limited to those persons with LC, there were 85 LC deaths in the IG vs 67 in the CG.\textsuperscript{15} The following three aberrant findings stand out: (1) at the completion of the initial 3-year period, which incorporated a final screening for both groups, the number of LCs identified in the IG was nearly twice that in the CG; (2) more than twice as many LCs were identified in the second period (134) than in the first period (56); and (3) the number of individuals identified with LC at 6 years was 32\% higher in the IG than in the CG. The authors were unable to ascertain the reason for the difference in proportion with LC. They suggested that it might be due to chance. This is unlikely (two-sample test: \(Z = 1.86; p = 0.031\)).

In summary, both trials violated expectations by demonstrating higher rates of LCM and all-cause mortality in the IG than in the CG arms despite exhibiting the following two essential, initial prerequisites of a successful screening program: a higher proportion of persons with resectable LCs; and an increased survival rate in the IG. They thus mirror and amplify the conundrum of the study of Welch et al,\textsuperscript{1} raising the question of how this might come about.

**Long-term Sequelae of Lobectomy**

In contrast to other cancers for which screening is advocated (\(ie\), breast, cervix, and colon cancer), the optimal treatment of LC entails the partial removal of a vital organ. Furthermore, the risk of LC is higher among smokers with advanced COPD (which correlates with all-cause mortality rate, particularly with coronary heart disease\textsuperscript{16}). Consider the following hypothetical scenario. Six thousand prescreened male smokers with widely varying coronary risk factors and pulmonary function levels, and mean values of 60 years of age, height of 69 inches, and FEV\(_1\) of 2.2 L (70\% of the predicted value [3.1 L]) are randomly divided into two groups of 3,000 individuals each. One group (IG) is radiographically screened, and the other (CG) is not. LC develops in 10\% of individuals, equally distributed between the IG and the CG, all in the first year of screening. In 200 individuals, the LC is aggressive; in 200, it is nonaggressive; and in 200, it is overdiagnosed. Assume further that all resectable LCs occur in the left upper lobe and are treated by lobectomy, that excess deaths from respiratory failure, pneumonia, or coronary disease occur when the mean FEV\(_1\) falls to 1 L, that the rate of FEV\(_1\) decline (55 mL/yr\textsuperscript{17}) is unchanged by surgery, and that the risk of pneumonia is increased by the surgically induced anatomic distortion and pleural reaction. With these assumptions, surgery would reduce the mean life expectancy from 22 to < 11 years (\(ie\), the mortality rate would be doubled (see Appendix A). Survival and mortality would be nearly equal in the IG and CG among persons with aggressive LC. Survival would be markedly increased (due to lead-time bias), and the mortality rate would be decreased slightly in the IG compared to the CG among persons with nonaggressive LC. However, in subjects with overdiagnosed LC, the survival would increase in the IG, but the
mortality rate would be double that in the CG. The net effect would be an increased incidence, survival, and mortality rate in the IG compared with the CG, mirroring the MLP and Czech experience.

Overdiagnosis plausibly accounts for the apparent paradox of improved survival and higher mortality rate in screened individuals. Surgery in overdiagnosed individuals will improve survival (due to lead-time bias) and increase the all-cause mortality rate (by accelerating the course of coexistent lung and heart disease). For this reason, the all-cause mortality rate is more discriminating than a cause-specific mortality rate in assessing the effect of LC screening. The selective screening of individuals who are at the highest risk (ie, those with advanced COPD) would be expected to increase both the likelihood of and the mortality rate attendant on overdiagnosis because of the limited pulmonary reserve and comorbidities affecting these individuals. Moreover, because of these features, this selection criterion would minimize the average duration of life saved.

**Overdiagnosis of LC**

The synonymous terms, overdiagnosis, clinically irrelevant cancer, pseudodisease, iatrogenic pseudodisease, and lanthanide disease, denote a pathologic finding lacking clinical import, detectable only by special means (in this instance, screening or autopsy). Broadly speaking, this can occur in the following two ways: the neoplasm may have little to no lethality; or the neoplasm may be phenotypically aggressive, but the affected individual may succumb to a competing morbidity. To individuals familiar with the often-aggressive biology of diagnosed LC, the term clinically irrelevant LC will appear to be antithetical to experience. Several lines of evidence indicate, however, that both low lethality and lethal comorbidities result in overdiagnosis. Black18 (in an editorial accompanying the article by Marcus et al13) stated that overdiagnosis was a highly plausible explanation for the MLP findings. He added that fatal comorbidities are very high in smokers as follows: some sputum cytology-detectable LCs are minute, and have an unknown natural history (eg, in situ squamous cell carcinoma); some adenocarcinomas grow very slowly; and pathologic false-positive (FP) results occur. (Distinguishing inflammatory or hyperplastic changes from malignancy may be difficult or impossible in some circumstances; for example, atypical adenomatous hyperplasia, which may be a precursor of adenocarcinoma, cannot be reliably distinguished from the latter.19) Referencing the higher LCM rate in the IG of the MLP study, Black18 added that the excess might have been attributable to the long-term complications of surgery in individuals with pseudodisease.

Direct evidence of overdiagnosis, beyond that due to tissue overinterpretation (ie, pathologic FP test results), derives from the following three sources: autopsy studies; observations on tumor growth rate; and analysis of tumor cell biology. McFarlane et al20 identified a large “reservoir” of undiagnosed LC cases that had neither caused nor contributed to the deaths of their patients in their 2,996-case autopsy series at Yale New Haven Hospital. After age adjustment, the incidence of autopsy-detected LC that was unsuspected during life was almost fourfold the community incidence rate in men and was almost 15-fold higher in women. In an updated review from the same site, Chan et al21 reported that one in six LCs that were detected at autopsy had not been recognized before death. Previously unsuspected LCs, 70% of which were resectable (ie, TNM stages 0 to II) and were therefore presumably not causing symptoms, were discovered in 1% of men on whom autopsies were performed between 1973 and 1982. The MLP found 46 more cases of LC in the IG than in the CG (206 cases vs 160 cases, respectively). The incidence in the IG was 5.5 cases (per 1,000 person-years).11 The incidence of putatively overdiagnosed cancer is 1.2 (46/206 × 5.5 = 1.2), or 16% of the incidence of clinically unsuspected LC discovered as a surprise (crude rate, 7.721) in concurrent (1973 to 1982) male autopsies in the Yale-New Haven series.

Aoki et al22 reported that, among resectable adenocarcinomas < 3 cm in diameter, tumor volume doubling time (TVDT) ranged from 42 to 1,486 days, and half the tumors had TVDTs of > 1 year. With a TVDT of 1 year, under the assumption of a constant rate of growth, 8 years are required for a tumor diameter to increase from 5 mm to 3 cm, providing sufficient time to die of competing causes.

There is a biological difference in aggressiveness between screening-detected and incidentally detected LCs.23 The latter are more undifferentiated, as indicated by DNA flow cytometry (which showed a higher DNA synthesis phase fraction). However, this finding may be a temporal artifact, since malignancies become more phenotypically aggressive over time and screening-detected LCs are earlier in their natural history.

**Hazards Inherent in Screening**

The aforementioned reservations address only those patients who, with the exception of cases of tissue overinterpretation, have true-positive (TP) screening test results (ie, LC). An unintended and unavoidable consequence of screening is the identi-
fication of radiographic abnormalities, principally on prevalence screening, in a substantial proportion of participants that, on further analysis, are determined to be benign (ie, FP test results). Although the psychological burden of an indeterminate finding cannot be quantified, some individuals will elect to undergo thoracotomy when LC cannot be excluded with certainty. Eddy emphasized that, in the initial (prevalence) screening at the Sloan-Kettering Institute, about 10% of participants required further evaluation, that about half underwent one or more invasive procedures (including thoracotomy) to rule out LC, and that 99% of the individuals with positive SCR findings had FP test results.

The MLP reported 4.4% abnormal CRs in the prevalence screening of 10,933 participants; LC was found in 91 participants. Thoracotomy was required for indeterminate lesions in 56 individuals, and 28 individuals had benign disease. The prevalence screening in the Johns Hopkins Hospital study found 15% of participants with abnormalities requiring further investigation, and 0.6% of the participants had LC. "Further evaluation" represents a spectrum ranging from inconvenient and inexpensive (eg, a review of earlier CRs or repeat radiographs), to expensive and invasive (eg, bronchoscopy, which is particularly arduous in individuals with a positive cytology test result and a normal SCR finding, and transcutaneous or transbronchial needle biopsy), to diagnostic thoracotomy for indeterminate lesions. Unfortunately, the number and type of invasive procedures necessitated by FP SCR was not provided in this (or any other) series.

Participants receiving true-negative (TN) results have the intangible benefit of reassurance, the influence of which (in addition to the promise of continued surveillance), if any, on their motivation to cease smoking is unknown. For false-negative (FN) test results, which are surprisingly frequent (90% of the 45 peripheral carcinomas in the MLP were visible in retrospect despite triple reading), litigation for failure to diagnose is a potential hazard. FN interpretations are likely to be more frequent in community settings where diagnostic sensitivity will fall short of that achieved in investigative settings that employ double or triple readings by radiologists expressly and closely scrutinizing SCRs for minute coin lesions.

Screening Helical Low-Dose CT Scanning

Low-dose CT (LDCT) scanning is four times as sensitive as an SCR. Compared with standard helical CT scanning, it is less costly, faster, and entails lower radiation exposure (approximately 0.5 vs approximately 5.0 millisieverts [mSv]). (Millisieverts is an effective dose equivalent. For comparison, a standard posteroanterior and lateral CR entails approximately 0.1 mSv, and the average exposure to background radiation in Germany is 2.4 mSv/yr and in the United States, is 3 mSv/yr. High-resolution CT (HRCT) scanning imposes a lower exposure than standard CT scanning, reflecting the far lower volume of tissue imaged. The carcinogenic effect of these exposures, if any, in the age group undergoing screening is exceedingly small. For example, the lifetime nominal risk of excess cancer mortality for an exposure of 0.5 mSv, which is equivalent to exposure from one LDCT scan, is 2 cases per 100,000 population.

All of the reported LDCT studies have been single-cohort noncomparative design. Kaneko et al triple-screened 1,369 individuals with sputum cytology examinations, CRs, and LDCT scans. The latter two tests demonstrated an abnormality in 701 individuals, 229 of whom underwent HRCT scans. Of the 19 who underwent biopsies, 15 had LCs (vs 4 with LCs identified on CRs), which were at stage I in 14 individuals, with a mean diameter of 1.6 cm. Sone et al serially screened 5,483 volunteers, age ≥ 40 years of age, from the Nagano Prefecture in Japan with CRs and LDCT scans over a 3-year period. Forty-six percent of participants were women, 55% were ≥ 60 years of age, and 54% were never-smokers (defined as smoking < 1 pack-year). The authors identified 588 suspicious lesions (criteria not specified) on 13,786 LDCT scans (4.3%). Of 72 individuals undergoing thoracotomy, LC was confirmed in 56 (adenocarcinoma in 51), atypical adenomatous hyperplasia in 9, and 7 had benign disease. A total of 60 individuals with LC were identified by screening. Of these, 20 had received earlier FN results from screenings, 53 (88%) had tumors that were at surgical stage IA, and 39 had tumors exceeding 1 cm in size. The 3-year LC detection rate of 406 persons per 100,000 population was nearly fourfold the 3-year regional LCM rate (annual rate, 37/100,000 × 3 = 111 persons per 100,000 population), suggesting, as one possibility, the occurrence of overdiagnosis.

The Early Lung Cancer Action Project (ELCAP) prevalence-screened 1,000 medically fit asymptomatic smokers (age, ≥ 60 years; median age, 67 years) with CRs and LDCT scans. One or more noncalcified nodules were identified in 233 participants (23%) by LDCT scans and in 68 participants by CRs. One hundred eighty-three of these participants underwent either HRCT scanning for confirmation and further evaluation or were advised to have periodic HRCT scans at 3, 6, 12, and 24 months (the decision rested on the size and appearance of the tumor on
the HRCT scan). Twenty-seven of these participants (2.7%) had LCs (vs 7 with LC detected by CR), which were at pathologic stage IA in 23. Twenty-eight patients underwent uncomplicated fine-needle transthoracic biopsy or video-assisted thoracoscopy, and 27 of these had confirmed LCs (almost all of which were adenocarcinomas or variants). Twenty-six patients with resectable disease underwent lobectomy and mediastinal lymph node dissection (pathologic staging was N0 in all participants). Three participants who had undergone serial HRCT scans demonstrating stability elected to undergo biopsy (which was not recommended in the protocol); each had benign disease.

The authors recently updated their study with a report of the annual incidence screening.33 There were two interval LCs (ie, LCs identified because of symptoms, not from screening) that presented as endobronchial lesions, a non-small cell LC (NSCLC) [stage IIB], and a limited stage small cell LC. Of the original 1,000 participants, 841 received one or more rescreenings, totaling 1,184 imaging studies. Thirty rescreenings (2.5%) had positive results (vs 23% in the prevalence screenings). Two subjects with positive screening results died of cardiovascular disease; in 12 subjects the abnormality resolved either spontaneously or in response to antibiotic therapy, and in 8 subjects no further growth was observed on sequential screenings. In the remaining eight individuals, the demonstration of growth led to the identification of NSCLC in six (stage IA, five persons; stage IIIA, one person). In the two remaining subjects, one was found to have limited-stage small cell LC, and a fine-needle aspiration biopsy showed a noncancerous process in the other. Of the seven detected NSCLCs, two were not operated on because of medical contraindications, three underwent lobectomy, one underwent pneumonectomy, and one operable patient elected radiosurgery. All four individuals who had surgery underwent mediastinal lymph node resection, and none had contralateral lymphatic involvement. Pathologic staging showed stage N0 in three patients and stage N2 status in one patient. Indicative of their rapid growth, neither of the two interval cases nor the stage IIIA case was definitely visible on the prior LDCT scan. No patient who was free of LC underwent a thoracotomy, and there were no complications resulting from the invasive procedures.

Jett44 reported that 775 persons (51%) among an LDCT-prevalence-screened population of 1,520 present or former smokers ≥50 years of age had positive test results. LC was found in 19 persons (2.5%), of whom 2 had small-cell LC and 10 (53%) had stage IA NSCLC. The outcomes for the remaining 97.5% of participants with positive screening results is, at present, unknown. Ninety-three percent of the individuals requiring HRCT scans in the trial by Kaneko et al30 were not found to have LC during the trial. The rate for the study by Sone et al31 was 90%, and for the ELCAP prevalence trial32,33 it was 84%. The lower figure in the ELCAP may reflect the subsequent identification of TP findings in follow-up HRCT scans. These figures should be regarded as provisional estimates of the percentage of positive results that were not due to LC (FP test result) because serial HRCT scan results are pending in most patients with suspicious LDCT scan abnormalities. For this reason, specificity (ie, TN/[TN + FP]) cannot be assessed.

The incidence screening of the ELCAP study33 demonstrated that the number of FN results was surprisingly high in the prevalence screening despite dual readings. Three of the six NSCLC tumors had been “clearly visible (in retrospect) on the previous screening.” However, all three tumors remained at stage IA and were resected. Compared to the baseline study, in which 24 of 27 LCs were resected, only 5 of 9 LCs (the latter figure includes the two interval cases) were resected following the incidence screening (one patient, who had been judged operable and resectable, chose radiotherapy). The difference in the proportion of individuals in the prevalence and incidence screenings who underwent resection reflects a combination of the effects of comorbidities obviating surgery (two cases) in the population selected for screening and, probably, the favorable influence of length-biased sampling in the prevalence screening.

The design of the LDCT studies rests on the following three premises that are open to question (the first two have been considered): (1) the implausibility of overdiagnosis eliminates the requirement for a control arm; (2) survival improvement (compared to historical control subjects) suffices to demonstrate the efficacy of screening with LDCT scan; and (3) earlier diagnosis and greater detection ensure a lower mortality rate.

Patz et al35 found no significant correlation between tumor size and survival in 510 patients with stage IA LC. This suggests a high degree of biological variability (ie, LCs that reach 3 cm without mediastinal metastases represent a phenotypically favorable subtype of smaller tumors that are similarly confined). The authors cautioned that the detection of smaller LCs by means of LDCT scanning would not necessarily improve the mortality rate, adding, “...there are no clinical or experimental data to support that there is a [size] threshold that represents an essential prognostic determinate.”

The ELCAP protocol recommends serial HRCT scanning to assess growth for noncalcified nodules
pointed out by Fidler, the vast majority of circulat-
logical import of this observation is unknown. (As
following aspiration. As the authors affirm, the bio-
20 pleural washing specimens were cytology-positive
specimens were cytology-positive before, and 12 of
The authors reported that 2 of 20 pleural washing
biopsy of the lung has a potential to disseminate LCs.
biopsy.

Pantel et al, employing an ultrasensitive immu-
nocytotoxic technique, found micrometastases in
the bone marrow in 60% of patients with resectable
NSCLC, in 54% of those with negative lymph nodes,
and, collectively, in 55% of patients with stage T1
and T2 primary tumors. Furthermore, they found no
correlation between the size of the primary tumor
and the presence of bone marrow micrometastases.
Cote et al, employing a different marker, found
that 5 of 15 patients with surgical stage I LCs had
bone marrow micrometastases. Both studies
reported a strong correlation between the presence of
bone marrow micrometastases and tumor recur-

Weiss estimated that the natural history of LC
encompassed 40 volume doublings, 27 to reach a
diameter of 5 mm, and 3 further doublings to
achieve a diameter of 1 cm (see Appendix B). The
putative benefit of screening with LDCT scans rests
on the premise that intervention between these two
events (27 and 30 doublings) will substantially re-
duce the frequency of metastases. Assuming a TVDT
of 1 year, individuals with 5-mm tumors identified
by LDCT scanning will experience a 3-year improve-
ment in survival time (compared with individuals
identified with 1-cm tumors by SCRs) solely due to
lead-time bias.

In summary, theoretical and experimental evidence
dicates that earlier detection will increase
overdiagnosis and will fail to substantially influence
mortality rate. Tumor biology (phenotype) is deter-

OTHER SCREENING CONSIDERATIONS

Sawabata et al found that fine-needle aspiration
biopsy of the lung has a potential to disseminate LCs.
The authors reported that 2 of 20 pleural washing
specimens were cytology-positive before, and 12 of
20 pleural washing specimens were cytology-positive
following aspiration. As the authors affirm, the
biological import of this observation is unknown. (As
pointed out by Fidler, the vast majority of circulating
tumor cells are destroyed.) In experienced hands,
the rate of major bleeding from invasive diagnostic
procedures is 0.6%, and pneumothorax complicates
transcutaneous needle aspiration in 12% of patients,
one fourth of whom require chest tube drainage.

Three criteria determine the utility of screening
tests. Efficacious means that the procedure is supe-
rior in controlled trials (ie, it works). Effectiveness
refers to whether it works in a noninvestigative
setting. Efficiency is a cost-benefit analysis.

It is predictable, given the heightened sensitivity
of LDCT scanning in comparison to SCRs, that the
combination of lead-time-biased and length-biased
sampling will favorably affect survival. Reflecting
the issues of both overdiagnosis and the long-term
sequelae of pulmonary resection, it is equally predict-
able that lack of a control arm will lead some to
question whether survival improvement constitutes
decisive evidence of efficacy. Estimates of the aver-
age duration of life saved is beyond the scope of this
article, but it should be noted that the median age of
ELCAP participants was 67 years, that all partici-
pants were heavy smokers (mean, 45 pack-years) and
therefore were at high risk for heart and lung
disability and death, and that occult metastases are
present in about half of apparently curable LCs.

Whether community physicians employing LDCT
scanning would reproduce the notable sensitivity,
low morbidity, and minimization of invasive proce-
dures (including resection) displayed by the ELCAP
investigators is open to question.

Absent evidence of efficacy or effectiveness, an
estimate of efficiency cannot be undertaken. How-
ever, one may estimate the annual cost of screening
persons who are at risk. The 2000 US census found
that there were 51 million Americans between the
ages of 50 and 69 years. Approximately half of them
are present or former smokers. Assuming that 10%
would be unsuitable screening candidates because of
comorbidities, and a charge of $250 for an LDCT
scan, the annual cost of the initial screening of 23
million at-risk individuals would be $6 billion. An
additional $600 for an HRCT scan in 23% of subjects
found to have an abnormal LDCT scan would add $3
billion, and for the 90% of those with indeterminate
lesions that were not subjected to biopsy, up to four
additional HRCT scans would be required, entailing
an additional undiscounted cost of $11 billion over 2
years, three fourths of which would occur in the first
year. The aggregate cost, exclusive of professional
consultation fees as well as charges for diagnostic (ie,
fine-needle aspiration biopsy and video-assisted thor-
coscopy) and therapeutic interventions, which are
likely to be quite large in the aggregate, would
therefore be $20 billion.

A substantial cost reduction might be achievable
with dedicated screening systems, and incidence-
screening costs will be substantially lower than the
prevalence screening. Sone et al, employing a
dedicated system, estimated the cost of LDCT scan-
ning at 5,000 Japanese yen (approximately $50) per
person. The ELCAP follow-up study demonstrated
that the rate of positive test results (one to five

www.chestjournal.org
noncalcified nodules) diminished from 23% on the initial (prevalence) screening to 2.5% on the incidence screening. Thus, the cost for incidence screening would be about $7 billion; $6 billion for the annual LDCT scan plus approximately $1 billion for one or more HRCT scans in individuals with positive test results not evident on the prevalence screening. The reduced utility of the incidence screening, however, partially offsets the two-thirds reduction in cost. The yield of patients undergoing resection for stage IA LC diminished sevenfold between the prevalence screening (23 of 1,000 screenings) and the incidence screening (4 of 1,184 studies).

There is a subtle, implicit ethical difference in a physician’s obligation to patients seeking medical care for symptoms of disease and to healthy persons seeking to remain in that state. For the former, the physician is required to provide the best care his knowledge and skill can offer, even if he is uncertain of benefit; for the latter, he is ethically obligated to be certain of benefit. LC screening fails the latter test because survival improvement does not presage mortality rate improvement. Screening for LC clearly disadvantages those with FP test results, and FN test results risk litigation. For TP test results, earlier diagnosis has no proven benefit, for the overriding outcome determinant is tumor biology. On the contrary, based on the available evidence, it appears harmful. Even if efficacy were definitively demonstrated in TP test results, its costs would be disproportionate to its expected benefit, and it is likely to prove less cost-effective in community settings. Community LC screening should be discouraged, pending the demonstration of an all-cause mortality rate advantage either in the ongoing National Cancer Institute-sponsored Prostate, Lung, Colon and Ovarian study, which has enrolled nearly 155,000 men and women 60 to 74 years of age in a randomized controlled screening trial, or a similarly constituted and powerful study.

ACKNOWLEDGMENT: I wish to express my gratitude to Pamela Marcus, PhD, National Cancer Institute, for her meticulous review of the manuscript and for her numerous helpful comments.

APPENDIX A

Under the assumption that the apical-posterior segment is equivalent to the apical posterior segments of the right upper lobe, the left upper lobe constitutes 5/19 of the pulmonary segmental anatomy. Surgical extirpation would remove this fraction of the preoperative (2.2 L) FEV₁, leaving 14/19 of this value (1.6 L). At an annual FEV₁ decline of 55 mL, an additional 600 mL would be lost in 11 years, leaving 1 L. Absent surgery, it would require 22 years to attain this value.

APPENDIX B

Derivation and Application of a Formula to Calculate the Number of Tumor Volume Doublings Required to Achieve a Specified Tumor Diameter

The volume of a sphere is given by \( \frac{4}{3} \pi r^3 \). Volume doubling therefore requires an increase of the radius by a factor of the cube root of 2 (i.e., 1.26). Under the dual assumptions that tumor growth is exponential and that the cell diameter is 0.001 cm, the number of volume doublings required to achieve a specified diameter is given by: tumor diameter (in centimeters) = cell diameter (cube root of 2 \( \times \) number of volume doublings or 0.001(1.26)\(^3\)).

The log form simplifies computation:

\[
x = \log(1000 \times \text{diameter}) \div \log 1.26
\]

Thirty doublings, three quarters of the life history of the tumor, are required to achieve a 1-cm diameter tumor; the threshold for radiographic detection. Forty doublings would produce a 10-cm diameter tumor. Twenty-five doublings are required to achieve a 3-mm diameter tumor; the threshold for detection by LDCT scanning. As a convenient rule of thumb, three volume doublings (an eightfold change) are required to double the diameter of a spherical tumor.

REFERENCES

34 Jett JR. Spiral computed tomography screening for lung cancer is ready for prime time [editorial]. Am J Respir Crit Care Med 2001; 163:812–813
43 Meltzer MI. Introduction to health economics for physicians. Lancet 2001; 358:993–998