

Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up

Pamela M. Marcus, Erik J. Bergstralh, Richard M. Fagerstrom, David E. Williams, Robert Fontana, William F. Taylor, Philip C. Prorok

Background: The Mayo Lung Project (MLP) was a randomized, controlled clinical trial of lung cancer screening that was conducted in 9211 male smokers between 1971 and 1983. The intervention arm was offered chest x-ray and sputum cytology every 4 months for 6 years; the usual-care arm was advised at trial entry to receive the same tests annually. No lung cancer mortality benefit was evident at the end of the study. We have extended follow-up through 1996. **Methods:** A National Death Index-PLUS search was used to assign vital status and date and cause of death for 6523 participants with unknown information. The median survival for lung cancer patients diagnosed before July 1, 1983, was calculated by use of Kaplan-Meier estimates. Survival curves were compared with the log-rank test. **Results:** The median follow-up time was 20.5 years. Lung cancer mortality was 4.4 (95% confidence interval [CI] = 3.9–4.9) deaths per 1000 person-years in the intervention arm and 3.9 (95% CI = 3.5–4.4) in the usual-care arm (two-sided *P* for difference = .09). For participants diagnosed with lung cancer before July 1, 1983, survival was better in the intervention arm (two-sided *P* = .0039). The median survival for patients with resected early-stage disease was 16.0 years in the intervention arm versus 5.0 years in the usual-care arm. **Conclusions:** Extended follow-up of MLP participants did not reveal a lung cancer mortality reduction for the intervention arm. Similar mortality but better survival for individuals in the intervention arm indicates that some lesions with limited clinical relevance may have been identified in the intervention arm. [J Natl Cancer Inst 2000;92:1308–16]

The use of chest x-ray and sputum cytology as mass screening tools for lung cancer fell out of favor in the early 1980s when a number of clinical trials found no reduction in lung cancer mortality for screened individuals (1). Regardless of these results, chest x-ray as a screening tool is advocated by many people, primarily because of improvements in technology, design limitations of the early studies (2), and the absence of other early-detection procedures. Lung cancer is currently the leading cause of cancer-related death in both men and women in the United States (3), and because symptoms often do not appear before the disease is advanced (3), secondary prevention is an appealing option.

The Mayo Lung Project (MLP), a National Cancer Institute (Bethesda, MD)-funded, randomized, controlled clinical trial that was conducted between 1971 and 1983, observed no reduction in lung cancer mortality with an intense regimen of chest x-rays and sputum cytology (every 4 months for 6 years) (4). This finding has been questioned by proponents of lung cancer

screening, most often on the grounds that the trial did not have adequate statistical power to identify a very modest reduction in lung cancer mortality and on the presumption that substantial contamination in the control arm reduced power even further (5). In addition, statistical modeling and increased knowledge regarding lung cancer progression suggest that the follow-up time in the MLP may have been too short (an average of 3 years after the last screening) for observation of a screening benefit (6).

We have extended follow-up of the MLP participants through the end of 1996, with the goal of examining whether additional time would allow for a reduction in lung cancer mortality to be observed in the intervention arm. Because participants diagnosed with lung cancer in that arm experienced more favorable survival as of July 1, 1983, we also wanted to explore whether that trend continued, with an eye toward assessing the impact of lead-time bias (earlier diagnosis of disease but no postponement of death) and overdiagnosis (identification of lesions with limited clinical relevance that would not have been detected in the absence of screening) (7), two common screening biases that may be responsible for what some suggest are conflicting findings in the MLP (8).

METHODS

The Mayo Lung Project

The MLP was designed to evaluate whether an intense regimen of lung cancer screening by use of chest x-ray and sputum cytology would reduce lung cancer mortality in male smokers. From November 1971 through July 1976, 10 933 male Mayo Clinic (Rochester, MN) outpatients who smoked and were not suspected of having lung cancer were prevalence screened for the disease by use of chest x-ray and sputum cytology. Individuals who tested negative for lung cancer, who had a life expectancy of at least 5 years, and who had a respiratory reserve that was considered adequate to undergo lobectomy, if necessary, were invited to participate in a randomized, controlled clinical trial of lung cancer incidence screening. A total of 9211 men were randomly assigned to one of two study arms.

In the intervention arm (*n* = 4618), participants were offered (and reminded to receive) a free chest x-ray and sputum cytology every 4 months for 6 years. Noncompliant participants were contacted annually by letter. During the 6-year screening period, compliance with the scheduled testing averaged 75% (4). Only

Affiliations of authors: P. M. Marcus, R. M. Fagerstrom, P. C. Prorok, Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD; E. J. Bergstralh, W. F. Taylor (Biostatistics Section), D. E. Williams, R. Fontana (Division of Pulmonary and Critical Care Medicine), Mayo Clinic, Rochester, MN.

Correspondence to: Pamela M. Marcus, Ph.D., M.S., National Institutes of Health, 6130 Executive Blvd., MSC 7354, Suite 344, Bethesda, MD 20892-7354 (e-mail: pm145q@nih.gov).

See "Notes" following "References."

© Oxford University Press

12 subjects in the intervention arm were lost to follow-up (4). Compliant participants in the intervention arm were contacted annually by letter after completing their 6 years of screening as a means of identifying potential lung cancer diagnoses and deaths.

Participants in the usual-care arm ($n = 4593$) merely received, on enrollment in the trial, the Mayo Clinic's standard 1970 recommendation to receive an annual chest x-ray and sputum cytology. Throughout the study, these participants received the same annual questionnaire that intervention participants received after cessation of their screening regimen. A special questionnaire that included questions regarding chest x-ray was mailed to usual-care participants around the end of the project (i.e., July 1, 1983). On that questionnaire, 3309 (72%) participants reported having had their last chest x-ray between 1972 and 1984; another 261 (6%) reported a chest x-ray but were unsure of the year. Fourteen subjects in the usual-care arm were lost to follow-up (4).

The MLP ended on July 1, 1983. At that point, lung cancer mortality was similar in both arms: 3.2 lung cancer deaths per 1000 person-years in the intervention arm versus 3.0 lung cancer deaths per 1000 person-years in the usual-care arm. In the intervention and usual-care arms, 206 and 160 participants, respectively, had been diagnosed with lung cancer. In the intervention arm, a greater proportion of the participants who were diagnosed with lung cancer had early-stage disease.

For additional information regarding the MLP, see (4,9,10).

National Death Index Search

The National Death Index (NDI; Hyattsville, MD) was used to follow-up the 6523 MLP participants who were known to be alive on July 1, 1983, and for whom vital status and date and cause of death, as of December 31, 1996, were unknown. The Institutional Review Board of the Mayo Clinic approved this project.

Details of NDI matching procedures are discussed more thoroughly elsewhere (11,12). Briefly, national death-certificate files are searched to identify possible matches for each record submitted by use of criteria based on Social Security number, name, middle initial, date of birth, and father's surname (for females). For each potential matching death certificate (usually more than one is identified), a value is assigned to each of three variables. The SCORE variable (as it is referred to in NDI documentation) is a probabilistic score that reflects the degree of matching. CLASS is a five-level categorical variable that reflects the number of items submitted, the numbers of items that match, and the specific items that matched. SEQ is a ranking that is based on the value of SCORE, with $SEQ = 1$ indicating the most likely match for a given submitted record. A dichotomous variable, the STATUS variable, is also assigned; it reflects NDI's suggested (but not required) criteria for true and false matches. (These matching criteria are referred to as "the recommended NDI algorithm" throughout this article.)

Death-certificate files for the years 1983 through 1996 were searched against the submitted records of the 6523 MLP participants. A total of 23 651 potential matches were returned, with the following information provided for each: date of death, exactness of match (i.e., whether fields matched and which specific characters within the fields matched), and the SCORE, CLASS, STATUS, and SEQ variables. Because we requested an NDI-PLUS search, we also received information on the coded cause of death, including the underlying cause of death. NDI-PLUS also returns codes that indicate secondary causes of death and medical conditions that may have contributed to death (the entity and record codes).

Matching Algorithms and Determination of Cause of Death

The algorithm that we developed to identify "true" death-certificate matches for the records of the 6523 MLP participants used the CLASS, SCORE, and SEQ variables. All matches with SEQ greater than 1 were considered to be false matches. To classify the remainder of the matches as true or false, for each value of CLASS we chose a cut point that maximized the percent of correctly classified matches in the calibration sample provided in the NDI documentation. For $CLASS = 2$, the match was true if SCORE was greater than or equal to 49.5; for $CLASS = 3$ and $CLASS = 4$, the corresponding values of SCORE were 37.5 and 27.5, respectively.

The underlying cause of death listed on the death certificate was considered to be the primary cause of death. Unless otherwise noted, all analyses used this set of deaths. Causes of death were grouped according to the following International Classification of Diseases, 9th Revision, codes (13): lung cancer, code 162; other cancers, codes 140–161 or 163–208; chronic obstructive pulmonary disease (COPD), codes 490–496; ischemic heart disease (IHD), codes 402–404 or 410–

414; and other respiratory causes, codes 480–486. We used the entity and record codes to determine all deaths in which lung cancer may have played a part (i.e., as either an underlying or a contributing cause) and included these possible lung cancer deaths in our sensitivity analyses.

Other Data

Baseline characteristics, including lung cancer risk factors, were obtained from the baseline questionnaire filled out by participants at entry in the trial. The date and cause of death for the 1977 deceased individuals whose records were not sent to the NDI were obtained from death-certificate information recorded in MLP records. For this group, the number of deaths contributed to the current analyses varies slightly from the number of deaths reported in previously published MLP analyses (9,14) for two reasons. Previous numbers of deaths were based on the findings of a mortality review board rather than on death-certificate information, and data on some deaths that occurred between July 1, 1983, and May 5, 1984, were available in MLP files. Data on compliance and contamination, as well as on stage of disease, tumor histology, diagnosis, and treatment, for the 366 participants diagnosed with lung cancer prior to July 1, 1983, also were obtained from MLP records.

Statistical Analysis

Lung cancer mortality rates were calculated by dividing the number of lung cancer deaths by person-years at risk of death. Person-years equaled the time from study entry to death or December 31, 1996, whichever was relevant. Rates also were generated for all-cause mortality as well as for death from COPD, IHD, other respiratory problems, cancers not of the lung, and all causes other than those listed. Ninety-five percent confidence intervals (CIs) for the mortality rates (15) are reported. All P values reflect two-sided tests.

By use of PROC PHREG of the SAS software package (SAS Institute, Inc., Cary, NC) (16), we examined whether adjustment for other lung cancer risk factors (age at study entry, pack-years smoked, exposure to lung carcinogens other than tobacco [asbestos, arsenic, nickel, chromium, or radioactive material], or history of pulmonary illness) in proportional hazards models would affect the lung cancer mortality hazard ratio for intervention versus usual care. The assumption of proportional hazards was checked by graphing the log of the negative log of the survival function versus the log of follow-up time for each value of the risk factor. Since the plotted lines were roughly parallel across time, we considered the assumption of proportional hazards to be valid.

PROC PHREG also was used to assess effect modification. We fit a proportional hazards model with a term for study arm, the potential effect modifier, and the interaction of the two. If the beta coefficient for the interaction was statistically significantly different ($P < .05$) from zero and the CI for the interaction risk ratio included 2.0 (suggesting a potential twofold difference in the screening risk ratios across strata of the potential effect modifier), effect modification was said to exist. In the case of polytomous variables, effect modification was said to exist if these criteria held for any of the interaction terms.

By use of SAS's PROC LIFETEST (17), we calculated the median and 5-year survival for participants diagnosed with lung cancer before July 1, 1983. Kaplan-Meier estimates were used. Log-rank tests were employed to compare survival curves. We could not include participants diagnosed with lung cancer after July 1, 1983, because the extended follow-up addressed mortality but not incidence.

RESULTS

The vital status of 2669 MLP participants on December 31, 1996, was available from the Mayo Clinic's records. Information on 6523 of the remaining 6542 participants (19 persons refused additional follow-up and were required by Minnesota statute to be excluded from all analyses) was sent to the NDI. Of the 6523 records, no match was obtained for 1590, and no true match was obtained for 1972. These men were, therefore, assumed to be alive on December 31, 1996. True matches were identified for 2961 participants. Table 1 shows vital status assignment by study arm. The two study arms were balanced on baseline characteristics, including age, smoking habits, exposure to non-tobacco lung carcinogens, and history of pulmonary illnesses (18).

Table 1. Assignment of vital status for Mayo Lung Project participants

	Intervention arm, No. (%)	Usual-care arm, No. (%)	Total, No. (%)
Total participants	4618	4593	9211
Vital status already known			
Dead as of around July 1, 1983	994 (22)	983 (21)	1977 (21)
Alive on December 31, 1996	372 (8)	320 (7)	692 (8)
Records sent to National Death Index			
Returned with no match*	796 (17)	794 (17)	1590 (17)
Returned, false match*	946 (20)	1026 (22)	1972 (21)
Returned, true match†	1499 (32)	1462 (32)	2961 (32)
Refusals	11 (0)	8 (0)	19 (0)
Total participants, without refusals	4607	4585	9192

*Assumed to be alive.

†Assumed to be dead.

Mortality

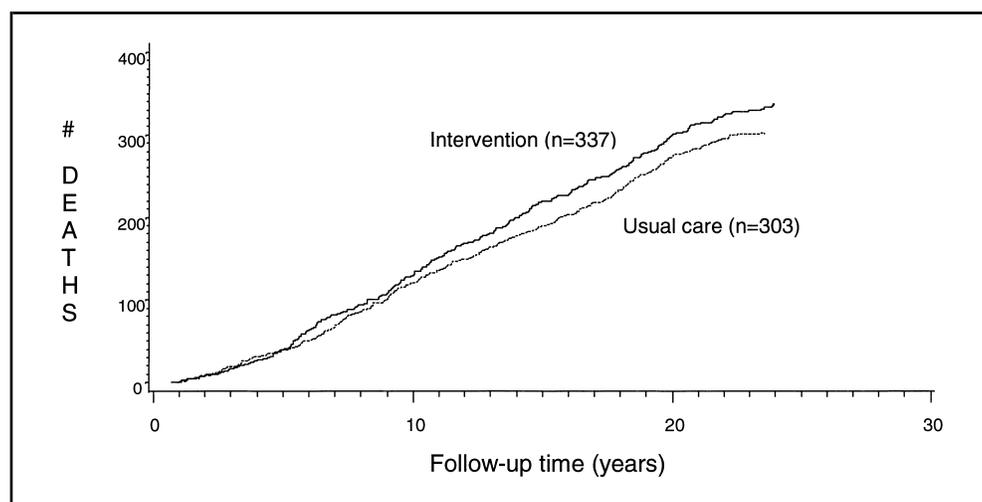
Our NDI search and our matching algorithm identified 396 lung cancer deaths, bringing the lung cancer death totals to 337 among participants in the intervention arm (76 760.7 person-years) and 303 among participants in the usual-care arm (76 772.4 person-years) as of December 31, 1996 (Fig. 1; Table 2). The median follow-up time was 20.5 years. The lung cancer

mortality rate was 4.4 deaths per 1000 person-years (95% CI = 3.9–4.9) in the intervention arm and 3.9 deaths per 1000 person-years (95% CI = 3.5–4.4) in the usual-care arm; the two rates were not statistically significantly different ($P = .09$; 95% CI for the observed 13% increase in lung cancer mortality in the intervention arm: -5% to 30%). All-cause mortality and mortalities from other cancers, COPD, IHD, and respiratory ailments other than COPD and lung cancer also did not differ by study arm (Table 2).

The finding of similar lung cancer mortalities in both study arms remained after adjustment for four established lung cancer risk factors (age, smoking [measured as pack-years smoked], exposure to non-tobacco lung carcinogens, and history of pulmonary illness) (unadjusted hazard ratio [HR] = 1.1 [95% CI = 1.0–1.3]; adjusted HR = 1.1 [95% CI = 1.0–1.3]). Furthermore, when assessed individually, neither age (HR = 1.0 for <55 years, HR = 1.1 for 55–64 years, and HR = 1.6 for ≥ 65 years), amount smoked (HRs = 1.1 for <50 pack-years, 50–99 pack-years, and ≥ 100 pack-years), exposure to non-tobacco lung carcinogens (HRs = 1.1 for both never and ever), nor history of other pulmonary illness (HR = 1.2 for never and HR = 1.0 for ever) acted as effect modifiers.

Of 933 participants noted in the Mayo Clinic registration system to have died after July 1, 1983 (with no available cause of death), our algorithm correctly identified 91%. Of these, 89% had exact agreement on date of death and 98% had agreement within 30 days.

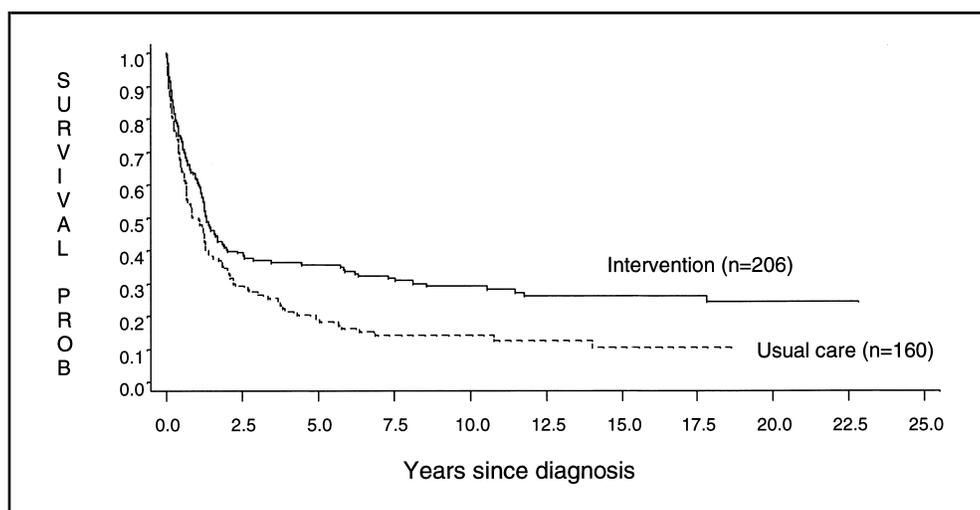
Fig. 1. Cumulative lung cancer deaths by study arm. Sample size was 4607 in the intervention arm (solid line) and 4585 in the usual-care arm (dashed line). Numbers in parentheses are the numbers of lung cancer deaths as of December 31, 1996. The National Death Index was used, as described in the text, to follow-up Mayo Lung Project participants for whom vital status on December 31, 1996, was unknown.

**Table 2.** Mortality in the Mayo Lung Project, as of December 31, 1996

Cause of death*	Deaths, No. (%)		Mortality rate (95% confidence interval) per 1000 person-years	
	Intervention arm (n = 4607)	Usual-care arm (n = 4585)	Intervention arm (76 760.7 person-years)	Usual-care arm (76 772.4 person-years)
Lung cancer	337 (7)	303 (7)	4.4 (3.9–4.9)	3.9 (3.5–4.4)
Causes other than lung cancer	2148 (47)	2133 (47)	28.0 (26.8–29.2)	27.8 (26.6–29.0)
Cancers other than lung cancer	403 (9)	391 (9)	5.3 (4.8–5.8)	5.1 (4.6–5.6)
Chronic obstructive pulmonary disease	156 (3)	149 (3)	2.0 (1.7–2.4)	1.9 (1.6–2.3)
Ischemic heart disease	816 (18)	816 (18)	10.6 (9.9–11.4)	10.6 (9.9–11.4)
Other respiratory causes	60 (1)	44 (1)	0.8 (0.6–1.0)	0.6 (0.4–0.8)
Other	712 (15)	733 (16)	9.3 (8.6–10.0)	9.5 (8.9–10.3)
All causes	2493 (54)	2445 (53)	32.5 (31.2–33.8)	31.8 (30.6–33.1)

*Seventeen participants (eight in the intervention arm and nine in the usual-care arm) had unknown causes of death.

Fig. 2. Survival of participants diagnosed with lung cancer prior to July 1, 1983, by study arm, analyzed by the Kaplan–Meier method. **Numbers in parentheses** are the numbers of participants diagnosed with lung cancer prior to July 1, 1983. The event of interest was death from lung cancer; participants who died of other causes were censored at their dates of death. The number of lung cancer deaths was 133 in the intervention arm and 119 in the usual-care arm. Two-sided $P = .0039$ (log-rank test). At 1 year after diagnosis, lung cancer survival was 61.7% (95% confidence interval [CI] = 54.8%–68.6%) among participants in the intervention arm and 50.1% (95% CI = 42.0%–58.3%) among those in the usual-care arm; 113 participants in the intervention arm and 70 participants in the usual-care arm were at risk. At 5 years after diagnosis, lung cancer survival was 35.6% (95% CI = 28.6%–42.6%) among participants in the intervention arm and 18.5% (95% CI = 11.5%–25.5%) among those in the usual-care arm; 53 participants in the intervention arm and 18 in the usual-care arm were at risk. At 10 years after diagnosis, lung cancer survival was 29.2% (95% CI = 22.2%–35.1%) among participants in the intervention arm and 14.2% (95% CI = 7.6%–20.7%) among those in the usual-care arm; 32 participants in the intervention arm and 10 in the usual-care arm were at risk. At 15 years after diagnosis, lung cancer survival was 26.2% (95% CI = 19.2%–33.2%) among participants in the intervention arm and 10.6% (95% CI = 4.0%–17.2%) among those in the usual-care arm; 16 participants in the intervention arm and three participants in the usual-care arm were at risk.



The recommended NDI algorithm identified 299 lung cancer deaths, for a total of 289 and 254 lung cancer deaths in the intervention and usual-care arms, respectively. Using this algorithm, lung cancer mortality was 3.6 (95% CI = 3.2–4.1) in the intervention arm and 3.2 (95% CI = 2.8–3.6) in the usual-care arm. This algorithm correctly identified 80% of the aforementioned 933 participants known to have died after July 1, 1983.

Case Survival

Lung cancer survival among participants diagnosed prior to July 1, 1983, differed by study arm ($P = .0039$; Fig. 2; Table 3).

Table 3. Survival of Mayo Lung Project participants diagnosed with lung cancer prior to July 1, 1983, as of December 31, 1996*

	Intervention arm	Usual-care arm
Participants diagnosed prior to July 1, 1983, No. (%)	206	106
Died prior to December 31, 1996		
COD: lung cancer	133 (65)	119 (74)
COD: other than lung cancer	57 (28)	37 (23)
COD: all causes	190 (92)†	156 (98)†
Alive on December 31, 1996	16 (8)	4 (3)†
Median survival, y		
COD: lung cancer	1.3	0.9
COD: other than lung cancer	10.8	6.9
COD: all causes	1.2	0.7
5-y survival, %		
COD: lung cancer	35	19
COD: other than lung cancer	72	59
COD: all causes	26	11
Two-sided P value‡		
COD: lung cancer		.0039
COD: other than lung cancer		.13
COD: all causes		.0012

*COD = cause of death.

†Percentages do not add up exactly because of rounding.

‡Comparison of survival curves for intervention versus usual care, log-rank test.

The median survival time was 1.3 years for those in the intervention arm compared with 0.9 years for those in the usual-care arm. For resected early-stage (T1 or T2) (4) disease, the lung cancer median survival for patients in the intervention arm was 16.0 years compared with 5.0 years for those in the usual-care arm, but the difference was not statistically significant ($P = .16$; Fig. 3; Table 4). For late-stage (T3 or T4) or unresected disease, lung cancer survival rates for participants in both study arms were the same. Percentages of resection in the study arms were almost identical for individuals with early-stage disease (81% in the intervention arm versus 80% in the usual-care arm) but varied somewhat for individuals with late-stage disease (25% in the intervention arm versus 15% in the usual-care arm).

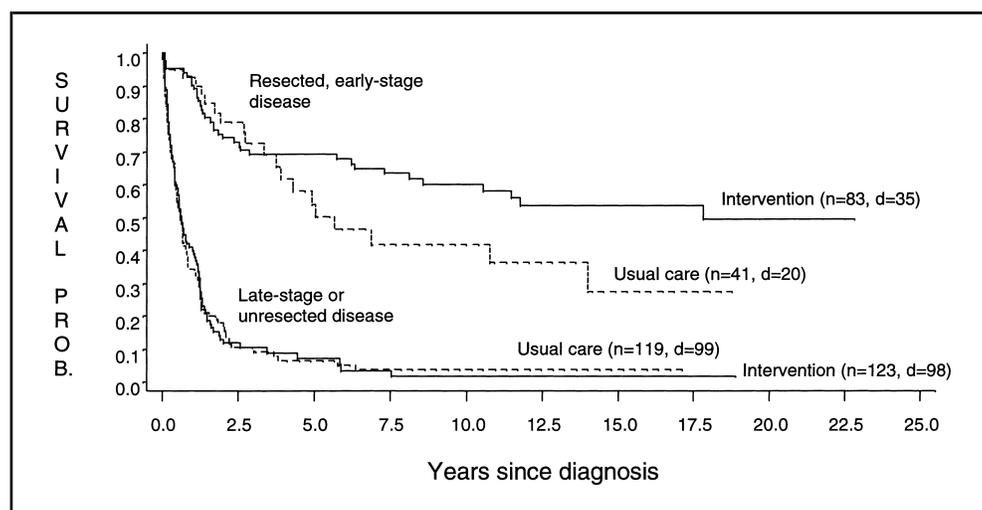
Among participants diagnosed with squamous cell histology, lung cancer survival was better in the intervention arm, and the difference was statistically significant ($P = .032$; Table 5). For small- and large-cell histologies, differences in lung cancer survival were observed, but they were not statistically significant. In the case of adenocarcinoma, the two arms had similar lung cancer survival.

DISCUSSION

Extended follow-up of the MLP participants indicates that an intense 6-year screening regimen of chest x-ray and sputum cytology did not reduce lung cancer mortality, the most important and meaningful end point in trials of mass screening. However, individuals in the intervention arm who were diagnosed with lung cancer prior to July 1, 1983, had better survival than their counterparts in the usual-care arm. These findings reflect 25 years of study time and more than 150 000 person-years of follow-up.

Use of the NDI provided us with an efficient and inexpensive method to determine which participants in the MLP had died and of what cause. Rather than using the recommended NDI algorithm, we developed our own algorithm, one that employed more liberal matching criteria for participants for whom we had no Social Security number. It is likely that neither our algorithm

Fig. 3. Survival of participants diagnosed with lung cancer prior to July 1, 1983, by study arm and stage of disease, analyzed by the Kaplan–Meier method. **Numbers in parentheses** are the numbers of participants who were diagnosed with lung cancer (n) prior to July 1, 1983, and the number of lung cancer deaths (d) that occurred prior to December 31, 1996. The event of interest was death from lung cancer; participants who died of other causes were censored at their dates of death. The numbers of lung cancer deaths were 35 (intervention arm; resected, early-stage disease), 20 (usual-care arm; resected, early-stage disease), 98 (intervention arm; late-stage or unresected disease), and 99 (usual-care arm; late-stage or unresected disease). Two-sided *P* values were .16 (resected early-stage disease, log-rank test) and .92 (late-stage or unresected disease, log-rank test).



For resected early-stage disease at 1 year after diagnosis, lung cancer survival and number at risk were 90.3% (95% confidence interval [CI] = 83.9%–96.7%) and 73, respectively, in the intervention arm and 92.5% (95% CI = 84.3%–100.0%) and 36, respectively, in the usual-care arm; at 5 years after diagnosis, lung cancer survival and number at risk were 69.1% (95% CI = 57.4%–77.8%) and 49, respectively, in the intervention arm and 54.1% (95% CI = 36.2%–71.9%) and 14, respectively, in the usual-care arm; at 10 years after diagnosis, lung cancer survival and number at risk were 59.9% (95% CI = 48.7%–71.0%) and 31, respectively, in the intervention arm and 41.6% (95% CI = 23.2%–60.1%) and eight, respectively, in the usual-care arm. For late-stage or unresected disease at 1 year after diagnosis, lung cancer survival and number at risk were 39.9% (95% CI = 30.7%–49.2%) and 40, respectively, in the intervention arm and 34.4% (95% CI = 25.4%–43.6%) and 34, respectively, in the usual-care arm; at 5 years after diagnosis, lung cancer survival and number at risk were 7.2% (95% CI = 1.4%–13.0%) and four, respectively, in the intervention arm and 6.5% (95% CI = 1.3%–11.7%) and five, respectively, in the usual-care arm.

Table 4. Survival of Mayo Lung Project participants diagnosed with lung cancer prior to July 1, 1983, as of December 31, 1996, by stage at diagnosis*

	Stage			
	Resected early-stage disease		Late-stage or unresected disease	
	Intervention arm	Usual-care arm	Intervention arm	Usual-care arm
Participants diagnosed prior to July 1, 1983, No. (%)	83	41	123	119
Died prior to December 31, 1996				
COD: lung cancer	35 (42)	20 (49)	98 (80)	99 (83)
COD: all causes	68 (82)	38 (93)	122 (99)	118 (99)
Alive on December 31, 1996	15 (18)	3 (7)	1 (1)	1 (1)
Median survival, y				
COD: lung cancer	16.0	5.0	0.6	0.6
COD: all causes	7.2	3.1	0.5	0.5
5-y survival, %				
COD: lung cancer	68	50	6	6
COD: all causes	59	32	3	4
Two-sided <i>P</i> value†				
COD: lung cancer	.16		.92	
COD: all causes	.0099		.79	

*COD = cause of death.

†Comparison of survival curves for intervention versus usual care, log-rank test.

nor the recommended NDI algorithm pinpointed the true number of deaths; however, regardless of which algorithm we used, no reduction in lung cancer mortality for the intervention arm was observed. Support for the effectiveness of our algorithm is given by the fact that examination by calendar year of the all-cause mortality hazard revealed a steady increase in that hazard over time (data not shown), suggesting that we did not miss a substantial number of deaths. Also, our algorithm correctly identified as deceased 91% of the participants noted in the Mayo Clinic registration system to have died after July 1, 1983; for those individuals, the match on the date of death was excellent.

Our lung cancer mortality rates were based on individuals with lung cancer noted as the underlying cause of death, but inclusion of 81 individuals (41 in the intervention arm and 40 in

the usual-care arm) with lung cancer as a contributing cause of death did not change the results (data not shown).

Could a true reduction in lung cancer mortality have been missed in the MLP? It has been suggested (8) that the finding of no mortality benefit could be due to population heterogeneity; i.e., individuals in the two randomly assigned study arms had different distributions of risk and prognostic factors for lung cancer. However, adjustment for four lung cancer risk factors did not change results in either this analysis or in analyses with follow-up through July 1, 1983 (18). Unknown or unmeasured risk or prognostic factors, particularly genetic factors, could cause a mortality benefit to be missed, but it is fully expected that such factors were balanced by the randomization. That such balance is highly likely follows not only from randomization and

Table 5. Survival of Mayo Lung Project participants diagnosed with lung cancer prior to July 1, 1983, as of December 31, 1996, by tumor histology*

	Tumor histology†							
	Squamous		Adenocarcinoma		Large cell		Small cell	
	Intervention arm	Usual-care arm						
Participants diagnosed prior to July 1, 1983, No. (%)	68	51	59	38	29	24	48	45
Died prior to December 31, 1996								
COD: lung cancer	35 (51)	33 (65)	39 (66)	25 (66)	19 (66)	20 (83)	40 (83)	40 (89)
COD: all causes	61 (90)	49 (96)	55 (93)	37 (97)	26 (90)	24 (100)	47 (98)	44 (98)
Alive on December 31, 1996	7 (10)	2 (4)	4 (7)	1 (3)	3 (10)	0 (0)	1 (2)	1 (2)
Median survival, y								
COD: lung cancer	7.5	1.3	1.3	1.8	1.3	0.7	1.0	0.7
COD: all causes	2.6	1.2	1.1	0.6	1.1	0.7	0.7	0.5
5-y survival, %								
COD: lung cancer	55	26	30	33	35	9	6	6
COD: all causes	40	14	21	21	29	6	4	2
Two-sided <i>P</i> value‡								
COD: lung cancer		.032		.97		.085		.28
COD: all causes		.0099		.77		.061		.41

*COD = cause of death.

†Four participants (two in the intervention arm and two in the usual-care arm) had other histologies.

‡Comparison of survival curves for intervention versus usual care, log-rank test.

balance on baseline factors (18) but also from balance on causes of death other than lung cancer (Table 3).

Contamination, which occurs when participants randomly assigned to the nonintervention arm receive screening, can mask a mortality benefit. Participants in the usual-care arm in the MLP did not participate in any formal screening regimen but were advised at the beginning of the Project to obtain an annual chest x-ray and sputum cytology. On a special questionnaire that was mailed to the participants near the end of the Project, 72% of usual-care participants reported having had a chest x-ray between 1972 and 1984. Because additional details for these individuals (in particular, the total number of chest x-rays and indication for the procedure), are unknown, we cannot assess how alike the participants in the usual-care and intervention arms were with regard to their screening practices. Nevertheless, it is hard to believe that a substantial number of usual-care participants would have received screening x-rays or sputum cytology as frequently as did the participants in the intervention arm. Extreme contamination, had it existed, could have masked a true screening mortality benefit. However, if that were the case, we should not have observed an excess of lung cancer diagnoses and improved lung cancer survival in the intervention arm. We know of no realistic scenario in which contamination would mask a true mortality benefit while not masking an accompanying incidence and case survival difference.

Other factors that could have resulted in the masking of a true mortality benefit include misclassification of lung cancer as the underlying cause of death, ineligibility of persons diagnosed with lung cancer during the prevalence screen, lack of statistical power, and dilution of the screening effect due to long-term follow-up after cessation of the screening regimen. The last of these can be dismissed easily, since, at no point in the MLP, was mortality meaningfully better in the intervention arm (Fig. 1).

Biased reporting of lung cancer as the underlying cause of death on the death certificate could have occurred when the underlying cause of death was unclear. Because lung cancer,

which is generally considered to be fatal in most instances, had been diagnosed, physicians might record that disease as the underlying cause of death if they were uncertain. Alternatively, tumors diagnosed as primary lung cancers might actually have been metastases of occult tumors of other organs. In both instances, the intervention arm would experience more of these misreports because more lung cancers were detected as a result of screening, and it is likely that our results are affected somewhat by this so-called sticking diagnosis bias.

Another sort of cause-of-death misclassification could exist if deaths truly due to lung cancer were erroneously classified as deaths not due to lung cancer. The degree of misclassification can help determine whether a mortality benefit was missed, but only limited data on misclassification are available. For deaths prior to July 1, 1983, assigned cause of death is available from both an expert review panel and from a death certificate. If we assume that the expert panel classified true lung cancer deaths correctly, the death certificates overestimated the true number of lung cancer deaths by about 5% in the intervention arm and underestimated that figure by about 2% in the usual-care arm. We used those percentages to adjust the number of lung cancer deaths during the extended follow-up period that we identified by use of death-certificate information. The numbers changed to 321 lung cancer deaths in the intervention arm and 308 in the usual-care arm, suggesting that no masking of a lung cancer mortality benefit occurred because of disagreement between death certificates and an expert review panel.

Of course, the expert panel may not have always been correct in its classification (in part because of sticking diagnosis bias). No additional data are available on the degree of misclassification for the expert panel. Using simple algebra, we calculated that a 16% overestimate in the number of lung cancer deaths for the intervention arm would have been necessary to mask a true 10% reduction in lung cancer mortality. Similarly, overestimates of 30% and 49% would be necessary to mask 20% and 30% reductions, respectively. These calculations assume that the adjusted number of lung cancer deaths in the

usual-care arm (308 deaths) is accurate, but any scenario that allows for an overestimate of lung cancer deaths (as would be the case with sticking diagnosis bias) would produce an even greater degree of misclassification in the intervention arm. We believe that the degrees of misclassification necessary to mask at least a 20% reduction in lung cancer mortality are highly improbable; therefore, we conclude that cause-of-death misclassification could have obscured a small mortality benefit from screening but would not have obscured a substantial benefit.

Ninety-one individuals were diagnosed with lung cancer as a result of the MLP eligibility screen and, as such, were excluded from the remainder of the trial. Exclusion of these individuals, who are often referred to as the prevalence cases of the MLP, could have masked a mortality benefit if the subset of lung cancer lesions amenable to a screening benefit were eliminated from the pool of lesions that was available for detection in the incidence-screening phase of the Project. It appears, however, that the lesions identified during the prevalence screen had prognoses similar to those diagnosed afterwards; the 5-year survival for the 91 persons diagnosed at the prevalence screen was close to 40% (4), while the 5-year survival for the 206 intervention-arm participants diagnosed prior to July 1, 1983, was 35% (Table 3). This comparison suggests that the lesions that were detected during the prevalence screen were as amenable to a screening benefit as were those that were detected after randomization.

Lack of statistical power has been noted repeatedly as a shortcoming of the MLP and one that may be responsible for the observation of no mortality benefit in the study. The Project was designed with 95% power to detect a 50% reduction in lung cancer mortality in the intervention arm but had only 48% power to detect a more realistic (according to present thinking) 20% reduction (4). Our extended follow-up, however, raised power to 88%, close to the generally accepted level of 90% (19). The 95% CI associated with our observed 13% increase in the lung cancer mortality rate for the intervention arm indicates that values compatible with the experience of the MLP range from a 5% decrease to a 30% increase in lung cancer mortality with screening; a 95% CI that makes allowance for 25% noncompliance in the intervention arm extends the lower bound to a 6% decrease. These numbers argue against the existence of a major reduction in lung cancer mortality but indicate that a small reduction could have been missed.

Although no reduction in lung cancer mortality was seen in the intervention arm of the MLP, a case-survival difference was observed. It has been suggested that this seemingly better survival in the intervention arm actually indicates that a true lung cancer mortality benefit exists but was somehow missed in the study (8). This argument fails to acknowledge that improved case survival can reflect either a mortality benefit from earlier detection or any one of several biases produced by screening. Since we observed no mortality benefit, we conclude that screening biases are responsible. One possibility is a length bias that would yield a transient survival benefit that disappears with complete lung cancer incidence follow-up and the occurrence of catch-up in the usual-care arm (20).¹ This possibility would require better survival among participants diagnosed after July 1, 1983, in the usual-care arm compared with those in the intervention arm. While this type of length bias is possible, we do not consider it as likely as overdiagnosis bias, the most ex-

treme form of length bias. By overdiagnosis, we are referring to detection, through screening, of cancers that never would have progressed to clinical disease during a person's lifetime, rather than the situation in which lesions with nonmalignant morphology are erroneously classified as malignant. The lesions associated with overdiagnosis have malignant morphologies but would never have been diagnosed in the absence of screening (21). A study of lung cancer aggressiveness (22) supports the existence of these indolent lesions, since some screen-detected tumors with malignant morphology had evidence of low biologic aggressiveness. An analysis by Sobue et al. (23) that indicated that very few untreated persons with early-stage disease survive long after radiographic lung cancer detection does not present strong evidence against the existence of overdiagnosis, since staging was determined clinically and that is a much less reliable method than surgical staging. It appears that both screening modalities used in the MLP may identify indolent lesions; survival among screened individuals diagnosed with resected early-stage disease was comparably long when disease was detected either by chest x-ray alone or by sputum cytology alone (data not shown).

The presence of an apparent survival difference absent a mortality benefit also can occur if screening results in earlier diagnosis of disease but no postponement of death (lead-time bias). However, we believe that not to be the case in the MLP, since a survival analysis that examined mortality of participants diagnosed with lung cancer beginning at the time of randomization rather than the date of diagnosis (to eliminate lead-time bias) still indicated a survival difference. Another possible explanation for the apparent survival benefit we observed, that the intervention arm received superior treatment, is unlikely, since the percentages of resection for early-stage disease were similar in the two study arms. Lung cancer incidence follow-up of the MLP participants would determine whether length bias or overdiagnosis bias is responsible for the survival difference, but such data are not available at this time.

With regard to our interpretations of the mortality and case-survival results, some researchers, a number of whom are clinicians, have stated that the existence of indolent lung cancer lesions is not possible. While clinicians clearly bring to the screening arena a valuable perspective on the progression of lung cancer, these observations, while often based on very important and telling clinical experiences, are unfortunately limited in their generalizability, especially in the presence of conflicting data collected in the context of a controlled experiment.

Other lines of reasoning regarding why overdiagnosis is implausible have been provided but are flawed. Strauss et al. (8) state that "the over-diagnosis hypothesis is counter to virtually all known data on the natural history and biologic behavior of lung cancer." Such conclusions, however, are based overwhelmingly on cancers that have been diagnosed primarily because of symptoms, not those that never present. High mortality, as is the case in lung cancer, does not imply that all lung cancer lesions are lethal; it implies only that known lesions are lethal. Strauss et al. (8) cite as additional evidence an autopsy study (24) of lung cancer that identified 26 "surprise" instances of lung cancer in 2886 decedents. The authors suggest that, because "58% [of the decedents] had regional and/or distant metastases and 37% had been too ill for medical evaluation, the majority died of lung cancer," reasoning with which we do not agree. That five (19%) of the tumors were classified as stage 0 and six (23%) were

classified as stage I indicates that a substantial percentage of identified lesions might have had limited clinical relevance. Autopsies, particularly those that are not performed for the sole purpose of identifying occult lung cancer lesions [as was the case in (24)], may be limited in their ability to identify small lesions, since the lungs may be too voluminous for thorough examination. The autopsy series discussed by Strauss et al. (8) may be further limited, since it did not comprise a random sample of deaths. Finally, neoplastic lesions with limited clinical relevance almost certainly exist in prostate (25) and breast (26) cancers; their existence is known primarily, if not entirely, because of screening programs. That indolent lesions in the lung could not exist or would not be identified in mass screening programs seems to be counter to what we know about the development and progression of neoplastic lesions in general.

We have presented evidence that we believe argues strongly against a large reduction in lung cancer mortality with an intense regimen of chest x-ray and sputum cytology screening and, more important, for the possible existence of lung cancer lesions with limited clinical relevance. Although the finding of improved survival in the intervention arm appears attractive, it does not indicate that lung cancer screening saved lives in the MLP. The true worth of any cancer screening modality lies in whether it can reduce mortality. Case patient survival is not a reliable measure of efficacy, especially in the setting of lung cancer where lead-time and overdiagnosis biases are probable (7,27–29).

Since the end of the MLP, advances have been made in both radiographic equipment and in the treatment regimens that are available for lung cancer patients, especially those with tumors of small-cell histology. It is plausible that these improvements are sufficient to result in a reduction of lung cancer mortality with the use of modern chest x-ray as a screening modality. The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (30) is presently collecting data to assess whether mass screening that uses today's chest x-ray technology is effective. Screened participants are receiving chest x-rays annually for either 3 or 4 years, while participants in the control arm are receiving no recommendation about screening. The sample size, 148 000 participants, allows for 90% power to detect a 10% reduction in lung cancer mortality in the presence of 40% contamination and 15% noncompliance. Time will tell whether a reduction in lung cancer mortality can be obtained with an annual chest x-ray.

Spiral computed tomography (CT) has been shown recently to identify early lung cancer lesions (31) and may ultimately prove to be a more useful screening modality than chest x-ray. However, if lung cancer lesions with little or no clinical relevance truly exist, spiral CT will identify them at a rate even higher than that of chest x-ray. Before spiral CT is accepted into medical practice, it is critical to determine whether this promising new screening modality ultimately does more good than harm in a randomized, controlled clinical trial with lung cancer mortality as the end point.

REFERENCES

- (1) Wagner H, Ruckdeschel JC. Lung cancer. In: Reintgen D, Clark R, editors. Cancer screening. St. Louis (MO): Mosby; 1996. p. 118–49.
- (2) Black WC. Lung cancer. In: Kramer BS, Gohagan JK, Prorok PC, editors. Cancer screening: theory and practice. New York (NY): Marcel Dekker; 1999. p. 327–77.
- (3) Cancer facts and figures—1999. Atlanta (GA): American Cancer Society; 1999.
- (4) Fontana RS, Sanderson DR, Woolner LB, Taylor WF, Miller WE, Muhm JR, et al. Screening for lung cancer: a critique of the Mayo Lung Project. *Cancer* 1991;67(4 Suppl):1155–64.
- (5) Wolpaw DR. Early detection in lung cancer. Case finding and screening. *Med Clin North Am* 1996;80:63–82.
- (6) Flehinger BJ, Kimmel M, Polyak T, Melamed MR. Screening for lung cancer. The Mayo Lung Project revisited. *Cancer* 1993;72:1573–80.
- (7) Morrison AS. Screening in chronic disease. 2nd ed. New York (NY): Oxford University Press; 1992.
- (8) Strauss GM, Gleason RE, Sugarbaker DJ. Screening for lung cancer. Another look; a different view. *Chest* 1997;111:754–68.
- (9) Fontana RS, Sanderson DR, Woolner LB, Taylor WF, Miller WE, Muhm JR. Lung cancer screening: the Mayo Program. *J Occup Med* 1986;28:746–50.
- (10) Fontana RS, Sanderson DR, Taylor WF, Woolner LB, Miller WE, Muhm JR, et al. Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study. *Am Rev Respir Dis* 1984;130:561–5.
- (11) National Death Index User's Manual. Hyattsville (MD): National Center for Health Statistics; 1997.
- (12) Horm J. Assignment of probabilistic scores to National Death Index Record matches. Hyattsville (MD): National Center for Health Statistics; 1996.
- (13) National Death Index Plus: coded causes of death. Supplement to the National Death Index User's Manual. Hyattsville (MD): National Center for Health Statistics; 1997.
- (14) Fontana RS, Sanderson DR. Screening for lung cancer: a progress report. In: Mountain CF, Carr DT, editors. Lung cancer: current status and prospects for the future. Austin (TX): University of Texas Press; 1986. p. 51–7.
- (15) Lawless JF. Statistical models and methods for lifetime data. New York (NY): John Wiley & Sons; 1982.
- (16) SAS/STAT software: changes and enhancements through release 6.11. Cary (NC): SAS Institute, Inc.; 1996.
- (17) SAS/STAT user's guide. Version 6. Vol 2. 4th ed. Cary (NC): SAS Institute, Inc.; 1990.
- (18) Marcus PM, Prorok PC. Reanalysis of the Mayo Lung Project data: the impact of confounding and effect modification. *J Med Screen* 1999;6:47–9.
- (19) Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics* 1983;39:499–503.
- (20) Connor RJ, Prorok PC. Issues in the mortality analysis of randomized controlled trials of cancer screening. *Control Clin Trials* 1994;15:81–99.
- (21) Prorok PC, Kramer BS, Gohagan JK. Screening theory and study design: the basics. In: Kramer BS, Gohagan JK, Prorok PC, editors. Cancer screening: theory and practice. New York (NY): Marcel Dekker; 1999. p. 29–53.
- (22) Hakama M, Holli K, Visakorpi T, Pekola M, Kallioniemi OP. Low biological aggressiveness of screen-detected lung cancers may indicate overdiagnosis. *Int J Cancer* 1996;66:6–10.
- (23) Sobue T, Suzuki T, Matsuda M, Kuroishi T, Ikeda S, Naruke T. Survival for clinical stage I lung cancer not surgically treated. Comparison between screen-detected and symptom-detected cases. The Japanese Lung Cancer Screening Research Group. *Cancer* 1992;69:685–92.
- (24) McFarlane MJ, Feinstein AR, Wells CK, Chan CK. The 'epidemiologic necropsy'. Unexpected detections, demographic selections, and changing rates of lung cancer. *JAMA* 1987;258:331–8.
- (25) Zappa M, Ciatto S, Bonardi R, Mazzotta A. Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence Screening Pilot Study. *Ann Oncol* 1998;9:1297–300.
- (26) Ernster VL, Barclay J. Increases in ductal carcinoma *in situ* (DCIS) of the breast in relation to mammography: a dilemma. *J Natl Cancer Inst Monogr* 1997;22:151–6.
- (27) Cole P, Morrison AS. Basic issues in population screening for cancer. *J Natl Cancer Inst* 1980;64:1263–72.
- (28) Prorok PC, Hankey BF, Bundy BN. Concepts and problems in the evaluation of screening programs. *J Chronic Dis* 1981;34:159–71.

- (29) Prorok PC. Epidemiologic approach for cancer screening. Problems in design and analysis of trials. *Am J Pediatr Hematol Oncol* 1992;14:117–28.
- (30) Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Concept. Bethesda (MD): Division of Cancer Prevention and Control, National Cancer Institute, Project Review, January 31, 1991 (available upon request).
- (31) Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999;354:99–105.

NOTES

¹During screening, the intervention arm is likely to have a higher cumulative incidence rate. Once screening ceases, however, the cumulative incidence rates in the two groups should eventually equalize if overdiagnosis does not exist. Catch-up occurs when the cumulative cancer incidence rate in the nonintervention arm equals, or “catches up” with, that in the intervention arm.

Manuscript received November 24, 1999; revised March 7, 2000; accepted June 17, 2000.