Overdiagnosis: An Underrecognized Cause of Confusion and Harm in Cancer Screening

William C. Black

The Mayo Lung Project (MLP) was a National Cancer Institute-funded randomized clinical trial designed to determine the effectiveness of intensive screening with chest radiography and sputum cytology in comparison with usual care (1). The trial was begun in 1971 and was completed in 1983; when the average follow-up after the last scan was about 3 years. Although the 5-year survival for lung cancer was much higher in the screened group than in the control group, there was no difference in lung cancer mortality. This apparent discrepancy between survival and mortality along with an excess of 46 lung cancer cases in the screened group (206, as compared with 160 in the usual-care arm) has been the source of much controversy. Marcus et al. (2), in an attempt to resolve this controversy, used the National Death Index–Plus search to extend the follow-up of the MLP participants through 1996. The investigators report their findings in this issue of the Journal (2).

After more than 76,000 person-years of observation in each group, there was still no statistically significant difference in lung cancer mortality (4.4 deaths per 1000 person-years in the intervention group versus 3.9 deaths per 1000 person-years in the control group); the mortality rates from all other causes were virtually identical. The authors acknowledge the possibility of contamination, noting that many of the subjects in the control group did have chest radiographs during the intervention period; they point out, however, that it is not known what proportion of these radiographs were obtained for screening rather than for evaluating specific symptoms. Furthermore, they point out that the markedly higher 5-year survival and excess cases in the screened group indicate that this group did, in fact, undergo more intensive screening than did the control group. In addition, the investigators found no baseline differences in age, smoking habits, or other lung cancer risk factors in the two groups (3). Thus, the authors provide compelling evidence that a major reduction in lung cancer mortality was not missed because of insufficient follow-up, contamination, or faulty randomization (4).

Marcus et al. (2) reason that the apparent discrepancy between survival and mortality is due largely, if not completely, to some combination of lead time, length, and overdiagnosis biases. Furthermore, they demonstrate that, when lung cancer survival is measured from the time of diagnosis rather than from the time of randomization and thereby adjusted for lead time, the survival advantage in the screened group persists. Thus, by process of elimination, they conclude that the discrepancy between survival and mortality is mainly due to the tendency for screening to detect the more slowly progressive forms of a disease (length bias), some of which would not have become clinically significant (overdiagnosis bias). (An analysis of lung cancer incidence after the completion of the trial could help determine the relative contribution of these two biases, but National Death Index–Plus does not provide incidence data.) Although it is sometimes argued that the dismal prognosis for lung cancer is inconsistent with the overdiagnosis hypothesis (4), this reasoning is flawed because it confuses symptomatic cases of lung cancer with asymptomatic cases, which are detectable only through screening.

Overdiagnosis occurs with the detection of “pseudodisease” (5), a subclinical condition that would not have produced signs or symptoms before the individual died of other causes. In any screening program, some proportion of screen-detected cases will be pseudodisease simply because of competing mortality. In the MLP, a substantial proportion of screen-detected cases were probably pseudodisease for three reasons: 1) the mortality rate from all causes in smokers is high, about threefold that in non-smokers (6); 2) some squamous cell carcinomas detectable by sputum cytology are very small; and 3) some primary adenocarcinomas detectable by chest radiography grow very slowly (7).

It should be pointed out that pseudodisease is almost impossible to document in a living individual. When pseudodisease is treated, as it almost always is, long-term survival is attributed to the treatment and is labeled a cure. In the rare instances when it is not treated because of old age or other contraindication, pseudodisease cannot be confirmed as such while the patient is still alive because, by definition, it must remain asymptomatic until the patient dies of other causes. These problems with documentation probably explain why pseudodisease has received relatively little attention. However, autopsy studies provide irrefutable evidence that pseudodisease is abundant, both for cancer in general (8) and for lung cancer in particular. In a 30-year review of all adult autopsies on hospital deaths at the Yale New Haven Hospital (New Haven, CT) (9), about one in six lung cancers observed at autopsy had not been recognized before the death of the patient. In the 10 most recent years of the review, about 1% of the men had had previously unsuspected lung cancer, most cases of which were resectable and presumably asymptomatic. In a more recent study of smokers being considered for lung reduction surgery (10), unsuspected primary lung cancer was found by preoperative chest radiography in 2% of the patients. Thus, it is not unreasonable to expect 6 years of intensive screening to detect 46 cases of pseudodisease among 4618 high-risk subjects in the intensively screened group of the MLP.

Overdiagnosis can also occur with the detection of a nonma-
lignant condition that is misclassified as malignant, that is, a pathologic false-positive error. Although the authors specifically exclude this type of error from their definition of overdiagnosis, pathologic false-positive results probably occur not infrequently in cancer screening. Even under the microscope, the distinction between malignancy and inflammation (11) or hyperplasia (12) can sometimes be very subtle, and the pretest probability of malignancy is usually low in screen-eligible subjects. In the MLP, the subset of patients with squamous cell carcinomas detected by sputum cytology alone, who had a 5-year survival of 83% (1), probably included some instances of pathologic false-positive results as well as pseudodisease.

Overdiagnosis plays havoc with our understanding of cancer statistics. Because overdiagnosis effectively changes a healthy person into a diseased one, it causes overestimations of the sensitivity, specificity, and positive predictive value of screening tests and the incidence of disease (13). As the MLP and a recent analysis of Surveillance, Epidemiology, and End Results (SEER) data illustrate (14), overdiagnosis also markedly increases the length of survival, regardless of whether screening or associated treatments are actually effective. However, overdiagnosis does not reduce disease-specific mortality because treating subjects with pseudodisease does not help those who have real disease. Consequently, disease-specific mortality is the most valid end point for the evaluation of screening effectiveness.

For individuals who undergo cancer screening, overdiagnosis is also highly relevant because it is the most serious side effect. False-positive results, which have received much more attention, may cause the screenee to worry for months about having cancer and may lead to an invasive procedure, such as a percutaneous needle biopsy, in the case of lung cancer screening. In contrast, overdiagnosis gives the screenee a false diagnosis of cancer for life and leads to definitive treatment, such as a lobectomy in the case of lung cancer screening. However, the public is much less informed about overdiagnosis than false-positive results. In a recent nationwide survey of women (15), 99% of the respondents were aware of the possibility of false-positive results from mammography but only 6% were aware of either ductal carcinoma in situ by name or the fact that mammography could detect a form of “cancer” that often doesn’t progress.

One apparent paradox in the MLP is that the lung cancer mortality was 11% higher in the screened group than in the control group. Although this excess mortality could be explained by chance alone ($P = 0.18$, two-tailed Fisher’s exact test), overdiagnosis could also have contributed to it in both real and spurious ways. Unnecessary surgery for pseudodisease or a pathologic false-positive result could have led to some deaths in the screened group that were correctly attributed to lung cancer. (In a randomized clinical trial of screening, deaths from treatment should be attributed to the target disease.) In addition, overdiagnosis could have led to a spurious increase in lung cancer deaths in the screened group because of misclassification of the cause of death, i.e., “sticking diagnosis bias.” It is not difficult to imagine that a diagnosis of lung cancer could have influenced subsequent testing and reporting in a patient’s medical record, which, in turn, could have influenced the cause of death that appeared on the death certificate. Deaths from various causes could have been misclassified as deaths from lung cancer, but there are two good reasons to suspect that this misclassification involved metastatic adenocarcinoma, in particular. The primary site of this disease is often difficult to determine. Moreover, adenocarcinoma was the only cancer cell type for which patients in the screened group actually had a shorter median survival than those in the control group (2), despite the effects of lead-time, length, and overdiagnosis biases.

Misclassification because of sticking diagnosis bias would have biased the MLP results against screening. However, because the mortality rates for other causes of death were virtually identical in the two groups, an equally large misclassification of death in favor of screening, probably related to treatment complications, must have also been present. For example, some deaths due to surgery may have been attributed to diseases other than lung cancer, such as pneumonia. Regardless, the fact that the all-cause mortality rates were nearly identical (2% higher in the screened group) makes it extremely unlikely that any major net benefit of screening was missed.

The negative results of the MLP and the problem of overdiagnosis do not exclude the possibility that screening for lung cancer with low-dose helical computed tomography (CT) could be highly effective and worthwhile. CT is far more sensitive than chest radiography. In a recent screening study (16), CT detected almost six times as many stage I lung cancers as chest radiography, and most of these tumors were 1.0 cm or less in diameter. However, for this very reason, overdiagnosis and false-positive results could be a much bigger problem with chest CT than they were with chest radiography. In a recent study of small (<3 cm) surgically resected peripheral adenocarcinomas that had been followed by CT (17), tumor volume doubling times ranged from 42 to 1486 days and one half of the tumors had doubling times over 1 year. With a volume doubling time of 1 year, it takes nearly 8 years for a tumor to increase in diameter from 5 mm to 3 cm, plenty of time for the screenee to die of other causes.

Because the potential for overdiagnosis and false-positive results will be so great with helical CT, it is essential that there be some mechanism in the screening process to minimize these side effects, such as a mandatory observation period for small nodules. Randomized clinical trials should be performed, and all causes of mortality should be closely monitored to avoid missing a major benefit or harm from the screening process. Finally, a balanced presentation of the potential benefits and risks—including overdiagnosis—should be made to all prospective screenees to ensure that they can make an informed decision about being screened or enrolled in a randomized trial of screening.

REFERENCES


NOTES

1Editor’s note: SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

I thank H. G. Welch for his careful review of earlier versions of this manuscript and his many helpful suggestions.