Analysis of Mortality Data From Cancer Screening Studies

Looking in the Right Window

James A. Hanley

Background: Appropriate statistical analysis is required to measure the impact of early detection and treatment of cancer. The current practice of using cumulative mortality ignores both (1) the delay between early treatment and the time that any averted deaths would have otherwise occurred, and (2) cessation of these delayed benefits some time after screening is discontinued.

Methods: We use time-specific mortality density ratios to estimate the mortality ratio in the "window of influence." We then use time-specific incidence density ratios to assess the extent to which the removal of polyps and other possibly precancerous lesions detected by fecal occult blood screening reduces the incidence of colorectal cancer.

Results: Applied to a theoretical example, the current practice of using cumulative mortality substantially underestimates the reduction in mortality achievable by early treatment. If there is sufficient time for the full impact to emerge, time-specific mortality patterns provide a more accurate measure. In a previous analysis of the screening study, the reduction in cumulative incidence in the screened groups was just under 20%. In our reanalysis, yearly incidence density ratios indicate that had screening not been interrupted, there might have been a 40% reduction in incidence.

Conclusions: Time-specific mortality ratios provide a more sensitive measure of the effects of early detection and treatment. Measures based on cumulative mortality are diluted by inclusion of deaths that occur soon after the initiation of screening as well as deaths that occur too long after the cessation of screening.

(*Epidemiology* 2005;16: 786–790)

In the design of trials to assess the mortality reduction resulting from screening-induced early interventions against cancer, considerable care is taken to generate high-quality data. The statistical analyses of these data usually measure the reduction in cumulative mortality. Unfortunately, by mixing "irrelevant experience with the relevant experience,"[1] these analyses underestimate the impact of early intervention. We discuss a data analysis principle, long established but seldom practiced until recently,[1–3] and illustrate its sharpness by an unusual example.

The purpose of cancer screening is to detect and treat a lesion now that if left to present itself at a later date would prove fatal x years from now. If such early treatment is successful, the resulting "cure" will contribute to a deficit of mortality x years from now, ie, there will be fewer cancer deaths at that time. Deaths that are averted by today's early treatment, but that would not have been averted by later treatment, create a delayed shortfall that will be distributed within some future time window. Outside this window, cancer mortality statistics will resemble those in a nonscreened population.

Figure 1 shows the reductions in cancer deaths in a hypothetical situation in which screening is carried out for 10 years. For example, as a result of the screening activities in year 1, the earlier detection and associated earlier treatment averted 1 death that would otherwise have occurred in year 5, 2 that would have occurred in year 6, and so on (13 in all). As a result of the *several* years of screening, the total numbers of deaths that would otherwise have occurred in years 5, 6, 7, . . . are 1, 3, 6, ... The totals remain in steady state (13 averted deaths) in years 10 to 14. Because of the cessation of screening in year 10, the "deficits" diminish from years 15 onward; the last deficit is visible in year 19. In the absence of 10 years of screening, there would be no averted deaths. The curve in the bottom of the figure contrasts the mortality in the presence and absence of screening (assuming equal amounts of experience): the mortality rate ratio is 25/25 = 1.0 for years 1 to 4; it falls to 24/25 = 0.96 in year 5, to 22/25 = 0.88 in year 6, and so on. Using *cumulative* mortality up to years 10, 20, and 30 (30
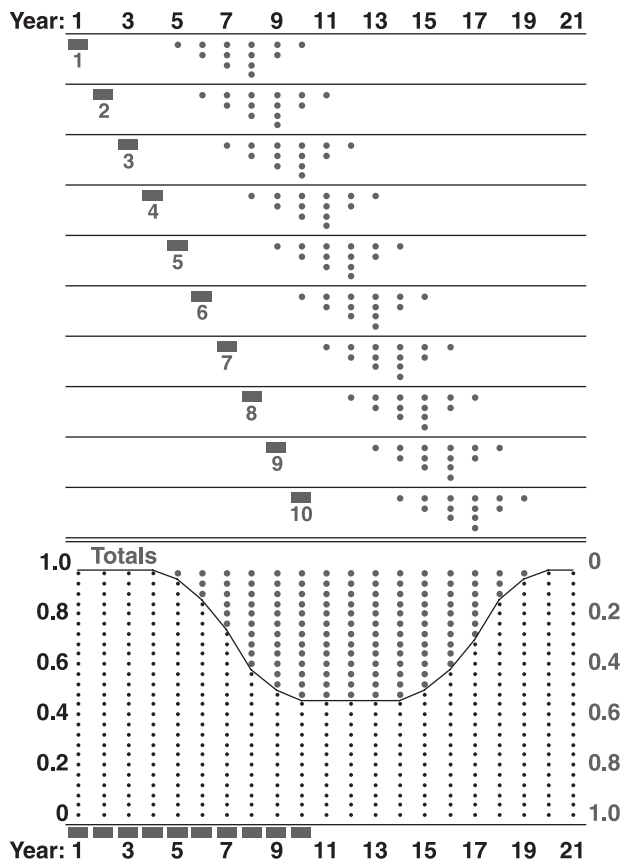
FIGURE 1. Reductions in cancer deaths in a hypothetical situation in which screening is carried out for 10 years. The dots in a specific row in the upper part of the figure represent the deaths averted by that year's screening; the dots in the region entitled "totals" in the lower portion of the figure represent the aggregated numbers of deaths averted, whereas the smaller dots represent deaths that are not averted. The curve represents the mortality rate ratio (left vertical axis) and its complement (right vertical axis).

not shown), the apparent reductions associated with screening are 1–205/250 = 18%, 1–370/500 = 0.26%, and 1–620/750 = 17%, respectively. In contrast, the reductions are 35% and 52% if averaged over years 5 through 19 (any manifestation of effect of early treatment) and 10 through 14 (maximal manifestation), respectively.

Relative to the yearly numbers of deaths in the absence of screening and early treatment, each separate cycle produces its own "deficit" or "trough." The left "lip" of each trough reflects the delay between the time when cancers are detected at a curable stage and when they would otherwise have been fatal. Deaths that occur earlier were not averted by the screening diagnosis and treatment, because the cancer was already incurable at the time of screening. The right lip (where again no deaths are averted) reflects the limits of the

"reach" of the screening instrument—a feature that is discussed subsequently. The width of each separate trough reflects the person-to-person variation in "x", whereas the volume of the trough reflects the overall impact of the single application. Continued regular cycles of an effective screening program eventually produce a steady state. If screening is discontinued, cancer mortality among the screened persons reverts to what one would observe with no screening as the last of the delayed deficits are expressed. The parametric relations in Figure 1 are described in more detail in Miettinen's analysis.[1]

The principle of looking in the appropriate window after initiation of screening is widely appreciated by those who examine nonexperimental data on screening. For example, investigators[4–8] and commentators[9] have assessed whether the extensive prostate-specific antigen (PSA)-based screening begun around 1990 has produced corresponding shortfalls in prostate cancer deaths in the early 2000s. Appropriately, none of these assessments considered the declining prostate cancer death rates in some countries in the early 1990s as evidence of the benefits of PSA-based early detection and treatment, nor did they take unchanged rates in other countries as evidence that earlier treatment had no impact. After all, PSA-based screening was not even available in the 1980s to detect—at a curable stage—the cancers that proved fatal in the early 1990s. The pattern of prostate cancer mortality soon after the introduction of PSA was uninformative and correctly ignored. Similarly, to study the impact of the NHS Breast Screening Programme, which was initiated in Wales in 1991, Fielder and colleagues[10] focused on deaths from breast cancer among women who were diagnosed after the program began and who died *after 1998*.

Curiously, it is in studies in which experimental data have been available—from randomized clinical trials of screening for cancer of the breast, colon, and lung—that the principle of "looking in the right window" has been more neglected. Morrison's textbook[11] devotes a few sentences to this principle; but it then goes on, in all of the examples, to compare cumulative mortality—over the *entire* period of screening and follow up—in the screened and unscreened groups, no matter how long the duration of screening. Until recently, other investigators have done the same.

Caro and McGregor[2] were apparently the first to use this data analysis principle. In a report to the Quebec health ministry, they state: "The difference in cumulative mortality obscures the effect of screening because there is a lag of several years between screening and the time that deaths would have otherwise occurred and, thus, mortality during these early years cannot be influenced by screening. To obtain more revealing estimates requires translating the reported figures to time-specific breast cancer mortality rates (incidence densities)."

The first to reiterate the principle explicitly in the open literature appears to have been Miettinen.[1,3] Much of the quote in the previous paragraph is a paraphrase of his arguments. When he applied this principle to the data from the Malmö mammographic screening trial, in which other authors could see little impact on mortality,[12] the impact became much clearer and stronger.

His reanalysis prompted me to revisit the data from another cancer screening study that we had previously used (without questioning the data analysis) in our graduate teaching in epidemiology.

EXAMPLE AND METHOD

In 1999, Mandel et al[13] reported the latest results of a large U.S. randomized trial of the effect of fecal occult blood screening on colorectal cancer mortality. In 2000,[14] they reported the effect on the incidence of colorectal cancer. A total of 46,551 people were recruited between 1975 and 1978 and randomly assigned to annual screening, biennial screening, or usual care. The incidence end point makes this a particularly sensitive model because of the shorter time scale between action and impact: the focus of the analysis was the impact of discovering and removing polyps and other precancerous lesions that might otherwise (in the absence of this screening and removal) become cancer. A second, unplanned feature of this trial was the pattern and duration of screening. Screening was conducted between 1976 and 1982 and, after a hiatus resulting from a lack of funding, resumed in 1986. All screening was completed in 1992.

The reanalysis presented here is based on the patterns of incidence of colorectal cancer in the first 18 years of the study. In the original report, the authors calculated the ratio of the 18-year cumulative incidence of colorectal cancer in each of the 2 screening groups to the incidence in the control group.[14] This ratio was used to measure the extent to which screening affected incidence. Relative to the control group, the 18-year cumulative incidence ratios were 0.80 and 0.83 for the annual screening and biennial screening groups, respectively.

Our analysis is based on the numbers of cases of colorectal cancer reported in Table 1 of the article (417, 435, and 507 respectively); the numbers at risk at years 0, 2, . . . ; 18 reported at the foot of Figure 1, and the plotted cumulative incidence for each year.[14] From these pieces of information, the numbers of new cases of colorectal cancer for each separate year after the introduction of the program were reconstructed. Because the patterns in the 2 screening arms did not differ much, they were combined. The yearly incidence ratios for the screening group relative to the control group were then calculated using the moving averages of the data for 3 adjacent years.[1,3] Because the focus here is on avoiding bias in point estimation, interval estimates[1] are not shown.

RESULTS

Part A of Figure 2 shows the cumulative incidence of colorectal cancer in the screened and unscreened study groups for each of the 18 years of follow up. The reported reduction in incidence in the screened groups (just under 20%) reported by Mandel and colleagues was based on the cumulative incidence at 18 years. Our yearly incidence density ratios, shown in part B, yield a stronger and more visible "signal." This new analysis highlights the lag from screening to impact, the lag from the discontinuation of screening to the loss of impact, and (after the resumption of screening) the lag from screening to impact. It suggests that had screening continued uninterrupted, there would have been a sustained reduction in incidence of at least 40%. This interpretation is different from that in a review,[15] which stated, "In the U.S. study, colorectal cancer incidence rates were reduced by 20% and 17% in the annually and biennially screened groups, but only after 18 years. No incidence reduction has been observed in either of the 2 European studies, both of which have offered the test at 2-yearly intervals, although the cohorts have been followed for only 13 years so far, and at that stage no effect on incidence was discernible in the US data."

DISCUSSION

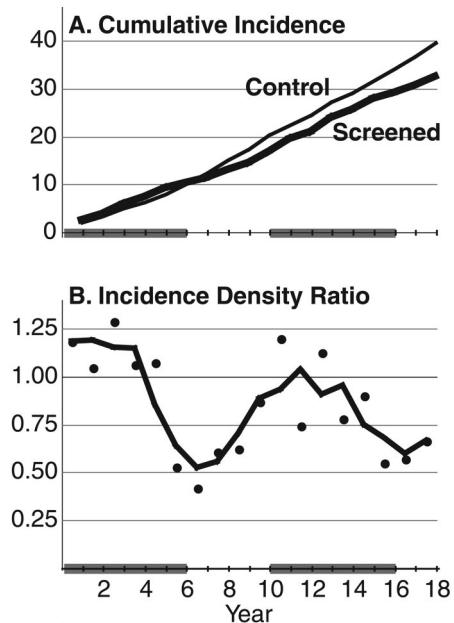In many studies focusing on cancer mortality, the reductions may be obscured or minimized by a number of

FIGURE 2. Colorectal cancer in the unscreened and screened study groups (annual and biennial combined) based on data in Mandel et al.[14] The 2 6-year periods when screening was conducted are shown as thicker lines on the time axis. Cumulative incidence (**A**) is per 1000. Yearly incidence density ratios (**B**) are shown as points.

factors: person-to-person variability in the delay until the averted deaths would have occurred, few screening cycles, limited uptake and adherence, and random variation because of small numbers of deaths. The study reanalyzed here focused on cancer incidence, and on the impact of detecting and removing polyps and other precancerous lesions that might otherwise become cancer. Although the several screening cycles and good compliance helped to create a large impact, the magnitude of this effect is underestimated if one measures it by reductions in cumulative incidence. In contrast, the yearly incidence density ratios provide an undiluted measure of the impact. In addition, the ratios allow the delay to be estimated directly from the data.

This particular cancer incidence example was chosen because the data were reported in sufficient detail for reanalysis. In addition, the unusual pattern of screening and follow up generated a complex "output function" that was much more readily discernible using uncumulated data. However, the principle is a general one; it applies with greater force (using its counterpart, yearly mortality density ratios) to studies that seek to quantify the reduction in mortality achieved by early detection and treatment of already malignant lesions. Indeed, mortality ratios leave less room for misinterpretation than incidence ratios: the reduction in colorectal cancer incidence might simply reflect an advance in the diagnosis of prevalent already malignant lesions rather than a true reduction in future incidence caused by the removal of precancerous lesions. The fact that the incidence density ratio does not exceed 1.0 when screening was reinstituted suggests that this alternative explanation does not account for all of the observed pattern of incidence density ratios.

It should be noted that the time-specific mortality density ratios do not require prior specification of the "window of influence." Rather, if there is sufficient screening and follow up, its location is revealed by the data themselves.

The fact that the pattern of observed mortality ratios is a function of the duration of screening and follow up has an important implication for metaanalysis of data from screening studies. Because each study screens for a different duration, with a different screening interval, and follows up subjects for a different length of time, the locus and shape of its mortality–density–ratio curve will reflect its unique time pattern of screening. If there is one comparative parameter that makes sense for metaanalysis, it is the maximal depth of the trough theoretically achievable with continued screening. However, one must first consider whether the screening and follow up lasted long enough to expose the maximal impact. This prerequisite is discussed in more detail in Miettinen's commentary on the pooling of results from 2 mammographic screening studies with very different screening and follow-up patterns.

In most instances, the impact of screening is obscured if the screening duration or follow up is too short. At the other

extreme, too much follow-up time after the discontinuation of screening, with cumulation of all deaths regardless of their temporal pattern, can also obscure the impact. For example, the report on the extended (24-year) follow up of the Mayo Lung Project examined "whether additional time would allow for a reduction in lung cancer mortality to be observed in this arm."[16] Lung cancer mortality in the intervention arm (intensive screening) over the entire block of 24 years was compared with the corresponding average rate in the usual care arm. The rate in each arm was based on *all l*ung cancer deaths from those in the very first year (deaths that could scarcely have been influenced by detection and slightly earlier treatment) through the end of intensive screening at 6 years up until the end of follow up 18 years after intensive screening was discontinued. Tumors that proved fatal in the later years of follow up must have been well beyond the temporal "reach" of screening during the first 6 years. This strategy of including deaths for several years beyond the impact of the last screening is the temporal analog of evaluating the benefits of screening sigmoidoscopy but including deaths from cancers located beyond the reach of the sigmoidoscope. Including these deaths outside of the "window of influence" associated with the screening dilutes whatever impact (beneficial or otherwise) the early detection and treatment might have already had on lung cancer mortality. If intensive screening and the resultant earlier treatment were indeed effective, time-specific mortality ratios would be more likely to show it; they would also show the length of the lag until the impact becomes apparent and the eventual loss of impact after discontinuing screening.

The emphasis here is the effectiveness of screening in organized trials, but the same principle of the appropriate time window applies to case–control studies,[17] which have the added challenge of minimizing any effects of subject self-selection. However, possibly because the approach is nonexperimental, and also possibly because of the "after-the-fact" perspective that is inherent in case–control studies, these investigators using the case–control approach seem to appreciate the importance of the appropriate window more fully than their clinical trials counterparts.

Although it can be difficult to decide what constitutes "recent" and "distant," the principle of ignoring irrelevant distant and recent exposure to a putative etiologic agent, based on the concept of "latency," is commonly applied to data analyses in (etiologic) research into the unintended effects of an agent. The analysis of data from trials of cancer screening needs to reflect the fact that when cancers are cured by today's early detection and treatment, but would not have been if detected and treated later, these cures only becomes apparent after some delay. Fortunately, if they are allowed to, the data will ultimately speak for themselves.

REFERENCES

1. Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? Available at: http://image.thelancet.com/extras/1093web.pdf. Accessed July 6, 2005.
2. Caro J, McGregor M. Screening for breast cancer in women aged 40–49 years. Montreal: CÉTS Report no. 22, 1993. 91p. Available at: http://www.aetmis.gouv.qc.ca/en/ Accessed July 6, 2005.
3. Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? *Lancet* 2002;359:404–406.
4. Shibata A, Whittemore AS. Prostate cancer incidence and mortality in the United States and the United Kingdom. *J Natl Cancer Inst*. 1998;90:1230–1231.
5. Oliver SE, Gunnell D, Donovan JL. Comparison of trends in prostate-cancer mortality in England and Wales and the USA. *Lancet*. 2000;355:1788–1789.
6. Oliver SE, May MT, Gunnell D. International trends in prostate-cancer mortality in the 'PSA ERA'. *Int J Cancer*. 2001;92:893–898.
7. Perron L, Moore L, Bairati I, et al. PSA screening and prostate cancer mortality *Can Med Assoc J*. 2002;166:586–591.
8. Lu-Yao G, Albertsen PC, Stanford JL, et al. Natural experiment examining impact of aggressive screening and treatment on prostate cancer mortality in two fixed cohorts from Seattle area and Connecticut. *BMJ*. 2002;325:740–743.
9. Albertsen PC. Prostate cancer mortality after introduction of prostate-specific antigen mass screening in the Federal State of Tyrol, Austria [Editorial]. *J Urol*. 2002;168:880–881.
10. Fielder HM, Warwick J, Brook D, et al. A case–control study to estimate the impact on breast cancer death of the breast screening programme in Wales. *J Med Screen*. 2004;11:194–198.
11. Morrison AS. *Screening in Chronic Disease*, 2nd ed. New York: Oxford University Press; 1992.
12. Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet*. 2000;355:129–134.
13. Mandel JS, Church TR, Ederer F, et al. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. *J Natl Cancer Inst*. 1999;91:434–437.
14. Mandel JS, Church TR, Bond JH, et al. The effect of fecal occult-blood screening on the incidence of colorectal cancer. *N Engl J Med*. 2000;343:1603–1607.
15. Atkin W. Options for screening for colorectal cancer. *Scand J Gastroenterol Suppl*. 2003;237:13–16.
16. Marcus PM, Bergstralh EJ, Fagerstrom RM, et al. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst*. 2000;92:1308–1316.
17. Walter SD. Mammographic screening: case–control studies. *Ann Oncol*. 2003;14:1190–1192.

IN THE NEXT ISSUE

Coming in January (Selected papers)

Estrogen metabolism and breast cancer risk

Analgesic drug use and risk of ovarian cancer

***EPHX1* polymorphisms and the risk of lung cancer: A HuGE review**

Breastfeeding and overweight in adolescence

Exposure to nonpersistent insecticides and male reproductive hormones

Exposure misclassification in epidemiologic studies of agricultural pesticides

Polychlorinated biphenyls and neurodegenerative disease mortality

Impact of the 2003 heat wave on all-cause mortality in 9 French cities

Xenobiotic-metabolizing genes and small-for-gestational-age births: Interaction with maternal smoking

Smokeless tobacco use and risk of stillbirth