JNCI J Natl Cancer Inst (2017) 109(10): djx064

doi: 10.1093/jnci/djx064
First published online May 22, 2017
Editorial

EDITORIAL

Challenges in Quantifying Overdiagnosis

Stuart G. Baker, Philip C. Prorok, Barnett S. Kramer

Affiliation of authors: Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.

Correspondence to: Stuart G. Baker, ScD, Division of Cancer Prevention, National Cancer Institute, 9609 Medical Center Dr, Room 5E606, MSC 9789, Bethesda, MD 20892-9789 (e-mail: sb16i@nih.gov).

An important concern in cancer screening is overdiagnosis, the detection of cancer that would not have become clinically evident over the patient's lifespan in the absence of screening. Ripping et al. (1) provide a useful overview of strengths and many key limitations of various approaches to quantifying overdiagnosis. We complement their disquisition with additional considerations.

Lead time is the time from detection on screening to the time when the cancer would surface clinically in the absence of screening. With a good estimate for the distribution of lead times, one can use simulation to estimate the fraction of persons detected with cancer on screening who are overdiagnosed (2). For each screen-detected cancer, the simulation generates the time of clinical detection in the absence of screening based on the lead time distribution and a time to death from other causes. The cancer is overdiagnosed if the former time is smaller than the latter. The resulting estimate of overdiagnosis fraction is sensitive to the choice of lead time distribution.

The challenge with this estimation of the overdiagnosis fraction is obtaining a good estimate of the distribution of lead times. To appreciate the challenge, note that preclinical cancers in a population are a mixture of progressive cancers (that would become clinical in the absence of screening) and nonprogressive cancers (that would never become clinical). The estimated lead time distribution is derived from the estimated distribution of time with progressive cancer, which, in turn, is estimated from the distribution of the duration of preclinical cancer, which also ideally includes the probability of nonprogressive cancer. It is a mathematical theorem that the commonly employed exponential distribution for the lead time is identical to the exponential distribution for the time with progressive cancer. However, a strong assumption is typically made that there is no nonprogressive cancer so that the distribution of the duration of preclinical cancer used to estimate the mean lead time is also exponential—an assumption that risks a substantially biased estimate of the overdiagnosis fraction (3–5). With a more realistic distribution of the duration of preclinical cancer, it is statistically difficult, and likely impossible in many cases, to separate the distribution of time to progressive cancer from the probability of nonprogressive cancer so as to obtain a reliable estimate of the overdiagnosis fraction.

In a stop-screen randomized trial, one randomization group receives no screening and the other randomization group receives periodic screening until the start of a follow-up period. As noted by Ripping et al. (1), the preferred method for estimating the number of persons overdiagnosed in a stop-screen trial is the excess cumulative incidence, the difference in the cumulative incidences of cancer between randomization groups. Unbiased estimation requires that the length of follow-up exceed the longest lead time (1,6,7). However, because the lead time distribution is difficult to estimate, it is difficult to determine the bias associated with estimating the overdiagnosis fraction at a particular follow-up time.

Another commonly used method to estimate the overdiagnosis fraction is the difference between the annual incidence of cancer in a population receiving screening and the estimated annual incidence if, counterfactually, the population screened were not screened. For the latter quantity, Ripping et al. (1) discuss important limitations of the following estimates: annual incidence based on extrapolating a prescreening trend, annual contemporaneous incidence in an unscreened geographic region, and annual contemporaneous incidence among persons who did not accept the screening invitation. Another potentially useful estimate is the annual incidence of late-stage cancer that is presumably minimally affected by screening.

A limitation with this method is lead time bias. The annual incidence in the population receiving yearly screening involves the sum of clinical cancers in the interval before the screen, overdiagnosed cancers, and progressive cancers detected on screening, which has moved the detection time forward by the lead time. The annual incidence if the screened population were not screened involves the sum of the following: clinical cancers that became screen-detectable in the interval and clinical cancers that would have been detected on hypothetical previous screening and whose lead time would have led to clinical incidence in the interval. A rigorous adjustment requires a good

Received: March 7, 2017; **Accepted:** March 13, 2017

estimate of the distribution of lead times, which, as discussed previously, is difficult to obtain.

Apart from noting the challenges to bias, Ripping et al. (1) call for a uniform definition of overdiagnosis fraction: one in which the denominator is screen-detected plus interval cancers. However, different estimates of overdiagnosis are useful in different contexts. All have as a numerator the number of overdiagnosed cancers. For an easy to understand metric of the fraction of screen-detected cancers that are overdiagnosed, the denominator would be screen-detected cancers. For the patient or policy-maker who wants to know the probability of overdiagnosis in a screening program, the denominator would be the number of persons entering the screening program. A challenge to estimating this latter overdiagnosis fraction in a randomized trial is noncompliance. When noncompliance is all-or-none (such as refusing screening), a simple adjustment is possible under plausible assumptions (9,11).

Apart from bias from a misspecified lead time distribution and the differences in the definition of the overdiagnosis fraction, a third explanation for the wide range of estimates of the overdiagnosis fraction in the literature is random variability. Variability arises in estimates of the lead time distribution and in estimates of excess cumulative incidence, and it is seldom formally addressed.

To quote William Bruce Cameron (12), "Not everything that counts can be counted. . ." Overdiagnosis is a good example of an event that is not directly observable yet is important to assess, even if a "precise count" is difficult to achieve. Because overdiagnosis is not observable, statistical methods and assumptions regarding the lead time distribution are needed to estimate the overdiagnosis fraction. Unfortunately, the estimates can be sensitive to these assumptions. Realistically, the best that can be hoped for is an order of magnitude quantitation about whether the amount of overdiagnosis is large or small (13). We recommend a clear definition of the type of overdiagnosis fraction estimated and the setting where it would be useful, a detailed discussion of the assumptions underlying estimation, and a report of confidence intervals whenever sampling is involved.

Note

The authors have no conflicts of interest to declare.

References

1. Ripping TM, ten Haaf K, Verbeek ALM, van Ravesteyn NT, Broedersm MJM. Quantifying overdiagnosis in cancer screening: A systematic review to evaluate the methodology. *J Natl Cancer Inst*. 2017;109(10): djx060.
2. Etzioni R, Xia J, Hubbard R, Weiss NS, Gulati R. A reality check for overdiagnosis estimates associated with breast cancer screening. *J Natl Cancer Inst*. 2014; 106(12): dju315.
3. Baker SG, Prorok PC, Kramer BS. Lead time and overdiagnosis. *J Natl Cancer Inst*. 2014;106(12): dju346.
4. Zahl PH, Jørgensen KJ, Gøtzsche PC. Lead-time models should not be used to estimate overdiagnosis in cancer screening. *J Gen Intern Med*. 2014;29(9): 1283–1286.
5. Zahl PH, Jørgensen KJ, Gøtzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. *Br J Cancer*. 2013;109:2014–2019.
6. Etzioni R, Gulati R, Mallinger L, et al. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Ann Intern Med*. 2013;15811:831–838.
7. Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: The importance of length of observation period and lead time. *Breast Cancer Res*. 2013; 15:R41.
8. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: An independent review. *Lancet*. 2012;380(9855):1778–1786.
9. Jacklyn G, Glasziou P, Macaskill P, Barratt A. Meta-analysis of breast cancer mortality benefit and overdiagnosis adjusted for adherence: Improving information on the effects of attending screening mammography. *Br J Cancer*. 2016;114(11):1269–1276.
10. Ripping TM, Verbeek AL, Broeders MJ. Overdiagnosis in cancer screening: The need for a standardized denominator. *J Med Screen*. 2016; 23(2):111–113.
11. Baker SG, Kramer BS, Lindeman KL. Latent class instrumental variables: A clinical and biostatistical perspective. *Stat Med*. 2016;35(1):147–160.
12. Cameron WB. *Informal Sociology: A Casual Introduction to Sociological Thinking*. New York, NY: Random House; 1963.
13. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. *N Engl J Med*. 2016; 375(15):1438–1447.