Oversimplifying Overdiagnosis

Ruth Etzioni, PhD and Roman Gulati, MS

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Estimating the frequency of overdiagnosis due to cancer screening is not a simple task, because overdiagnosis is almost never observable. Since we treat most cancers and their precursors, we rarely get to follow a tumor over its natural course and learn whether that tumor would have persisted as an indolent lesion, spontaneously regressed, or progressed to the point of producing symptoms.

To accurately assess the frequency of overdiagnosis, we need to really understand disease natural history. And this is a hard problem. In fact, one of the greatest controversies in the debate about overdiagnosis is how best to estimate this elusive quantity.

In this issue of JGIM, Zahl et al.[1] comment on this controversy. To summarize, there are two main camps.[2] One recognizes that overdiagnosis is a result of disease natural history that cannot be observed. Therefore, this camp uses statistical models to infer the most plausible underlying natural history. In a sense, this work uses observed data to make inferences about unobservable disease progression. Zahl et al.[1] term this the "lead-time" approach, because lead time is the main feature of disease natural history that is typically studied to reach conclusions about overdiagnosis (Fig. 1).

The second camp uses observed quantities as proxies for the unobserved frequency of overdiagnosis. This approach commonly estimates the incidence of overdiagnosis as the "excess incidence" of disease in the presence of screening relative to observed (or extrapolated) incidence in the absence of screening.

As we[2] and others have shown, the two approaches typically yield quite different results. Estimates based on the excess-incidence approach routinely exceed those based on the lead-time approach. Therefore, in judging the validity of any overdiagnosis study, it is absolutely critical to know which approach has been used, and to understand its merits.

In their article, Zahl et al.[1] champion the excess-incidence approach, rejecting the lead-time approach in all settings. We respectfully disagree. Indeed, the main premise of this Editorial is that, when attempting to estimate an unobservable quantity such as overdiagnosis, an empirical calculation such as that based on excess incidence can be quite misleading.

A growing body of evidence is confirming that the excess-incidence proxy for overdiagnosis is incorrect and generally leads to estimates that are biased upwards.[3,4] The reason for this is that, because all cancers are latent for at least some time, the introduction of screening always produces a rapid increase in disease incidence. This increase in incidence is attributable to a combination of overdiagnosed cases and non-overdiagnosed cases whose date of detection has been brought forward in time by the screening test. The excess-incidence approach cannot partition the increase in incidence into the portion that is rightly attributable to overdiagnosis and that due to early detection of disease that would have manifested in the absence of screening.

Recognizing this problem, some practitioners of the excess incidence approach wait until incidence has stabilized and then compute the difference between the observed incidence under screening and observed (or extrapolated) incidence without screening. However, a simulation study[4] showed that this method may require waiting for many years. A recent report from the Canadian National Breast Screening study[5] waited 10 years after the end of the initial 5-year screening period. However, the accuracy of the resulting estimate of overdiagnosis is unclear, because screening practices among participants on both arms of the trial were not documented during this time. In the absence of a control group, extended extrapolation of the incidence that would be expected without screening is required, and this often amounts to a guessing game with different guesses greatly impacting results. This was an issue in a study of breast cancer incidence in the US,[6] which extrapolated the incidence without screening from a baseline in 1976 all the way to 2008.

There are other significant problems[2,7] with the excess-incidence approach. Although ad-hoc adjustments have been proposed to address these, they have never been shown to eliminate the resulting bias. Indeed, Biesheuvel et al.,[3] referring to their review of many studies using this approach, conclude that "all available estimates of overdetection due to mammography screening are seriously biased."
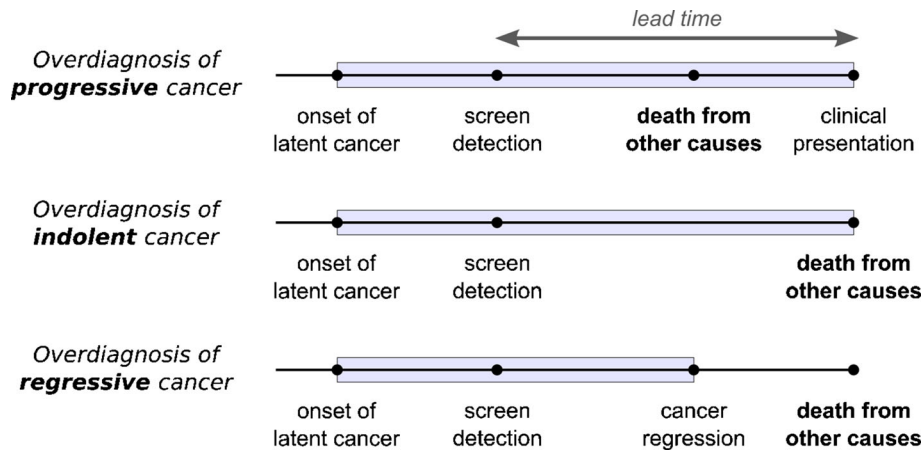
Figure 1.. Schematic illustration of overdiagnosis of progressive, indolent, and regressive cancers detected by screening. Shaded boxes indicate windows of detectability. The lead time is the time from screen detection to the point at which disease would have presented clinically in the absence of screening. In the case of progressive cancer, overdiagnosis occurs when other-cause death happens during the lead time. Therefore, the longer the lead time, the greater the chance of overdiagnosis. Indolent or persistent cancers and regressive cancers never reach the point of clinical presentation; thus, they are always overdiagnosed.

What then of the lead-time approach? Or, more generally, any approach that attempts to infer the underlying natural history? Is it doomed as well?

Unfortunately, there is no simple answer to this question. We believe, based on our extensive experience with this approach, that models can be developed that will reliably inform about key features of disease natural history. However, there is more to it than this.

In the most basic case, when the majority of tumors are progressive, the minimal description of disease natural history that enables estimation of overdiagnosis includes a risk of disease onset, a risk of progression from latent to clinical disease in the absence of other-cause death, and the sensitivity of the screening test. So long as good data on disease incidence with and without screening are available and screening patterns in the population are known, lead-time approaches can provide reliable estimates of these quantities. There is a large body of work in the cancer[8,9] and HIV[10] literatures on such lead-time models.

Zahl et al.'s main objection is that these basic lead-time models (that assume all tumors are progressive) are often used in settings such as breast cancer where, they maintain, many tumors regress. They argue that, by not accommodating the possibility of regression, these models are wrong and consequently that the empirical excess-incidence approach must be preferred.

We have three main responses to this assertion.

First, even if disease does regress in a non-trivial fraction of cases, it has not been demonstrated that the more basic lead-time models always produce incorrect estimates of overdiagnosis. In the case where the interval to regression is short relative to the interval to progression, the presence of regressive cases will not greatly change incidence under screening and will not

significantly impact estimates of lead time or overdiagnosis.[11] When the interval to regression is longer, the incidence under screening will be impacted and this will generate a higher estimate of the average lead time than in the absence of regression and, consequently, a greater projected frequency of overdiagnosis. More thorough investigation is needed to establish how biased these basic lead-time models are likely to be when some tumors regress. As the oft-cited adage by George Box goes, "all models are wrong, but some are useful."

Second, there are versions of lead-time models that do address a more complex biology[11–13] and do not simply assume that all tumors are progressive. Such models generally need to incorporate external information on parts of the disease progression and detection process, because typical screening studies do not provide enough information to identify all of the risks describing a complex disease natural history. Thus, for example, a recent cervical cancer model[11] allowed for regression, persistence, and progression of CIN 2/3 lesions, but incorporated external information on the risk of CIN 2/3 onset based on the age-specific prevalence of HPV lesions in the population. In the case where appropriate external information is available and a coherent statistical method is used to integrate this information with data on disease incidence with and without screening, the lead-time modeling approach can provide important insights into disease natural history and overdiagnosis.

Third, we would challenge Zahl et al.'s insistence that a significant number of breast tumors must regress. This conclusion is based on two studies[14,15] that empirically compared two overlapping age-matched groups of women. The first group was screened several times and the second was screened only once, at the end of a comparable-length interval. The incidence of detected

disease was higher in the first group than in the second, and Zahl et al. attributed the excess to detection of tumors that would have regressed in the absence of screening. However, since mammography is not perfectly sensitive, particularly in detecting smaller tumors, the increased incidence in the group with multiple screens is to be expected even if tumors do not regress. Indeed, the observed excess could be partly (or even largely) due to increased intensity of screening and consequently better program sensitivity to detect small, latent tumors in this group. That the difference between the two groups persists with further follow-up is also to be expected because, owing to the way in which the comparison groups were constructed there will always be more intensive screening in the first group than in the second group. Thus, there may be plausible alternative explanations for the empirical observation of increased detection in the group with multiple screening tests.

In general, we agree with Zahl et al. that when disease can regress, application of lead-time models may represent a simplification of disease biology and may produce biased results, or at least results that are contingent on other inputs to the analysis. But this does not mean that the empirical excess-incidence approach, with all of its known deficiencies, should be used instead. In this endeavor to estimate the unobservable, we stand by a well-known quote attributed to Albert Einstein: "Everything should be made as simple as possible, but not simpler."

Conflict of Interest: *The authors declare that they do not have a conflict of interest.*

Corresponding Author: *Ruth Etzioni, PhD; Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B230, Seattle, WA 98109-1024, USA (e-mail: retzioni@fhcrc.org).*

REFERENCES

1. **Zahl PH, Jorgensen KJ, Gotzsche CO.** Lead-time models should not be used to estimate overdiagnosis in cancer screening. J Gen Intern Med doi:10.1007/s11606-014-2812.2.
2. **Etzioni R, Gulati R, Mallinger L, Mandelblatt J.** Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. Ann Intern Med 2013;158:831–8. PMC: 3733533.
3. **Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L.** Effects of study methods and biases on estimates of invasive breast cancer over detection with mammography screening: a systematic review. Lancet Oncol 2007;8:1129–38. PMC.
4. **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. PMC: 3706885.
5. **Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA.** Twenty-five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomized screening trial. BMJ 2014;348:g366. PMC: 3921437.
6. **Bleyer A, Welch HG.** Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 2012;367:1998–2005. PMC.
7. **Duffy SW, Lynge E, Jonsson H, Ayyaz S, Olsen AH.** Complexities in the estimation of overdiagnosis in breast cancer screening. Br J Cancer 2008;99:1176–8. PMC.
8. **Zelen M, Feinleib M.** On the theory of screening for chronic diseases. Biometrika 1969;56:601–14. PMC.
9. **Pinsky PF.** Estimation and prediction for cancer screening models using deconvolution and smoothing. Biometrics 2001;57:389–95. PMC.
10. **De Gruttola V, Lagakos SW.** Analysis of doubly-censored survival data, with application to AIDS. Biometrics 1989;45:1–11. PMC.
11. **Vink MA, Bogaards JA, van Kemenade FJ, de Melker HE, Meijer CJ, Berkhof J.** Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data. Am J Epidemiol 2013;178:1161–9. PMC.
12. **Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL.** The Wisconsin Breast Cancer Epidemiology Simulation Model. J Natl Cancer Inst Monogr 2006:37–47. PMC.
13. **Seigneurin A, Francois O, Labarere J, Oudeville P, Monlong J, Colonna M.** Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. BMJ 2011;343:d7017. PMC: 3222945.
14. **Zahl PH, Gotzsche PC, Maehlen J.** Natural history of breast cancers detected in the Swedish mammography screening programme: a cohort study. Lancet Oncol 2011;12:1118–24. PMC.
15. **Zahl PH, Maehlen J, Welch HG.** The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008;168:2311–6. PMC.