Published in final edited form as:

Acad Radiol. 2017 February ; 24(2): 226–229. doi:10.1016/j.acra.2016.05.020.

Missteps in Current Estimates of Cancer Overdiagnosis

Christoph I. Lee, MD, MSHS^a and Ruth Etzioni, PhD^b

^aDepartment of Radiology University of Washington School of Medicine; Department of Health Services, University of Washington School of Public Health; 825 Eastlake Avenue East, G3-200, Seattle, WA, 98109-1023, USA

^bDivision of Public Health Sciences, Fred Hutchinson Cancer Research Center; 1100 Fairview Avenue N, M2-B230, PO Box 19024, Seattle, WA 98109-1024, USA

Abstract

The balance between the benefits and harms of imaging-based cancer screening continues to be an area of controversy and widespread media attention. Of the potential harms, overdiagnosis from screening is likely the most elusive in estimating and quantifying. This article describes the major methodological issues with recently reported estimates of overdiagnosis that are based on excess cancer incidence, and suggests that modeling focused on tumor lead-time can serve as a complementary method for excess incidence-based overdiagnosis estimates. Radiologists should be conversant on the topic of overdiagnosis, and understand the limitations of different methods used to estimate its magnitude.

Introduction

With recent updates to both the American Cancer Society (ACS) and U.S. Preventive Services Task Force (USPSTF) mammography recommendations, the balance between the benefits and harms of routine cancer screening are again in the media spotlight.^{1,2} While stakeholders can easily grasp notions of benefits such as decreased mortality and morbidity and harms such as false-positive tests and unnecessary biopsies, one of the more abstract and confusing factor for patients, physicians, radiologists, and policymakers is overdiagnosis. Simply mentioning this potential harm without a sense for its magnitude leaves patients and physicians without actionable information to use in shared decision-making, something recommended by both the ACS and USPSTF.

Overdiagnosis can be defined as a screen-detected cancer that would not have become clinically significant during the patient's lifetime. While the definition is simple, its measurement is quite complex. Since all screen-detected cancers are treated under current standard of care, whether a case has been overdiagnosed or not cannot be directly observed.

Corresponding Author: Christoph I. Lee, MD, MSHS, 825 Eastlake Avenue East, G3-200, Seattle, WA 98109-1023, (206) 288-6783 office, (206) 288-6473 fax, stophlee@uw.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Instead, the magnitude of overdiagnosis can only be estimated with different techniques requiring varying assumptions. Not surprisingly, estimates for breast cancer overdiagnosis vary over a wide range in the medical literature, and are as low as 5% and as high as 42%.^{3,4} Some of this variability may be due to the use of different denominators or references (e.g., only cases detected by screening or all cancer cases).³ However, much of the reason for variability lies in the methodologies used for estimation.

In this article, we examine two major methodologies used to estimate cancer overdiagnosis in the context of breast cancer screening. We will describe why a commonly used approach, based on excess incidence (EI) under screening, is prone to overestimation. We will also describe how an alternative approach based on estimation of the lead-time (LT), while imperfect, can provide useful complementary information. In reviewing the validity of these approaches, we hope to better elucidate the assumptions that generate discrepant overdiagnosis estimates. We conclude that refinements are needed in current approaches to estimation if we are to provide information that can properly inform shared decision-making for cancer screening.

The Counterfactual Incidence Problem

There has been extensive media coverage regarding breast cancer overdiagnosis based on the EI approach.⁶ At first glance, the use of excess incidence for estimating overdiagnosis is certainly appealing given its seeming directness and simplicity. Using the EI approach, breast cancer incidence trends with and without screening are compared to provide estimates of cancers that would not have presented clinically in the absence of screening. Unfortunately, this direct method of estimating overdiagnosis has multiple limitations that call into question its validity.

The most obvious limitation with the EI approach is that once screening is started, it is not possible to observe the true baseline incidence in the absence of screening. In other words, the counterfactual incidence without screening is never directly measurable. Instead, since all cancers are treated after detection, the cancer incidence without screening has to be imputed or extrapolated in some fashion. Studies using the EI approach have attempted to compensate for the lack of data on the true counterfactual incidence using ad hoc corrections or extrapolations.

In one of the most widely publicized EI approach studies estimating breast cancer overdiagnosis, Bleyer and Welch estimated that 31% of all breast cancers diagnosed in 2008 were overdiagnosed.⁶ They used Surveillance, Epidemiology, and End Results (SEER) data to examine trends in the incidence of early- and late-stage breast cancer among women 40 years of age and older from 1976 through 2008. Similar to other studies using the EI approach,^{4,7} Bleyer and Welch extrapolated the counterfactual incidence from breast cancer incidence trends in a different patient population not offered screening. Specifically, they used incidence trends in women < 40 years of age during the 30-year study period to estimate the trend over the same time period in the counterfactual incidence for women 40 years and older.^{6,8,9} To obtain their final year estimate of overdiagnosed cases for calendar year 2008, Bleyer and Welch were required to extrapolate their best-guess incidence

estimates over three decades, during which time very small changes in the counterfactual incidence trend would have significantly changed their results. This study illustrates how extrapolation of data over a long period of time can undermine the reliability of EI-based estimates of overdiagnosis.

The Ecological Fallacy Problem

Recently, Harding et al examined the association between mammography screening rates and incidence of breast cancer across U.S. counties to suggest widespread overdiagnosis using an ecological study.¹¹ In this study, investigators merged data from the SEER cancer registry with estimates of screening mammography rates published by the National Cancer Institute's Small Area Estimates for Screening Behaviors program. By following rates of breast cancer diagnosis in each county starting in 2000 for the next 10 years, the study team found increased breast cancer diagnosis in counties where more women reported undergoing screening mammography but no statistically significant difference in subsequent mortality between counties reporting higher or lower mammography use.¹¹

Estimating overdiagnosis based on county-level differences should be approached with suspicion. Ecological studies, by definition, relate frequency of exposure to an intervention (in this case, reported mammography use) to a population outcome (breast cancer incidence and mortality by county). Such studies are limited by the concept of ecological fallacy, where conclusions are inappropriately made about the nature of individuals based on the groups to which those individuals belong. Thus, even though Harding et al found no difference in breast cancer mortality between counties with higher and lower mammography relative use, they could not demonstrate that the women exposed to more screening were the same women with greater cancer incidence and unchanged mortality rates.

With breast cancer screening, there are a myriad of contextual factors beyond the imaging test such as patterns of mammography use, comorbidities, patient behaviors, and treatment patterns that vary geographically and can influence overdiagnosis estimates.¹² These alternative explanations for outcomes are often unaccounted for in ecological studies that rely on county-level population comparisons. In fact, even among ecological studies addressing breast cancer screening outcomes, there is wide variability in reported findings with other studies conducted at that state level showing lower breast cancer mortality associated with higher rates of mammography use.^{13,14}

The Insufficient Follow-Up Problem

Beyond the counterfactual incidence problem, the EI approach for estimating overdiagnosis has been hampered by the need for adequate follow-up time to determine which cancers are clinically significant and which are truly overdiagnosed. This limitation has been frequently cited as a probable cause for the EI approach leading to overestimation of overdiagnosis.^{15–17} Since all cancers have some latent period, the introduction of screening causes an automatic increase in cancer incidence. However, this immediate increase in incidence represents both overdiagnosed and clinically relevant cancers detected earlier due to screening.¹⁷

In principle, the EI approach requires waiting until the cancer incidence has stabilized in a population and then computing differences between observed incidence with screening and extrapolations of incidence as if there were no screening. Duffy et al suggests that the proper follow-up time needed after full adoption of a screening program is at least as long as the longest lead-time for adequate excess incidence estimates.¹⁸ In their analysis postulating a fifteen-year screening program among women in England and Wales and a lead-time ranging up to ten years, a follow-up period of 25 years or more would be needed to accurately estimate overdiagnosis via the EI approach.¹⁸ Many studies using the EI approach, thus, likely do not provide adequate follow-up to truly differentiate excess cancers into clinically relevant and overdiagnosed cases.

The Trial Design Problem

The EI approach has been applied to randomized controlled screening trial data as well as population-based observational studies. Given the problem of the counterfactual incidence, some researchers suggest that overdiagnosis estimates should be based on incidence data from screening trials, since these study designs provide a control group of non-screened women.¹⁹ While randomized controlled trials are certainly the gold standard for determining screening efficacy, they are not necessarily the gold standard for overdiagnosis estimation even if sufficient follow-up time is established.

Only certain screening trial designs have the potential to yield unbiased estimates for cancer incidence in the absence of screening. Cumulative excess incidence estimates from continuous screening trials, for instance, are prone to producing biased estimates. In continuous screening trials, screening continues into the follow-up period. This leads to overestimation of excess cancer incidence in the screening arm since relevant cases that would have become clinically apparent after the follow-up period are captured. In contrast, similar cancers detected in the control arm will not be included in the excess incidence calculation.¹⁷

In contrast, stop-screen trials do have the potential to yield more unbiased estimates. In these studies, mammography screening occurs only during the first few years of the study and then the follow-up period involves no additional screening. Yet, given that routine mammography became the standard of care after these randomized controlled trials, actual monitoring of screening behavior in both the interventional and control groups after the intervention period is extremely difficult.

One of the more highly publicized EI studies for estimating overdiagnosis was based on follow-up data from the Canadian Breast Cancer Screening Trial.²⁰ In this follow-up report for breast cancer incidence and mortality from a randomized trial conducted in the 1980s, study authors concluded that an excess incidence of 106 cancers, or 22% (106/484) of all screen-detected cancers, were observed in the intervention arm and attributable to overdiagnosis.²⁰ Unfortunately, investigators were not able to monitor screening mammography behavior in both arms after the initial five-year screening period.¹⁷ Their estimate of overdiagnosis would only be valid if routine mammography stopped in the screening arm and if mammography screening never started in the control arm. However,

given that screening likely occurred in both arms after the trial potentially at different rates, any follow-up cumulative excess incidence estimate could be biased.

An Alternative Lead-Time Approach

Given the aforementioned limitations and problems with current EI approaches, others have estimated cancer overdiagnosis using an alternative approach based on lead-time measurements (the LT approach). Overdiagnosis can be conceptualized as a competition between the lead-time from screening and the time between screening and other-cause death.⁵ The frequency of overdiagnosis, then, is the likelihood that the time between screen detection and other-cause death is shorter than the time to non-screen detected cancer diagnosis.

While less direct than the EI approach, the LT approach has the potential to avoid the major problems of insufficient follow-up time and trial design limitations. The LT approach requires incidence data on screen-detected cancers and interval cancers, as well as utilization rates of mammography in a screened population in order to infer underlying cancer incidence and tumor progression.^{21,22} Tumor lead-times are derived from sojourn time distributions, which can be estimated via established approaches.^{21,23} While more complex, the LT approach follows disease natural history, potentially allowing for estimates of overdiagnosis under any screening strategy with known sensitivity. Limitations of the modeling approach include the fact that it is sensitive to assumptions made during model development, as well as any assumptions that are made about model input parameters.

The Progressive Disease Assumption

Since the lead-time is not directly observable, a formal estimation approach is required to derive lead-time measures based on sojourn time (the latent pre-clinical duration).^{24,25} Most estimation approaches assume that disease is progressive, although sojourn time can be highly variable. Some modeling studies using the LT approach to estimate overdiagnosis have been criticized for assuming that all breast cancers are progressive. While controversial, reports suggest that a proportion of small breast cancers may actually regress.²⁶ Moreover, a certain proportion of tumors may remain indolent during a woman's lifetime.

Models of disease natural history should ideally account for the presence of non-progressive tumors that may remain indolent during a person's lifetime. Thus, in model development, the sojourn time distribution must allow for some cancers to have infinite lead-time. Ignoring non-progressive cancers may lead to inaccurate estimates of overdiagnosis.⁵ Some recently developed lead-time models, however, are able to address more complex tumor biology and do not assume that all tumors are progressive.^{27,28} Nevertheless, the reliability of the model estimates from such approaches may be difficult to establish as different mixtures of indolent and progressive cancers could yield similar patterns of screen- and interval-detection in practice. The potential non-uniqueness of model estimates in this setting is referred to as an identifiability problem.

Modeling as a Complement to the EI Approach

The LT approach, given its focus on disease natural history, can be used to provide a reliability test for existing studies of overdiagnosis based on the EI approach. Recently, the modeling approach, leveraging the close link between lead-time and overdiagnosis, challenged the reliability of the purported 31% overdiagnosis rate for breast cancer reported by Bleyer and Welch.²⁹ If the chances of overdiagnosis and other-cause survival are known, then the modeling approach can calculate the associated lead-time distribution. Calculating the associated lead-time from the overdiagnosis estimates reported by Bleyer and Welch yields a mean lead-time of 9 years for invasive breast cancer cases.²⁹ This estimate is much longer than the 2–4 year consensus estimate previously published for mean lead-time of invasive breast cancers,¹⁸ and suggests that the published estimates of overdiagnosis using the EI approach could be excessive.

In addition to providing a method for appraising the reliability of overdiagnosis estimates made by the EI approach, the LT approach can go beyond just calculating the proportion of cancers attributable to overdiagnosis. Other benefits of the LT approach include how overdiagnosis estimates would vary in subpopulations or under different imaging protocols,^{30,31} so long as the identifiability problem can be resolved. More work is needed to establish circumstances under which this approach can provide reliable results when allowing for a non-trivial fraction of indolent cases in the population.

Conclusion

Efforts to generate more reliable estimates of overdiagnosis first require recognizing inherent limitations in the methods currently used. Estimates of breast cancer overdiagnosis based on cumulative excess incidence, both using observational and trial data, are likely to be inflated. Current limitations include the problems of unknown counterfactual incidence, insufficient follow-up time, and trial design limitations. While complex modeling methods that calculate rates of overdiagnosis based on tumor lead-time are a less direct method for estimation, modeling can provide a complementary role for ensuring the reliability of overdiagnosis estimates based on excess incidence. Models based on reliable natural history estimates have the potential for informing screening recommendations and policies, as they can be used to project rates of overdiagnosis under variable scenarios. Radiologists should be conversant about the limitations of the current methods used to estimate overdiagnosis so that they can serve as active stakeholders in providing more reliable information for shared decision-making regarding cancer screening.

Acknowledgments

Funding Acknowledgments

This work was funded by the National Cancer Institute (1R01CA192402-01A1). Dr. Lee's time was also funded in part by the American Cancer Society (126947-MRSG-14-160-01-CPHPS).

References

1. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA*. 2015; 314:1599–1614. [PubMed: 26501536]
2. Siu AL. Force USPST. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2016
3. de Gelder R, Heijnsdijk EA, van Ravesteijn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev*. 2011; 33:111–121. [PubMed: 21709144]
4. Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control*. 2010; 21:275–282. [PubMed: 19894130]
5. Etzioni R, Gulati R. Recognizing the Limitations of Cancer Overdiagnosis Studies: A First Step Towards Overcoming Them. *J Natl Cancer Inst*. 2016; 108
6. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med*. 2012; 367:1998–2005. [PubMed: 23171096]
7. Njor SH, Olsen AH, Blichert-Toft M, Schwartz W, Vejborg I, Lynge E. Overdiagnosis in screening mammography in Denmark: population based cohort study. *BMJ*. 2013; 346:f1064. [PubMed: 23444414]
8. Kopans DB. Arguments against mammography screening continue to be based on faulty science. *Oncologist*. 2014; 19:107–112. [PubMed: 24536051]
9. Smith RA. Counterpoint: Overdiagnosis in breast cancer screening. *J Am Coll Radiol*. 2014; 11:648–652. [PubMed: 24794765]
10. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med*. 2012; 156:491–499. [PubMed: 22473436]
11. Harding C, Pompei F, Burmistrov D, Welch HG, Abebe R, Wilson R. Breast Cancer Screening, Incidence, and Mortality Across US Counties. *JAMA Intern Med*. 2015; 175:1483–1489. [PubMed: 26147578]
12. Elmore JG, Etzioni R. Effect of Screening Mammography on Cancer Incidence and Mortality. *JAMA Intern Med*. 2015; 175:1490–1491. [PubMed: 26147329]
13. Cooper GS, Yuan Z, Bowlin SJ, et al. An ecological study of the effectiveness of mammography in reducing breast cancer mortality. *Am J Public Health*. 1998; 88:281–284. [PubMed: 9491023]
14. Das B, Feuer EJ, Mariotto A. Geographic association between mammography use and mortality reduction in the US. *Cancer Causes Control*. 2005; 16:691–699. [PubMed: 16049808]
15. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review. *Lancet Oncol*. 2007; 8:1129–1138. [PubMed: 18054882]
16. Puliti D, Miccinesi G, Paci E. Overdiagnosis in breast cancer: design and methods of estimation in observational studies. *Prev Med*. 2011; 53:131–133. [PubMed: 21658405]
17. Etzioni R, Gulati R. Oversimplifying overdiagnosis. *J Gen Intern Med*. 2014; 29:1218–1220. [PubMed: 24830739]
18. Duffy SW, Agbaje O, Tabar L, et al. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res*. 2005; 7:258–265. [PubMed: 16457701]
19. Freidlin B, Korn EL. A model too far. *J Natl Cancer Inst*. 2014; 106:djt368. [PubMed: 24399851]
20. 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: randomised screening trial. *BMJ*. 2014; 348:g366. [PubMed: 24519768]
21. Pinsky PF. An early- and late-stage convolution model for disease natural history. *Biometrics*. 2004; 60:191–198. [PubMed: 15032789]

22. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics*. 2008; 64:10–19. [PubMed: 17501937]
23. Davidov O, Zelen M. Overdiagnosis in early detection programs. *Biostatistics*. 2004; 5:603–613. [PubMed: 15475422]
24. Shen Y, Zelen M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *J Clin Oncol*. 2001; 19:3490–3499. [PubMed: 11481355]
25. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst*. 2009; 101:374–383. [PubMed: 19276453]
26. Zahl PH, Jorgensen KJ, Gotzsche PC. Lead-time models should not be used to estimate overdiagnosis in cancer screening. *J Gen Intern Med*. 2014; 29:1283–1286. [PubMed: 24590736]
27. 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. [PubMed: 17032893]
28. 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. [PubMed: 22113564]
29. 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
30. Lansdorp-Vogelaar I, Gulati R, Mariotto AB, et al. Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. *Ann Intern Med*. 2014; 161:104–112. [PubMed: 25023249]
31. van Ravesteyn NT, Schechter CB, Near AM, et al. Race-specific impact of natural history, mammography screening, and adjuvant treatment on breast cancer mortality rates in the United States. *Cancer Epidemiol Biomarkers Prev*. 2011; 20:112–122. [PubMed: 21119071]