Special Report

Missteps in Current Estimates of Cancer Overdiagnosis

Christoph I. Lee, MD, MSHS, Ruth Etzioni, PhD

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.

Key Words: Overdiagnosis; overtreatment; mammography; cancer screening.

INTRODUCTION

With recent updates to both the American Cancer Society and the U.S. Preventive Services Task Force mammography recommendations, the balance between the benefits and harms of routine cancer screening is again in the media spotlight (1,2). Whereas 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 American Cancer Society and the U.S. Preventive Services Task Force.

Overdiagnosis can be defined as a screen-detected cancer that would not have become clinically significant during the patient's lifetime. Although the definition is simple, its measurement is quite complex. Because all screen-detected cancers are treated under current standard of care, whether a case has been overdiagnosed or not cannot be directly observed. 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 (eg, 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), although 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 (5). At first glance, the use of EI 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 di-

Acad Radiol 2016; ■:■■–■■

From the Department of Radiology, University of Washington School of Medicine, 825 Eastlake Avenue East, G3-200, Seattle, WA 98109-1023 (C.I.L.); Department of Health Services, University of Washington School of Public Health (C.I.L.), Seattle, Washington; Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington (R.E.). Received March 14, 2016; accepted May 6, 2016. Funding: 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). **Address correspondence to:** C.I.L. e-mail: stophlee@uw.edu

http://dx.doi.org/10.1016/j.acra.2016.05.020

rectly measurable. Instead, because 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 (5). They used the Surveillance, Epidemiology, and End Results 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 (4,6), 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 (5,7,8). 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 (9). In this study, the investigators merged data from the Surveillance, Epidemiology, and End Results cancer registry with estimates of screening mammography rates published by the National Cancer Institute's Small Area Estimates for Screening Behaviors program. By following the 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 (9).

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 (10). 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 (11,12).

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 (13–15). Because 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 (15).

In principle, the EI approach requires waiting until the cancer incidence has stabilized in a population, and then computing the differences between observed incidence with screening and extrapolations of incidence as if there were no screening. Duffy et al. suggest that the proper follow-up time needed after full adoption of a screening program is at least as long as the longest LT for adequate EI estimates (16). In their analysis postulating a 15-year screening program among women in England and Wales and an LT ranging up to 10 years, a follow-up period of 25 years or more would be needed to accurately estimate overdiagnosis via the EI approach (16). 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, as these study designs provide a control group of non-screened women (17). Although 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.