Influence of Study Features and Methods on Overdiagnosis Estimates in Breast and Prostate Cancer Screening

Ruth Etzioni, PhD; Roman Gulati, MS; Leslie Mallinger, MPH; and Jeanne Mandelblatt, MD, MPH

Knowledge of the likelihood that a screening-detected case of cancer has been overdiagnosed is vitally important to make treatment decisions and develop screening policy. An overdiagnosed case is an excess case detected by screening. Estimates of the frequency of overdiagnosis in breast and prostate cancer screening vary greatly across studies. This article identifies features of overdiagnosis studies that influence results and shows their effect by using published research. First, different ways to define and measure overdiagnosis are considered. Second, contextual features and how they affect overdiagnosis estimates are examined. Third, the effect of estimation approach is discussed. Many studies use excess incidence under screening as a proxy for overdiagnosis. Others use statistical models to make inferences about lead time or natural history and then derive the corresponding fraction of cases that are overdiagnosed. This article concludes with questions that readers of overdiagnosis studies can use to evaluate the validity and relevance of published estimates and recommends that authors of studies quantifying overdiagnosis provide information about these features.

Ann Intern Med. 2013;158:831-838. www.annals.org

For author affiliations, see end of text.

Screening is a deeply established component of efforts to reduce cancer mortality in the United States. The goal is detection of disease at an early and treatable stage to prevent morbidity and mortality associated with late-stage presentation. Mammography is now a routine element of a woman's health care program, and most older men are screened regularly with prostate-specific antigen (PSA) measurement. However, concern has recently increased about the adverse effects of cancer screening and, in particular, the problem of overdiagnosis.

Overdiagnosis occurs when screening detects a tumor that would not have presented clinically in the absence of screening. Thus, an overdiagnosed case is an excess case because it has only been identified through screening. Treatment of such a case is harmful because it cannot, by definition, improve disease outcomes. Accurately quantifying the frequency of overdiagnosis is important for informed decision making and clinical policy development because the scale of harms relative to benefits determines the individual and societal value of screening.

Overdiagnosis has long been a concern in prostate cancer screening. Early autopsy studies of the prostate revealed a high prevalence of latent prostate cancer among older men (1), and subsequent work (2) indicated that roughly 75% of all cases of prostate cancer are latent and never surface clinically—these findings indicate an enormous pool of persons at risk for overdiagnosis. Concerns about overdiagnosis in breast cancer screening are more recent but have been growing. Several high-profile articles published in the past few years (3–5) have focused attention on the possibility that overdiagnosis in breast cancer screening may be an issue of much greater magnitude than previously recognized.

Knowing how many cases of cancer are overdiagnosed by screening is clearly necessary to inform policy and clinical practice. However, in both prostate and breast cancer literature, estimates of overdiagnosis vary greatly across studies. Estimates of prostate cancer overdiagnosis range from 23% (6) to more than 60% (7) of screening-detected cases. Estimates of breast cancer overdiagnosis are also inconsistent, ranging from 10% or fewer (8, 9) to 30% or more (3, 4, 10). To reach a consensus for breast cancer overdiagnosis, *The Lancet* recently commissioned an independent review of the evidence (11); however, the report concluded that, despite the plethora of studies, further research was still needed to accurately assess the magnitude.

In this article, we examine features of published studies that influence estimates of overdiagnosis. The most influential features are definition (12) and measurement (8), study design and context (13), and estimation approaches (14, 15). Understanding how these features affect results will facilitate proper interpretation and use of published estimates of overdiagnosis by policymakers, clinicians, and patients. Prior reviews and methodology publications (8, 14–16) have discussed the importance of these features and even highlighted how certain choices in study design and execution can bias outcomes, but these have tended to focus on a specific type of cancer (either breast or prostate).

Our goal is not to recapitulate these articles but to integrate studies from the breast and prostate cancer literature to emphasize general principles that go beyond disease-specific considerations. To identify these studies, we searched PubMed between 1 January 1995 and 30 November 2012 using the search terms "overdiagnosis" and either "prostate cancer" or "PSA screening." In parallel, we searched for the terms "overdiagnosis" and either "mammography" or "breast cancer screening." Our search identified 172 publications in the prostate literature and 195 publications in the breast literature, from which we selected examples of quantitative studies to highlight different definitions and measures, study designs and contextual factors, and estimation approaches. **Tables 1** and **2** summarize the selected studies.

© 2013 American College of Physicians

Table 1. **Selected Studies of Overdiagnosis Due to Prostate Cancer Screening***

Study, Year (Reference)	Study Years	Estimate of Overdiagnosis	Measure	Context
Lead-time approach				
Etzioni et al, 2002 (20)	1988–1998	29% in white persons, 44% in black persons	Cases overdiagnosed/screening-detected cases	U.S. SEER9; age 60–84 y; population study
Draisma et al, 2003 (13)	1994–2000	48%	Cases overdiagnosed/screening-detected cases	ERSPC Rotterdam; age 55–67 y; screened every 4 y; PSA cutoff level, 3.0 μg/L; clinical trial
Telesca et al, 2008 (6)	1973–2000	23% in white persons, 34% in black persons	Cases overdiagnosed/screening-detected cases	U.S. SEER9; age 50–84 y; population study
Draisma et al, 2009 (25)	1985–2000	23%, 28%, 42%	Cases overdiagnosed/screening-detected cases	U.S. SEER9; age 50–84 y; population study
		9%, 12%, 19%	Cases overdiagnosed/all detected cases	
Wu et al, 2012 (17)	1996–2005	3.4%	Cases overdiagnosed/total persons screened	ERSPC Finland; age 55–67 y; screened every 4 y; PSA cutoff level, 4.0 μg/L; RCT
Excess-incidence approach†				
Welch and Albertsen, 2009 (34)	1986–2005	1.3 million	Excess cases	U.S. SEER9; age ≥20 y; population study
Schröder et al, 2009 (22)	1991–2006	48	Excess cases/lives saved	ERSPC; age 55–69 y; screened every 2 or 4 y; clinical trial
Pashayan et al, 2006 (7)	1996–2002	40%–64%	Cases overdiagnosed/cases detected by PSA testing	Cambridge, United Kingdom; PSA measurement used as a diagnostic test, not for routine screening; population study
Zappa et al, 1998 (18)	1992–1995	51%	Excess cases/cases in absence of screening	Florence, Italy; age 60 y at entry; 6 biennial screenings; population study

ERSPC = European Randomized Study of Screening for Prostate Cancer; PSA = prostate-specific antigen; RCT = randomized, controlled trial; SEER9 = core 9 catchment areas of the Surveillance, Epidemiology, and End Results program.
* Studies were selected to reflect a range of choices of overdiagnosis measures, contexts, and methods and to include recent, high-profile publications.
† Includes early years of screening use.

Definition of Overdiagnosis

There are 2 major concepts of overdiagnosis in the literature. The first and most commonly used defines "overdiagnosis" as screening-detected cancer that would have remained latent for the remainder of the patient's life in the absence of screening. According to this definition, an overdiagnosed case is a true excess case of cancer, essentially "caused" by screening. Such a case may be biologically indolent and therefore clinically nonprogressive. Alternatively, it may be progressive, but the patient's life expectancy at the time of screening detection may be short enough that death due to other causes occurs before the disease can cause symptoms.

Figure 1 shows that when the interval from screening detection to clinical presentation (diagnosis in the absence of screening), also known as the "lead time," is longer (for example, if the tumor is biologically indolent), there is a greater chance that death from another cause will occur first and, therefore, a higher risk for overdiagnosis. Conversely, for a given lead time, a higher risk for death from another cause (advanced age or poor health) implies a higher risk for overdiagnosis. Estimates of the frequency of overdiagnosis based on this definition generally increase with age at screening detection.

The second concept defines overdiagnosed cancer as only biologically indolent tumors on the basis of their clinical or pathologic characteristics. This definition does not account for life expectancy at the time of screening detection, and the resulting estimates do not exhibit the same age dependency as those based on the first definition.

In general, the 2 definitions may yield very different estimates of overdiagnosis. In the case of prostate cancer, for example, a study (19) that defined overdiagnosis in terms of clinical and pathologic tumor features concluded that most cases of prostate cancer detected by PSA screening were probably clinically significant, but studies that estimated the frequency of overdiagnosis on the basis of the first concept indicated that overdiagnosis is common (**Table 1**). Bach (12) noted a similar divergence of estimates using the different definitions in studies of overdiagnosis due to lung cancer screening. In the rest of this article, we use the first concept, because it is used in most overdiagnosis studies in breast and prostate cancer.

Measuring Overdiagnosis

Studies vary considerably in how the frequency of overdiagnosis is measured and presented (8, 15). de Gelder and colleagues (8) cite no fewer than 7 different measures of the extent of overdiagnosis. The many options arise because estimates of overdiagnosis are generally presented as a ratio, with the numerator being the estimated number of

cases overdiagnosed and with many choices for the denominator. Some studies (6, 20) report overdiagnosis as a fraction of screening-detected cases. Others (4, 21) present the number overdiagnosed as a fraction of the total number of cases detected or the total number invited to screening.

Many studies (5, 10) consider the number of cases overdiagnosed relative to the number of cases expected without screening as an expression of the magnitude of screening-induced excess diagnoses. At least 1 study (22) presents the ratio of overdiagnosed cases to deaths prevented by screening, but this measure effectively conflates screening harm and benefit and is highly sensitive to follow-up duration (23).

Different metrics can produce very different results. de Gelder and colleagues (8) concluded that estimates of overdiagnosis could vary by a factor of 3.5 when different denominators were used. However, some studies present results by using only a single metric, which can make comparison with other studies that use a different metric difficult if not impossible.

Study Design and Context

Study design and context are important sources of variation in published frequencies of overdiagnosis. Here, "study design and context" refers to the type of study (observational or clinical trial), the population used to estimate overdiagnosis, and the diagnostic intensity (fraction of the population tested or who had biopsy) as reflected by the incidence with and without screening.

The type of study is important because it dictates the specific protocol used for screening and whether there is a control group. Screening trials generally implement protocols that require regularly scheduled examinations, whereas opportunistic screening in a population setting may be more variable. The specific screening strategy used can strongly influence the frequency of overdiagnosis (24).

The presence of a control group can be valuable when overdiagnosis is being estimated using information on the excess incidence of cancer under screening because it provides an appropriate baseline for computing the extent of the excess. The absence of a control group necessitates projecting baseline incidence rates of disease in the absence of screening, which presents its own challenges.

The population under study is important because populations vary in terms of underlying disease prevalence and natural history. Puliti and coworkers (15) highlight this point in their review of European studies of overdiagnosis due to mammography screening. Differences in underlying disease risk across populations are a concern when overdiagnosis is being estimated on the basis of excess incidence in screened versus unscreened countries or regions.

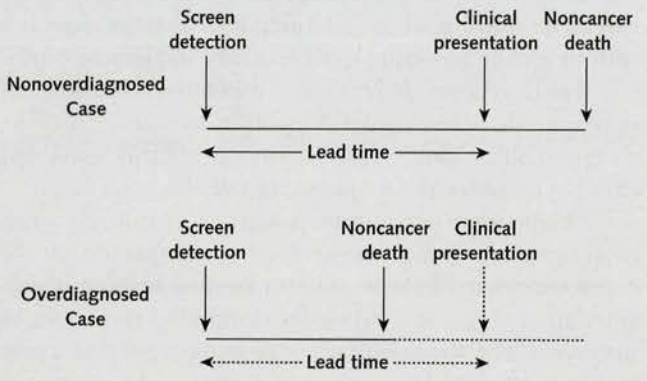
The difference in diagnostic intensity as reflected by incidence with and without screening is important because an overdiagnosed case is an excess case relative to no

Table 2. **Selected Studies of Overdiagnosis Due to Breast Cancer Screening***

Study, Year (Reference)	Study Years	In Situ Cases Included	Estimate of Overdiagnosis	Measure	Context
Excess-incidence approach					
Morrell et al, 2010 (10)	1999–2001	No	30%–42%	Excess cases/cases expected without screening	New South Wales, Australia; age 50–69 y; biennial screening; population study
Gøtzsche and Nielsen, 2011 (3)†	Multiple	Yes	30%	Excess cases/cases expected without screening	7 clinical trials of mammography
Kalager et al, 2012 (5)†	1996–2005	No	15%–25%	Excess cases/cases expected without screening	Norway; age 50–69 y; biennial screening; population study
Bleyer and Welch, 2012 (4)†	1976–2008	Yes	31%	Excess cases/detected cases	U.S. SEER9; age ≥40 y; population study
Independent UK Panel on Breast Cancer Screening, 2012 (11)	Multiple	Yes	11% 3	Excess cases/cases detected in the invited group Excess cases/lives saved	United Kingdom; age 50–70 y; 3 clinical trials of mammography; screened every 3 y
Lead-time approach					
Paci et al, 2006 (9)	1986–2001	Yes No	4.6% 3.2%	Cases overdiagnosed/cases expected without screening	6 areas of Italy; age 50–74 y; biennial screening; population study
Olsen et al, 2006 (21)	1991–1995	Yes	4.8%	Cases overdiagnosed/detected cases	Copenhagen, Denmark; age 50–69 y; biennial screening (first 2 rounds); population study
de Gelder, 2011 (8)	1990–2006	Yes	8.9% 4.6% 5%	Cases overdiagnosed/screening-detected cases Cases overdiagnosed/detected cases Cases overdiagnosed/cases expected without screening	The Netherlands; age 40–69 y; biennial screening; population study

SEER9 = core 9 catchment areas of the Surveillance, Epidemiology, and End Results program.
* Studies were selected to reflect a range of choices of overdiagnosis measures, contexts, and methods and to include recent, high-profile publications.
† Includes early years of screening use.

Figure 1. **Link between overdiagnosis and lead time.**



A nonoverdiagnosed case (*top*) is a screening-detected case that would have presented clinically before dying of causes not related to cancer. An overdiagnosed case (*bottom*) is a screening-detected case that would have died of causes not related to cancer before presenting clinically (i.e., a screening-detected case that would not have been diagnosed without screening). The lead time is the time from screening detection to clinical diagnosis. The longer the lead time, the greater the chance of overdiagnosis. Similarly, the higher the risk for death from another cause, the greater the chance of overdiagnosis.

screening; lower incidence in the absence of screening creates a larger pool of latent cases in the population and a greater potential for overdiagnosis. Therefore, higher incidence in the presence of screening due to better adherence to screening invitations or biopsy recommendations increases the reach into the latent pool and, with it, the likelihood of overdiagnosis.

In the case of prostate cancer screening, design and contextual factors cause much of the difference between overdiagnosis estimates from the U.S. population and from the European population based on the European Randomized Study of Screening for Prostate Cancer (ERSPC), 1 of the 2 large randomized trials of prostate cancer screening. Draisma and colleagues (13, 25) found that overdiagnosis among men aged 50 to 84 years was 66% based on a model developed for the Rotterdam section of the ERSPC but was 42% when the same model was adjusted to reflect incidence in the United States (26).

There are several identifiable contextual reasons for this difference. First, prostate cancer incidence was considerably higher in the United States than in ERSPC centers before screening was introduced (27). Thus, at the start of the ERSPC there was a relatively greater pool of latent cases with the potential to be detected by screening than in the U.S. population. Second, participants in the screening group of the ERSPC were highly adherent to the trial protocol; adherence to biopsy referral was 86% on average across ERSPC centers (22) compared with approximately 40% in the United States (28). Furthermore, the ERSPC centers generally had a lower threshold for referral to biopsy (PSA level >3.0 μg/L) than was standard in the United States. These differences in contextual factors produced a greater potential for and a correspondingly higher estimate of the frequency of overdiagnosis in the ERSPC compared with the U.S. population setting.

Estimation Approaches

The definition of an overdiagnosed case as an excess or extra diagnosis leads to estimation approaches based on the excess incidence of disease in the presence of screening. There are 2 main approaches. The first uses the observed excess incidence—the difference between incidence in the presence and in the absence of screening—as a proxy for overdiagnosis. We refer to this as the "excess-incidence approach." The second uses disease incidence under screening to make inferences about the lead time or the natural history of the disease and estimates the corresponding frequency of overdiagnosis. We refer to this as the "lead-time approach." **Figure 2**, which is adapted from a study by Feuer and Wun (29) that linked patterns of breast cancer incidence with lead time and overdiagnosis, shows the motivation for each approach.

Figure 2 presents 3 scenarios in which screening is introduced into a population with a background annual incidence rate of 100 cases per 100 000 persons. In the first scenario, the screening test has a 1-year lead time; in the second and third scenarios, the screening test has a 2-year lead time. Scenario 3 is the only one in which there is overdiagnosis, with 3 of every 13 (23%) screening-detected cases being overdiagnosed.

In each scenario, the introduction of screening generates a peak in disease incidence followed by a decline. The initial part of the incidence peak arises from a series of incidence "gains" caused by screening detection of cases from the latent incidence pool. Incidence gains are greatest in the beginning and then reach a steady state as the latent pool declines and screening use in the population stabilizes. The initial gains are followed by a pattern of incidence "deficits," because those cases whose diagnosis was advanced by screening are no longer present to be detected. The lead time determines the interval between the initial gains and the subsequent deficits and, consequently, the magnitude of the corresponding incidence peak.

For a disease with a given latent prevalence and a test with a given sensitivity, a longer lead time causes the incidence peak to be higher and wider. Regardless of the lead time, if there is no overdiagnosis, incidence eventually returns to its original background level as annual gains and deficits equalize. In **Figure 2**, the background incidence is constant, but in reality the level may change over time. This figure shows that the excess incidence under screening is highly informative about overdiagnosis and lead time; both the excess-incidence and lead-time approaches to estimation are based on this observation. However, this figure also shows that these approaches have important caveats and limitations.

The excess-incidence approach is the predominant one used in the breast cancer literature, and the methodological considerations of this approach have been the topic of several recent reviews (14–16). First, it may yield a biased result, particularly if the early years of screening dissemination are included. This is because, as scenario 3 in **Figure 2** shows, the excess incidence in the early years of screening consists of a mixture of overdiagnosed and nonoverdiagnosed cases and therefore overestimates the frequency of overdiagnosis. In this figure, the excess incidence estimate of the fraction overdiagnosed among screening-detected cases amounts to 40% (47/117) when the first 6 years of the screening program is included but 23% (6/26)—the correct answer—when the estimate is restricted to the last 2 years. de Gelder and associates (8) similarly estimated overdiagnosis on the basis of breast cancer incidence within intervals relative to the implementation of screening in the Netherlands and found that estimates based on the years before the program was fully implemented were 4 times higher than those based on the years afterward. Thus, it is important for excess-incidence estimates of overdiagnosis to be appropriately timed relative to the dissemination of screening.

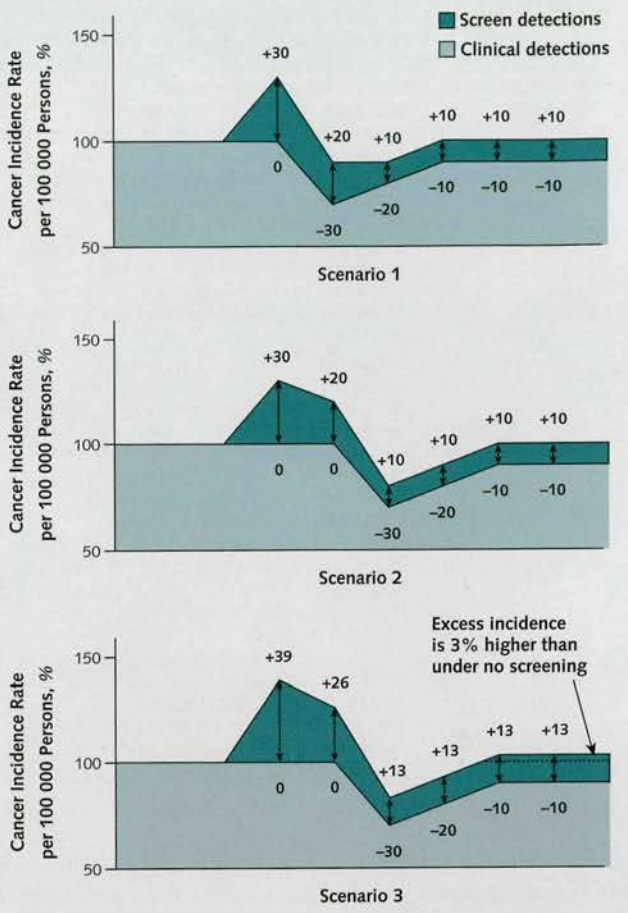
Second, computing the incidence under screening only on the basis of age groups eligible for screening will overestimate excess cases because it will not account for deficits in older age groups due to screening detection at younger ages. Thus, excess incidence calculations should account for what is commonly termed a "compensatory drop" in incidence at older ages.

Third, the baseline or control incidence trend must reflect the incidence for the screened population that would be expected in the absence of screening. In **Figure 2**, the baseline incidence is set to a constant level, but in reality this is frequently an unreasonable assumption. It can be challenging to project baseline incidence correctly when it is not observable (for example, through a concurrent control group from the same population). In this case, baseline incidence may be based on historical trends (5, 10) or concurrent trends among age groups that are not eligible for screening (4, 10).

In countries where screening programs have been implemented in some regions but not others, baseline incidence may be based on the trends in the regions not offered screening (5, 30). In some cases, trends or changes in disease risk factors (for example, over time or across regions) may have to be assessed and used to adjust estimated baseline incidence. Further, to account for lead time, the observed incidence in a screened population of age A should theoretically be compared with baseline incidence in a population of age $A + L$, where L is the lead time for that age.

Variation across studies in how the baseline incidence is estimated can substantially affect results. Even in a single study, the effects of different choices in the modeling of baseline incidence can be considerable. As an example,

Figure 2. **Incidence trends after the introduction of screening in 3 scenarios with the same baseline incidence (constant, 100 cases per 100 000 persons per year), disease prevalence, test sensitivity, and screening test dissemination.**



For each scenario, the number of cases per 100 000 screening-detected cases each year is preceded by a plus sign. The number of deficits in incidence each year due to earlier detection by screening is preceded by a minus sign. *Scenario 1.* Lead time equals 1 year, no overdiagnosis. *Scenario 2.* Lead time equals 2 years, no overdiagnosis. *Scenario 3.* Lead time (among nonoverdiagnosed cases) equals 2 years, with 3 screening-detected cases overdiagnosed per 13 screening-detected cases (23%). Adapted from reference 29.

Morrell and coworkers (10) used 2 methods to project expected breast cancer incidence in the absence of screening between 1999 and 2001, after organized mammography screening had become well-established in New South Wales, Australia.

The first method interpolated between the oldest and youngest age groups in the population who did not undergo screening. The second method extrapolated prescreening incidence trends, accounting for changes in risk factors for breast cancer, such as obesity and use of hormone replacement therapy. Corresponding overdiagnosis estimates expressed relative to the projected incidence

Table 3. **Questions That Readers of Overdiagnosis Studies Should Ask and That Writers of Overdiagnosis Studies Should Address to Facilitate Better Understanding and Comparison of Results Across Studies**

Study Feature	Questions
Definition and measurement	1. How does this study define overdiagnosis? 2. How does this study measure overdiagnosis? What is the numerator, and what is the denominator?
Study design and context	3. Where was the study conducted? 4. In what years was the study conducted? 5. What were the characteristics of the study population that might affect overdiagnosis (e.g., age range, risk factors for disease, and general health status)? 6. What was the screening protocol, and what were the criteria for biopsy referral? 7. How adherent were participants to screening and biopsy protocols? 8. What was the incidence in the absence of screening?
Estimation approaches	9. What estimation approach was used (excess incidence or lead time)? 10. If lead-time approach: a. Are model assumptions clearly stated and justified? b. Does model-projected incidence under screening reasonably match observed incidence? c. What assumptions are made about concurrent disease incidence without screening if this is not observed? 11. If excess-incidence approach: a. When is the study conducted relative to the start of screening? Does it include the early years of screening? b. How is concurrent control incidence projected if this is not observed? c. If screening is limited to specific age groups, does estimation of excess incidence account for compensatory drop at older ages? d. Does calculation account for lead-time differential between screened and control incidence?

without screening were 42% for the first method and 30% for the second method, indicating that the baseline incidence corresponding to the interpolation method was lower than that corresponding to the extrapolation method.

Rather than using empirical differences between observed and baseline incidence, the lead-time approach uses modeling techniques to infer the lead time and the corresponding fraction of cases overdiagnosed from the pattern of excess incidence under screening. This approach has the advantage that it can be applied to data from the beginning of screening dissemination. However, like the excess-incidence approach, it requires an estimate of the expected baseline incidence in the absence of screening, and results can be sensitive to this estimate. In addition, this approach requires knowledge of screening dissemination and practice patterns in the population.

Although the statistical literature provides theoretical underpinnings (31, 32) and some standard methods (33), there are many ways to estimate lead time and overdiagnosis by modeling. Some studies use observed incidence under screening to estimate elaborate models of the underlying progression of disease, effectively imputing times of disease onset, progression to metastasis, and transitions from a latent to a symptomatic state (8, 13). In these models, the baseline incidence is generally estimated along with the underlying disease progression. Then, the fraction of overdiagnosed cases given specified screening patterns is calculated empirically from the imputed disease histories and age-specific risks for death from other causes. Other studies (6, 20, 21) simply focus on estimating the lead time and frequency of overdiagnosis that are most consistent with the observed trends in disease incidence under screening.

Like the excess-incidence studies, choices made about model structure and assumptions affect results. For example, Draisma and coworkers (25) used 3 models to estimate lead time and overdiagnosis corresponding to the peak in prostate cancer incidence observed in the early 1990s. The models used similar assumptions about the background trend in incidence but had different underlying structures. The resulting overdiagnosis estimates ranged from 23% to 42% of screening-detected cases.

Tables 1 and **2** show examples of prostate and breast cancer studies that used the lead-time and excess-incidence approaches. In the case of prostate cancer (**Table 1**), estimates derived using the excess-incidence approach are generally considerably higher than those derived using the lead-time approach. However, it is difficult to say whether the higher estimates are due to context or estimation method because most excess-incidence studies were conducted in non-U.S. settings and most lead-time studies were based on U.S. data.

The single excess-incidence study based in the United States that we reviewed reported that 1.3 million cases had been overdiagnosed from 1987 (when PSA screening began) through 2005 (34). This corresponds to a frequency of overdiagnosis of 37% among all detected cases based on inflated counts of prostate cancer incidence in the U.S. SEER (Surveillance, Epidemiology, and End Reports) population. However, this study included the early years of PSA screening, which probably inflated the results. In contrast, the lead-time approach used by Draisma and colleagues (25) resulted in a range of overdiagnosis from 9% to 19% among all detected cases, corresponding to an absolute number of overdiagnosed cases that is at most about half of the 1.3 million (34). **Table 2** shows an even clearer dichotomy between excess-incidence and lead-time studies in breast cancer screening, with the latter studies generating overdiagnosis estimates that are markedly lower than those of the former.

DISCUSSION

Our examination of variation in study features and methods leads us to wonder whether it is possible to compare and integrate results across published studies of overdiagnosis. It is clear that the conceptual and analytic choices made by study investigators can dramatically affect overdiagnosis estimates. For consumers of the overdiagnosis literature, therefore, a necessary step is to understand

what choices were made. **Table 3** shows questions that readers of overdiagnosis studies may ask to clarify these choices to better understand study results.

The first set of questions addresses the definition of overdiagnosis used in the study and the measure of its frequency. The second set addresses contextual factors. To shed light on differences in diagnostic intensity without versus with screening, information about diagnostic practices in the absence and presence of screening will be valuable. The final set of questions addresses estimation approach, which is undoubtedly the most complex feature that we have considered and requires careful examination.

The main limitation of the excess-incidence approach is that observed excess incidence is not an unbiased estimate of the incidence of overdiagnosis. Ad hoc adjustments often need to be applied to the empirical measures, and understanding these adjustments is key to evaluating these studies. The main limitation of the lead-time approach is that the links among model choices, assumptions, and results are often not transparent, and this can make evaluation of these studies difficult. Prior publication of the model in the peer-reviewed statistics or biostatistics literature can be a strong positive indicator of model validity, and ongoing efforts (35) aim to improve and standardize model reporting in the interests of greater transparency.

Given that there are identifiable features of overdiagnosis studies that influence and even bias results, what does this mean for policymakers, clinicians, and patients? First, knowledge of these features should help all consumers of overdiagnosis studies to avoid using clearly biased estimates. Second, that studies use different measures of overdiagnosis should direct consumers to the ones that most meet their needs. Policymakers may want to focus on those that present results in terms of the number of overdiagnoses per invited participant; a patient newly diagnosed after a screening test may be more interested in overdiagnoses expressed as a fraction of screening-detected cases.

Recognition of the importance of contextual factors should enable consumers to select studies that are most suited to their setting and screening protocol. For example, in consideration of prostate cancer screening policies for the U.S. population setting, it would not be appropriate to use overdiagnosis estimates from a trial conducted in Europe with a PSA cutoff level that is lower than that typically used in the United States and with a much higher rate of adherence to biopsy referral. Finally, knowledge of the limitations of the different approaches may help with selecting estimates on the basis of the approach that uses relevant data sources and makes clinically reasonable assumptions.

In conclusion, we remain far from a consensus about how best to estimate the likelihood of overdiagnosis. Our goal has not been to make conclusions about the frequency of overdiagnosis in breast and prostate cancer screening but to help consumers of publications on overdiagnosis navigate this growing, confusing, and often controversial literature. Our focus on overdiagnosis as a harm of screening has precluded discussion of screening benefit, an equally controversial topic and one in which study features and methods also almost certainly influence results.

We encourage investigators publishing overdiagnosis studies to ensure that their reports address the questions in **Table 3** and, if possible, include the numbers needed to translate results across commonly used metrics. Doing so will provide the transparency to adequately compare and integrate studies of what is possibly the most important potential harm of screening.

From Fred Hutchinson Cancer Research Center, Seattle, Washington, and Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, DC.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute, the National Institutes of Health, or the Centers for Disease Control and Prevention.

Financial Support: By award numbers U01CA157224 and U01CA088283 and U01CA152958 from the National Cancer Institute and the Centers for Disease Control and Prevention. Additional funding was provided by award numbers KO5CA96940 and P01CA154292.

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M12-3000.

Requests for Single Reprints: Ruth Etzioni, PhD, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024; e-mail, retzioni@fhcrc.org.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Carter HB, Piantadosi S, Isaacs JT.** Clinical evidence for and implications of the multistep development of prostate cancer. J Urol. 1990;143:742-6. [PMID: 2313798]

2. **Etzioni R, Cha R, Feuer EJ, Davidov O.** Asymptomatic incidence and duration of prostate cancer. Am J Epidemiol. 1998;148:775-85. [PMID: 9786232]

3. **Gøtzsche PC, Nielsen M.** Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2011:CD001877. [PMID: 21249649]

4. **Bleyer A, Welch HG.** Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med. 2012;367:1998-2005. [PMID: 23171096]

5. **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;156:491-9. [PMID: 22473436]

6. **Telesca D, Etzioni R, Gulati R.** Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. Biometrics. 2008;64:10-9. [PMID: 17501937]

7. **Pashayan N, Powles J, Brown C, Duffy SW.** Excess cases of prostate cancer and estimated overdiagnosis associated with PSA testing in East Anglia. Br J Cancer. 2006;95:401-5. [PMID: 16832417]

8. **de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ.** Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol Rev. 2011;33:111-21. [PMID: 21709144]

9. **Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, Falcini F, et al.** Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead

time. A service screening study in Italy. Breast Cancer Res. 2006;8:R68. [PMID: 17147789]

10. **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;21:275-82. [PMID: 19894130]

11. **Independent UK Panel on Breast Cancer Screening.** The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380:1778-86. [PMID: 23117178]

12. **Bach PB.** Overdiagnosis in lung cancer: different perspectives, definitions, implications [Editorial]. Thorax. 2008;63:298-300. [PMID: 18364441]

13. **Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RA, Schröder FH, et al.** Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003;95:868-78. [PMID: 12813170]

14. **Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L.** Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. Lancet Oncol. 2007;8:1129-38. [PMID: 18054882]

15. **Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, et al; EUROSCREEN Working Group.** Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen. 2012;19 Suppl 1:42-56. [PMID: 22972810]

16. **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. [PMID: 18766185]

17. **Wu GH, Auvinen A, Määttänen L, Tammela TL, Stenman UH, Hakama M, et al.** Number of screens for overdetection as an indicator of absolute risk of overdiagnosis in prostate cancer screening. Int J Cancer. 2012;131:1367-75. [PMID: 22052356]

18. **Zappa M, Ciatto S, Bonardi R, Mazzotta A.** Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence Screening Pilot Study. Ann Oncol. 1998;9:1297-300. [PMID: 9932159]

19. **Humphrey PA, Keetch DW, Smith DS, Shepherd DL, Catalona WJ.** Prospective characterization of pathological features of prostatic carcinomas detected via serum prostate specific antigen based screening. J Urol. 1996;155:816-20. [PMID: 8583583]

20. **Etzioni R, Penson DF, Legler JM, di Tommaso D, Boer R, Gann PH, et al.** Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. J Natl Cancer Inst. 2002;94:981-90. [PMID: 12096083]

21. **Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW.** Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. Breast J. 2006;12:338-42. [PMID: 16848843]

22. **Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al; ERSPC Investigators.** Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360:1320-8. [PMID: 19297566]

23. **Gulati R, Mariotto AB, Chen S, Gore JL, Etzioni R.** Long-term projections of the harm-benefit trade-off in prostate cancer screening are more favorable than previous short-term estimates. J Clin Epidemiol. 2011;64:1412-7. [PMID: 22032753]

24. **Gulati R, Gore JL, Etzioni R.** Comparative effectiveness of alternative prostate-specific antigen–based prostate cancer screening strategies: model estimates of potential benefits and harms. Ann Intern Med. 2013;158:145-53. [PMID: 23381039]

25. **Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, et al.** Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst. 2009;101:374-83. [PMID: 19276453]

26. **Wever EM, Draisma G, Heijnsdijk EA, Roobol MJ, Boer R, Otto SJ, et al.** Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. J Natl Cancer Inst. 2010;102:352-5. [PMID: 20142584]

27. **Shibata A, Ma J, Whittemore AS.** Prostate cancer incidence and mortality in the United States and the United Kingdom. J Natl Cancer Inst. 1998;90:1230-1. [PMID: 9719085]

28. **Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK; Prostate, Lung, Colorectal and Ovarian Project Team.** Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. J Urol. 2005;173:746-50; discussion 750-1. [PMID: 15711261]

29. **Feuer EJ, Wun LM.** How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization? A dynamic population model approach. Am J Epidemiol. 1992;136:1423-36. [PMID: 1288272]

30. **Jørgensen KJ, Zahl PH, Gøtzsche PC.** Overdiagnosis in organised mammography screening in Denmark. A comparative study. BMC Womens Health. 2009;9:36. [PMID: 20028513]

31. **Davidov O, Zelen M.** Overdiagnosis in early detection programs. Biostatistics. 2004;5:603-13. [PMID: 15475422]

32. **Pinsky PF.** Estimation and prediction for cancer screening models using deconvolution and smoothing. Biometrics. 2001;57:389-95. [PMID: 11414561]

33. **Walter SD, Day NE.** Estimation of the duration of a pre-clinical disease state using screening data. Am J Epidemiol. 1983;118:865-86. [PMID: 6650488]

34. **Welch HG, Albertsen PC.** Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. J Natl Cancer Inst. 2009;101:1325-9. [PMID: 19720969]

35. **Caro JJ, Briggs AH, Siebert U, Kuntz KM; ISPOR-SMDM Modeling Good Research Practices Task Force.** Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Med Decis Making. 2012;32:667-77. [PMID: 22990082]

Copyright © American College of Physicians 2013.