



Published in final edited form as:

Ann Intern Med. 2013 June 4; 158(11): 831–838. doi:10.7326/0003-4819-158-11-201306040-00008.

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

Ruth Etzioni^{1,*}, Roman Gulati¹, Leslie Mallinger¹, and Jeanne Mandelblatt²

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center

²Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine

Abstract

Knowledge of the likelihood that a screen-detected cancer case has been overdiagnosed is vitally important for treatment decision making and screening policy development. An overdiagnosed case is an excess case detected because of cancer screening. Estimates of the frequency of overdiagnosis in breast and prostate cancer screening are highly variable across studies. In this article we identify features of overdiagnosis studies that influence results and illustrate their impact using published studies. We first consider different ways to define and measure overdiagnosis. We then examine contextual features and how they affect overdiagnosis estimates. Finally, we discuss the effect of estimation approach. 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. We conclude with a list of questions that readers of overdiagnosis studies can use to evaluate the validity and relevance of published estimates and recommend that authors of publications quantifying overdiagnosis provide information about these features of their studies.

Introduction

Cancer screening is a deeply established component of efforts to reduce cancer mortality in the US. The goal of screening is the detection of disease at an early and treatable stage to prevent morbidity and mortality associated with late stage disease presentation. Mammograms are now routine elements of a woman's health care program and the majority of older men are screened regularly with PSA. However, in recent years there has been growing concern 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 in the sense that it is only identified because of 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 since 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 prostate autopsy studies revealed a high prevalence of latent prostate cancer among older men (1), and subsequent work (2) indicated that roughly 75% of all prostate cancers are latent and never

*Correspondence to: Ruth Etzioni, PhD, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024; Tel: +1.206.667.6561; Fax: +1.206.667.7264; retzioni@fhcrc.org.

surface clinically, indicating an enormous pool of individuals at risk of overdiagnosis. Concerns about overdiagnosis in breast cancer screening are more recent but have been growing. Several high-profile articles published in the last 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.

Clearly, knowing how many cancers are overdiagnosed by screening is necessary to inform policy and clinical practice. However, in both the prostate and breast cancer literatures, overdiagnosis estimates are highly variable across studies. In prostate cancer, estimates range from as low as 23% (6) to over 60% (7) of screen-detected cases. In breast cancer, estimates are also inconsistent, ranging from 10% or fewer (8, 9) to 30% or more (3, 4, 10) cases overdiagnosed. To reach a consensus for breast cancer, The Lancet recently commissioned an independent review of the evidence (11), but the resulting report concluded that, despite the plethora of studies, further research was still needed to accurately assess the magnitude of overdiagnosis.

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 policy makers, 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 cancer (either breast or prostate). Our goal is not to recapitulate these articles; rather, we 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 from January 1, 1995, to November 30, 2012, using the search terms “overdiagnosis” and either “prostate cancer” or “PSA screening.” In parallel, we also searched on 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. The selected studies are summarized in Tables 1 and 2.

Definition of overdiagnosis

There are two major concepts of overdiagnosis in the literature. The first and most commonly used defines overdiagnosis as a screen-detected cancer that would have remained latent for the remainder of the patient’s lifetime 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 cancer may be biologically indolent and therefore clinically non-progressive. Alternatively, it may be progressive, but the patient’s life expectancy at the time of screen detection may be short enough that death due to other causes occurs before the disease can cause symptoms. Figure 1 shows that when the time interval from screen detection to clinical presentation (diagnosis in the absence screening), also known as the lead time, is longer (for instance, if the tumor is biologically indolent), there is a greater chance that other-cause death will occur first and therefore a higher risk of overdiagnosis. Conversely, for a given lead time, a higher risk of other-cause death (due to advanced age or poor health), implies a higher risk of overdiagnosis. Estimates of the frequency of overdiagnosis based on this definition generally increase with age at screen detection.

The second concept defines as overdiagnosed only biologically indolent tumors, based on their clinical and/or pathologic characteristics. This definition does not take into account life expectancy at the time of screen detection, and the resulting estimates do not exhibit the

same age dependency as those based on the first definition. In general, the two definitions may yield very different estimates of overdiagnosis. In the case of prostate cancer, for example, a study (17) that defined overdiagnosis in terms of clinical and pathologic tumor features concluded that most prostatic cancers detected by PSA screening were likely to be clinically significant, but studies that estimated the frequency of overdiagnosis based on the first concept indicated that it is not uncommon (Table 1). A similar divergence of estimates using the different definitions was noted by Bach in studies of overdiagnosis due to lung cancer screening (12).

In the rest of this article we use the first definition since this is the one used in most studies of overdiagnosis in breast and prostate cancer.

Measuring overdiagnosis

Studies vary considerably in how the frequency of overdiagnosis is measured and presented (8, 15). De Gelder et al (8) cite no fewer than seven 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, 18) report overdiagnosis as a fraction of screen-detected cases. Others (4, 19) present the number overdiagnosed as a fraction of the total number of cases detected or the total number invited to screening. Many (5, 10) consider the number overdiagnosed relative to the number of cases expected without screening as an expression of the magnitude of screening-induced excess diagnoses. At least one study (20) 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 (21). Different metrics can produce very different results. De Gelder et al (8) concluded that estimates of overdiagnosis could vary by a factor of 3.5 when different denominators were used. However some studies present results using only a single metric, and this can make comparison with other studies that use a different metric difficult if not impossible.

Study design and context

The study design and context are important sources of variation in published overdiagnosis frequencies. Here, study design and context refers to: (1) the type of study (observational or clinical trial), (2) the population used to estimate overdiagnosis, and (3) the diagnostic intensity (fraction of the population tested and/or biopsied) as reflected by the incidence with and without screening.

The type of study is important because it will dictate the specific protocol used for screening and whether there is a concurrent control group. Screening trials generally implement protocols calling for 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 (22). The presence of a concurrent control group can be of great value when estimating overdiagnosis using information on the excess incidence of cancer under screening since it will provide an appropriate baseline for computing the extent of the excess. The absence of a control group necessitates projecting what baseline disease incidence rates would have been in the absence of screening and this presents its own set of challenges.

The population under study is important because populations vary in terms of their underlying disease prevalence and natural history. Puliti et al (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 estimating overdiagnosis based on 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 screening; lower incidence in the absence of screening creates a larger latent pool of cases in the population and a greater potential for overdiagnosis. Consequently, higher incidence in the presence of screening due to more complete compliance with screening invitations or better adherence to biopsy recommendations will increase the reach into the latent pool and with it the likelihood of overdiagnosis.

In the case of prostate cancer screening, design and contextual factors are responsible for much of the difference between overdiagnosis estimates from the US population and from the European population based on the European Randomized Study of Screening for Prostate Cancer (ERSPC), one of the two large randomized trials of prostate cancer screening. Draisma and colleagues (13, 23) found that overdiagnosis among men aged 50–84 was 66% based on a model developed for the Rotterdam section of the ERSPC but 42% when the same model was adjusted to reflect incidence in the US setting (24). There are several identifiable contextual reasons for this difference. First, prostate cancer incidence was considerably higher in the US than in ERSPC centers before screening was introduced (25). Thus, at the start of the ERSPC trial, there was a relatively greater pool of latent cases with the potential to be screen detected than in the US population. Second, participants on the screening arm of the ERSPC were highly compliant with the trial protocol; compliance with biopsy referral was 86% on average across ERSPC centers (20) compared with approximately 40% in the US (26). Furthermore, the ERSPC centers generally had a lower threshold for referral to biopsy (PSA > 3 µg/L) than was standard in the US. 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 US population setting.

Estimation approaches

The definition of an overdiagnosed case as an excess or extra diagnosis gives rise to estimation approaches based on the excess incidence of disease in the presence of screening. There are two main approaches. The first uses the observed excess incidence—the difference between incidence in the presence and incidence 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. The motivation for each approach is shown in Figure 2, which is loosely based on a study by Feuer and Wun (27) that linked patterns of breast cancer incidence with lead time and overdiagnosis.

Figure 2 presents three scenarios in which screening is introduced into a population with a background annual incidence rate of 100 cases per 100,000 individuals. In the first scenario, the screening test has a 1-year lead time and 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%) screen-detected cases being overdiagnosed. In each scenario, the introduction of screening generates a peak in disease incidence, which is followed by a decline in incidence. The initial part of the incidence peak arises from a series of incidence “gains” caused by screen 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 as 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 the figure, the background incidence is constant, but in reality the level may well change over time.

Figure 2 demonstrates 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, the figure also shows that both the excess-incidence and lead-time approaches have important caveats and limitations.

The excess-incidence approach is the predominant approach used in the breast cancer literature, and the methodological considerations of this approach have been the topic of several recent reviews (14–16). This approach may yield a biased result, particularly if the early years of screening dissemination are included. This is because, as shown in the third panel of the figure, the excess incidence in the early years of screening consists of a mixture of overdiagnosed and non-overdiagnosed cases and therefore will overestimate the frequency of overdiagnosis. In this figure, the excess incidence estimate of the fraction overdiagnosed among screen-detected cases amounts to 40% (47/117) when including the first six years of the screening program, but 23% (6/26)—the correct answer—when restricting the estimate to the last two years. Similarly, De Gelder et al (8) estimated overdiagnosis based on breast cancer incidence within different intervals relative to the implementation of screening in the Netherlands and found that estimates based on years before the program was fully implemented were four times higher than those based on years after this point. Thus, it is important for excess-incidence estimates of overdiagnosis to be appropriately timed relative to the dissemination of screening.

Second, if the incidence under screening is computed based only on age groups eligible for screening, it will overestimate excess cases because it will not account for deficits in older age groups due to screen detection while at younger ages. Thus, excess incidence calculations must 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 the figure, 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 (e.g., via 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 not eligible for screening (4, 10). In countries where screening programs have been implemented in some regions but not in others, baseline incidence may be based on the trends in the regions not offered screening (5, 28). In some cases trends or changes in disease risk factors (e.g. 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 significantly impact results. Even within a single study the effects of different choices in the modeling of baseline incidence can be considerable. As an example, Morrell et al (10) used two different methods to project expected breast cancer incidence in the absence of screening in 1999–

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 pre-screening incidence trends accounting for changes in breast cancer risk factors such as obesity and use of hormone replacement therapy. Corresponding overdiagnosis estimates expressed relative to the projected incidence without screening were 42% and 30%, indicating that the baseline incidence corresponding to the interpolation method was lower than the baseline incidence 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.

While the statistical literature provides theoretical underpinnings (29, 30) and some standard methods (31), there are many different ways to estimate lead time and overdiagnosis by modeling. Some modeling 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 symptomatic state (8, 13). In these models, the baseline incidence is generally estimated along with the underlying disease progression. Then, the fraction overdiagnosed given specified screening patterns is calculated empirically from the imputed disease histories and age-specific risks of other-cause death. Other studies (6, 18, 19) simply focus on estimating the lead time and overdiagnosis frequency that are most consistent with the observed trends in disease incidence under screening.

Like the excess-incidence studies, choices made regarding model structure and assumptions will affect results. As an example, Draisma et al (23) used three different 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 quite different underlying structures. The resulting overdiagnosis estimates ranged from 23% to 42% of screen-detected cases.

Tables 1 and 2 provide 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. It is difficult, however, to say whether the higher estimates are due to context or estimation method, since most excess-incidence studies were conducted in non-US settings, while most lead-time studies were based on US data. The single US-based excess-incidence study that we reviewed reported that 1.3 million cases had been overdiagnosed from 1987 (when PSA screening began) through 2005 (32). This corresponds to an overdiagnosis frequency of 37% among all detected cases based on inflating counts of prostate cancer incidence in SEER to the US population. However, this study included the early years of PSA screening which likely inflated the results. In contrast, the lead-time approach used by Draisma et al (23) results in a range from 9% to 19% overdiagnosed among all detected cases, corresponding to an absolute number overdiagnosed that is at most about half of the 1.3 million (32). Table 2 shows an even clearer dichotomy between excess-incidence and lead-time studies in breast cancer screening, with the lead-time studies

generating overdiagnosis estimates that are markedly lower than those from the excess-incidence studies.

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. Clearly the conceptual and analytic choices made by study investigators can dramatically impact overdiagnosis estimates. For consumers of the overdiagnosis literature, therefore, a necessary step is to understand what choices were made. In Table 3 we propose a list of questions (Table 3) that readers of overdiagnosis studies may ask to clarify these choices so as 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 of questions 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 of value. The final set of questions addresses estimation approach. This is undoubtedly the most complex of the features 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. Often, ad-hoc adjustments 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 between 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 (33) aim to improve and standardize model reporting in the interests of greater transparency.

Given that there are identifiable features of overdiagnosis studies that will influence and even bias results, what does this mean for policy makers, clinicians, and patients? First, knowledge of these features should help all consumers of overdiagnosis studies to avoid using clearly biased estimates. Second, the fact that studies use different measures of overdiagnosis should direct consumers to the ones that most meet their needs. Policy makers may want to focus on studies that present results in terms of the number of overdiagnoses per invited participant; a patient newly diagnosed following a screening test may be more interested in overdiagnoses expressed as a fraction of screen-detected cases. Recognition of the importance of contextual factors should enable consumers to select those studies that are most suited to their setting and their screening protocol. For example, in considering prostate cancer screening policies for the US population setting, it will not be appropriate to use overdiagnosis estimates from a trial conducted in Europe with a PSA cutoff that is lower than that typically used in this country and with a much higher rate of compliance with biopsy referral. Finally, knowledge of the limitations of the different approaches may help with selecting estimates based on the approach that uses relevant data sources and makes clinically reasonable assumptions.

In conclusion, we remain far from a consensus regarding 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 overdiagnosis publications 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 across studies of what is possibly the most important potential harm of screening.

Acknowledgments

Financial support: This work was supported by Award Numbers U01CA157224 (RE, RG, LM) and U01CA088283 and U01CA152958 (JSM) from the National Cancer Institute and the Centers for Disease Control and Prevention. Additional funding provided by Award Numbers KO5CA96940 and P01CA154292 (JSM). 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.

References

1. Carter HB, Piantadosi S, Isaacs JT. Clinical evidence for and implications of the multistep development of prostate cancer. *J Urol.* 1990; 143(4):742–6. [PubMed: 2313798]
2. Etzioni R, Cha R, Feuer EJ, Davidov O. Asymptomatic incidence and duration of prostate cancer. *Am J Epidemiol.* 1998; 148(8):775–85. [PubMed: 9786232]
3. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane database of systematic reviews (Online).* 2011; (1):CD001877.
4. Bleyer A, Welch HG. Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence. *New England Journal of Medicine.* 2012; 367(21):1998–2005. [PubMed: 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. *Annals of Internal Medicine.* 2012; 156(7):491–9. [PubMed: 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(1):10–9. [PubMed: 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. *British Journal of Cancer.* 2006; 95(3): 401–5. [PubMed: 16832417]
8. de Gelder R, Heijnsdijk EAM, van Ravesteijn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiologic Reviews.* 2011; 33(1):111–21. [PubMed: 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 research: BCR.* 2006; 8(6)
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 and Control.* 2010; 21(2):275–82. [PubMed: 19894130]
11. Independent UKPoBCS. The benefits and harms of breast cancer screening: an independent review. *Lancet.* 2012; 380(9855):1778–86. [PubMed: 23117178]
12. Bach PB. Overdiagnosis in lung cancer: different perspectives, definitions, implications. *Thorax.* 2008; 63(4):298–300. [PubMed: 18364441]
13. Draisma G, Boer R, Otto SJ, van der Crujisen IW, Damhuis RA, Schröder FH, et al. Lead times and overdiagnosis due to prostate-specific antigen screening: Estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst.* 2003; 95(12):868–78. [PubMed: 12813170]
14. 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(12):1129–38. [PubMed: 18054882]
15. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *Journal of Medical Screening.* 2012; 19 (Suppl 1):42–56. [PubMed: 22972810]

16. Duffy SW, Lyng E, Jonsson H, Ayyaz S, Olsen AH. Complexities in the estimation of overdiagnosis in breast cancer screening. *British Journal of Cancer*. 2008; 99(7):1176–8. [PubMed: 18766185]
17. 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. *Journal of Urology*. 1996; 155(3):816–20. [PubMed: 8583583]
18. 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(13):981–90. [PubMed: 12096083]
19. Olsen AH, Agbaje OF, Myles JP, Lyng E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J*. 2006; 12(4):338–42. [PubMed: 16848843]
20. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009; 360(13):1320–8. [PubMed: 19297566]
21. 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. *Journal of Clinical Epidemiology*. 2011; 64(12):1412–7. [PubMed: 22032753]
22. 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. *Annals of Internal Medicine*. 2013; 158(3):145–53. [PubMed: 23381039]
23. 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(6):374–83. [PubMed: 19276453]
24. Wever EM, Draisma G, Heijnsdijk EAM, 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. *Journal of the National Cancer Institute*. 2010; 102(5):352–5. [PubMed: 20142584]
25. 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(16):1230–1. [PubMed: 9719085]
26. Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK. Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial. *J Urol*. 2005; 173(3):746–50. discussion 50–51. [PubMed: 15711261]
27. 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(12):1423–36. [PubMed: 1288272]
28. Jorgensen KJ, Zahl P-H, Gotzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC women's health*. 2009; 9:36. [PubMed: 20028513]
29. Davidov O, Zelen M. Overdiagnosis in early detection programs. *Biostatistics*. 2004; 5(4):603–13. [PubMed: 15475422]
30. Pinsky PF. Estimation and prediction for cancer screening models using deconvolution and smoothing. *Biometrics*. 2001; 57(2):389–95. [PubMed: 11414561]
31. Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol*. 1983; 118(6):865–86. [PubMed: 6650488]
32. 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(19):1325–9. [PubMed: 19720969]
33. Caro JJ, Briggs AH, Siebert U, Kuntz KM. On Behalf of the I-SMGRPTF. Modeling Good Research Practices--Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Medical Decision Making*. 2012; 32(5):667–77. [PubMed: 22990082]
34. Wu GH, Auvinen A, Maattanen 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. *International Journal of Cancer*. 2012; 131(6):1367–75.

35. 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. *Annals of Oncology*. 1998; 9(12):1297–300. [PubMed: 9932159]

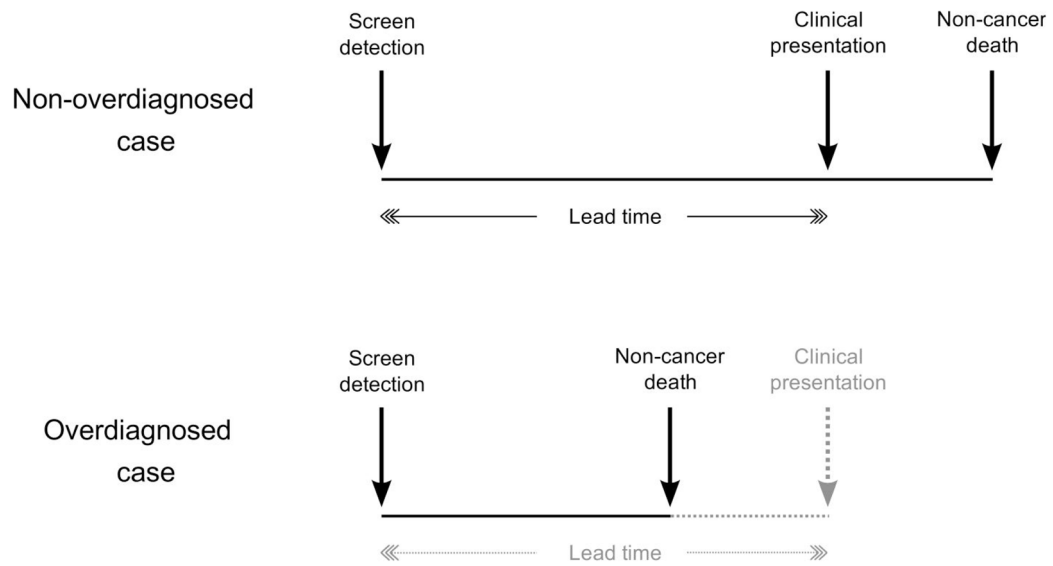


Figure 1.

Link between overdiagnosis and lead time. A non-overdiagnosed case (upper panel) is a screen-detected case who would have presented clinically before dying of non-cancer causes. An overdiagnosed case (lower panel) is a screen-detected case who would have died of non-cancer causes before presenting clinically, i.e., a screen-detected case who would not have been diagnosed without screening. The lead time is the time from screen detection to clinical diagnosis. The longer the lead time, the greater the chance of overdiagnosis. Similarly, the higher the risk of other-cause death, the greater the chance of overdiagnosis.

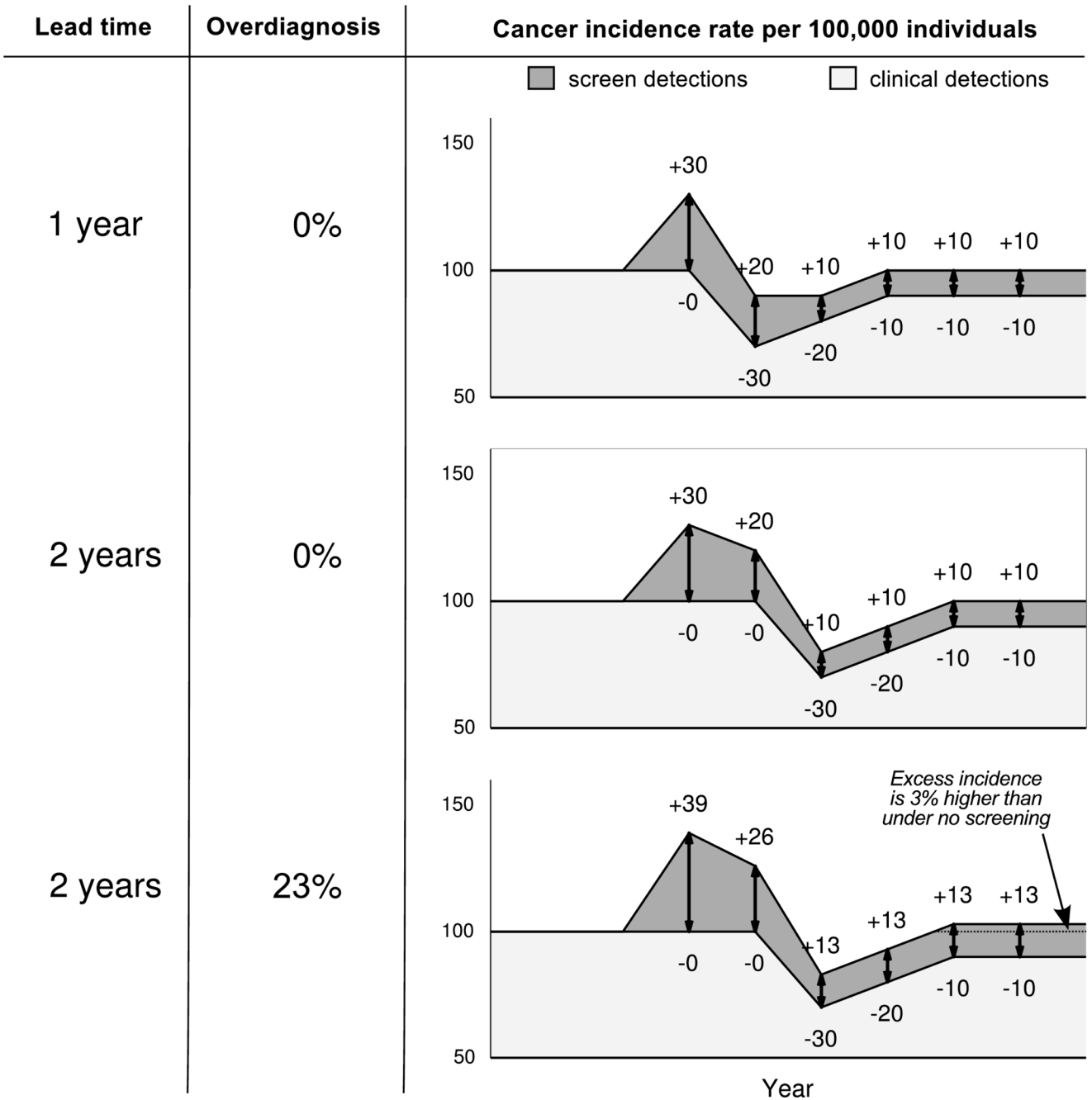


Figure 2. Incidence trends following the introduction of screening in three scenarios with the same baseline incidence (constant 100 cases per 100,000 individuals per year), disease prevalence, test sensitivity, and screening test dissemination. (1) Lead time equals 1 year, no overdiagnosis. (B) Lead time equals 2 years, no overdiagnosis (c) Lead time (among non-overdiagnosed cases) equals 2 years, 3 screen-detected cases overdiagnosed per 13 screen-detected cases (23%). In each panel, the number of cases per 100,000 screen-detected 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.

Table 1

A selection of studies of overdiagnosis due to prostate cancer screening, labeled according to the measure of overdiagnosis, the study design and context, and the estimation approach used. Studies were selected to reflect a range of choices of overdiagnosis measures, contexts and methods and to include recent, high-profile publications

Author	Years of study	Estimate	Measure	Approach	Context
Etzioni, 2002 (18)	1988–1999	29% in whites, 44% in blacks	Cases overdiagnosed/screen-detected cases	Lead-time	US SEER9 Age 60–84 Population study
Draisma, 2003 (13)	1994–2000	48%	Cases overdiagnosed/detected cases	Lead-time	ERSPC Rotterdam Age 55–67 Screened every 4 years PSA cutoff: 3.0 µg/L Clinical trial
Telesca, 2008 (6)	1973–2000	23% in whites, 34% in blacks	Cases overdiagnosed/screen-detected cases	Lead-time	US SEER9 Age 50–84 Population study
Draisma, 2009 (23)	1985–2000	23%, 28%, 42% 9%, 12%, 19%	Cases overdiagnosed/screen-detected cases Cases overdiagnosed/all detected cases	Lead-time	US SEER9 Age 50–84 Population study
Wu, 2012 (34)	1996–2005	3.4 %	Cases overdiagnosed/Screenees	Lead-time	ERSPC Finland Age 55–67 Screened every 4 years PSA cutoff 4.0 µg/L RCT
Welch, 2009 (32)	1986–2005	1.3 million	Excess cases	Excess incidence*	US SEER9 Age 20+ Population study
Schröder, 2009 (20)	1991–2006	48	Excess cases/lives saved	Excess incidence*	ERSPC Age 55–69 Screened every 2 or 4 years Clinical trial
Pashayan, 2006 (7)	1996–2002	40–64%	Cases overdiagnosed/cases detected by PSA test	Excess incidence*	Cambridge, UK PSA used as diagnostic test, not for routine screening Population study
Zappa, 1998 (35)	1992–1995	51%	Excess cases/cases in absence of screening	Excess incidence*	Florence, Italy Age 60 at entry 6 biennial screens Population study

PSA = prostate-specific antigen; SEER9 = core 9 catchment areas of the Surveillance, Epidemiology, and End Results program; ERSPC = European Randomized Study of Screening for Prostate Cancer

* Includes early years of screening use

A selection of studies of overdiagnosis due to breast cancer screening, labeled according to the measure of overdiagnosis, the study design and context, and the estimation approach used. Studies were selected to reflect a range of choices of overdiagnosis measures, contexts and methods and to include recent, high-profile publications

Table 2

Author	Years of study	In situ cases included ?	Estimate	Measure	Approach	Context
Morrell 2010 (10)	1999–2001	No	30–42%	Excess cases/cases expected without screening	Excess incidence	New South Wales, Australia Ages 50–69 Biennial screening Population study
Gotzsche, 2011 (3)	Multiple	Yes	30%	Excess cases/cases expected without screening	Excess incidence *	7 clinical trials of mammography
Kalager, 2012 (5)	1996–2005	No	15–25%	Excess cases/cases expected without screening	Excess incidence *	Norway Ages 50–69 Biennial screening Population study
Bleyer, 2012 (4)	1976–2008	Yes	31%	Excess cases/detected cases	Excess incidence *	US SEER9 Ages 40+ Population study
Independent UK Panel on Breast Cancer Screening (11)	Multiple	Yes	11% 3	Excess cases/cases expected without screening Excess cases/lives saved	Excess incidence	UK Ages 50–70 3 clinical trials of mammography Screened every 3 years
Paci, 2006 (9)	1986–2001	Yes No	4.6% 3.2%	Cases overdiagnosed/cases expected without screening	Lead-time	Five areas of Italy Ages 50–74 Biennial screening Population study
Olsen, 2006 (19)	1991–1995	No	4.8%	Cases overdiagnosed/detected cases	Lead-time	Copenhagen, Denmark Ages 50–69 Biennial screening (1st two rounds) Population study
De Gelder 2011 (8)	1990–2006	Yes	8.9% 4.6% 5%	Cases overdiagnosed/Screen-detected cases Cases overdiagnosed/detected cases Cases overdiagnosed/cases expected without screening	Lead time	The Netherlands Ages 40–69 Biennial screening Population study

SEER9 = core 9 catchment areas of the Surveillance, Epidemiology, and End Results program

* Includes early years of screening use

Table 3

Questions that readers of overdiagnosis studies should ask and writers of overdiagnosis studies should address to facilitate better understanding and comparison of results across studies

Study feature	Questions
Definition and measurement	<p>(1) How does this study define overdiagnosis?</p> <p>(2) How does this study measure overdiagnosis? What is the numerator and what is the denominator?</p>
Study design and context	<p>(3) Where was the study conducted?</p> <p>(4) In what years was the study conducted?</p> <p>(5) What were the characteristics of the study population that might affect overdiagnosis (e.g., age range, risk factors for disease, general health status)?</p> <p>(6) What was the screening protocol and what were the criteria for biopsy referral?</p> <p>(7) How compliant were participants with screening and biopsy protocols?</p> <p>(8) What was the incidence in the absence of screening?</p>
Estimation approaches	<p>(9) What estimation approach was used (excess incidence or lead time)?</p> <p>(10) If lead-time approach:</p> <ul style="list-style-type: none"> a. Are model assumptions clearly stated and justified? b. Does model-projected incidence under screening reasonable match observed incidence? c. What assumptions are made about concurrent disease incidence without screening if this is not observed? <p>(11) If excess-incidence approach:</p> <ul style="list-style-type: none"> a. When is study conducted relative to start of screening? Does it include the early years of screening? b. How is concurrent control incidence projected if this is not observed?