A Reality Check for Overdiagnosis Estimates Associated With Breast Cancer Screening

Ruth Etzioni, Jing Xia, Rebecca Hubbard, Noel S. Weiss, Roman Gulati

The frequency of overdiagnosis associated with breast cancer screening is a topic of controversy. Published estimates vary widely, but identifying which estimates are reliable is challenging. In this article we present an approach that provides a check on these estimates. Our approach leverages the close link between overdiagnosis and lead time by identifying the average lead time most consistent with a given overdiagnosis frequency. We consider a high-profile study that suggested that 31% of breast cancers diagnosed in the United States in 2008 were overdiagnosed and show that this corresponds to an average lead time of about nine years among localized cases. Comparing this estimate with the average lead time for invasive, screen-detected breast cancers of 40 months, around which there is a relative consensus, suggests the published estimate of overdiagnosis is excessive. This approach provides a novel way to appraise estimates of overdiagnosis given knowledge of disease natural history.

Overdiagnosis because of breast cancer screening is controversial, with estimates of its frequency varying greatly. [1,2] An overdiagnosed cancer is one detected by screening that would not have presented clinically during the patient's lifetime in the absence of screening, ie, the patient would have died from other causes with preclinical disease. Although there is no consensus regarding the most reliable estimates, an influential article[3] estimated that, in 2008, 31% of breast cancers among women older than age 40 years in the United States were overdiagnosed. This estimate was based on the excess incidence in 2008 relative to a projection of incidence in the absence of mammography.

Here we propose an approach for judging the plausibility of overdiagnosis estimates like this one, leveraging the link between overdiagnosis and lead time (LT), which is the time by which screening advances detection. For a patient, overdiagnosis occurs when the time to other-cause death is less than her LT. In a population, the fraction overdiagnosed is determined by the distributions of LT and other-cause survival. If we know the distributions of LT and other-cause survival, we can infer the chance of overdiagnosis. Similarly, if we know the chance of overdiagnosis and other-cause survival, we can infer the LT distribution.

Why is this useful? Because, for invasive breast cancers, the average LT is two to four years based on statistical models[4–9] fit to individual-level screening data. In accordance with a recent[10] summary estimate, we use a mean of 40 months as a consensus value.

Checking whether an overdiagnosis estimate is consistent with a consensus LT provides a check on the plausibility of the estimate. Here, we apply this approach to the estimate of 31% overdiagnosed in 2008. Among breast cancers in the Surveillance, Epidemiology, and End Results (SEER) registry in 2008, 22% were (SEER historic stage)[11] in situ, 49% were localized, and 29% were advanced (regional or distant).[12] It is likely that few advanced cancers are overdiagnosed; we assume that none were overdiagnosed in 2008. We also assume initially that all in situ cases were overdiagnosed. The remaining overdiagnosed cases must be localized, amounting to approximately 18% ([31−22]/49) of localized cases. We aim to identify the average LT yielding an overdiagnosis frequency of 18% for these cases. Assuming that all in-situ cases are overdiagnosed is conservative, minimizing the overdiagnosis frequency among localized cases and lowering the corresponding LT.

We estimate other-cause survival using SEER*Stat[13] given the age distribution for localized cases diagnosed in 2008. Because many localized cases have been screened, they have a lower risk of noncancer death than the general population.[14] For each year post diagnosis, we compute the ratio of the observed risk of other-cause death (O) to the expected risk of death in the age-matched population (E) ( ) and use this to derive a hazard ratio (HR) to

adjust US life tables for this case population. The estimated HR is 0.75 for localized cancers; ie, the annual risk of death is 25% lower among these cases than the age-matched female population.

Table 1. Calculation of hazard ratio for risk of other-cause death among localized, invasive breast cancer cases relative to age-matched population*

Year interval	Observed interval survival (OBS)	Expected interval survival (EXP)	Cause-specific interval survival (CS)	Other-cause interval probability (OC)	Hazard ratio†
< 1 yr	98.5%	98.1%	99.5%	1.0%	0.52
1 to <2 yr	97.9%	98.0%	99.2%	1.3%	0.64
2 to <3 yr	97.4%	97.9%	99.0%	1.6%	0.76
3 to <4 yr	97.4%	97.8%	99.0%	1.6%	0.72
4 to <5 yr	97.1%	97.6%	99.0%	1.9%	0.79
5 to <6 yr	97.1%	97.5%	99.1%	2.0%	0.80
6 to <7 yr	96.8%	97.4%	99.1%	2.3%	0.88
7 to <8 yr	96.5%	97.2%	99.2%	2.7%	0.96

*The table shows Surveillance, Epidemiology, and End Results 18 registries observed, expected, and cause-specific survival by year following diagnosis for localized, invasive breast cancer cases aged 40 and above diagnosed between 2003 and 2010. For each interval of follow-up we calculate the probability of other-cause (noncancer) death within the interval as OC=CS-OBS.

†The hazard ratio (HR) is given by log(1-OC)/log(EXP). The average HR over the first eight years is 0.75.

Competition between times to clinical diagnosis and other-cause death is implemented via simulation. We generate a virtual population of women with ages as in SEER localized cases diagnosed in 2008. For each woman we simulate two times: time to other-cause death (D) and LT with a specified mean (M). Then we compute the percent overdiagnosed as the empirical fraction of women with D < LT. We vary M to find the value that yields 18% overdiagnosis. We first assume that the LT follows an exponential distribution and also allow distributions with more and less extreme lead times to represent differing frequencies of indolent cancers.

provides mean LT for a range of overdiagnosis frequencies. Under an exponential LT distribution, for 18% of localized cancers to be overdiagnosed, the mean LT must be approximately 108 months. This increases to 136 months if only 90% of in situ cases are overdiagnosed, implying that 23% ([31-19.8]/0.49) of localized cases must be overdiagnosed. The mean LT for localized cases that yields 18% overdiagnosed is longer under Weibull (shape = 0.5) and slightly shorter under Weibull (shape = 2.0) than under the exponential.

Table 2. Frequency of overdiagnosis among localized invasive screen-detected cancers corresponding to given mean lead times under two different settings for the risk of other-cause death and under three different assumptions for the lead-time distribution

Mean lead time among localized invasive breast cancers, y	Frequency of overdiagnosis among localized invasive breast cancers (percent)		
	LThist*: Exponential	LThist*: Weibull (shape = 0.5)	LThist*: Weibull (shape = 2.0)
		HR† = 0.75	HR† = 0.75

	HR† = 0.75	HR† = 0.85		
2	3.3	4.0	3.3	3.5
4	7.0	8.2	6.9	7.3
6	11.3	13.5	10.6	11.9
8	15.5	17.3	14.1	17.0
10	20.4	21.3	16.2	20.2
12	24.3	24.9	19.5	24.4

*A Weibull distribution with shape 0.5 is more dispersed than an exponential distribution with the same mean. A Weibull distribution with shape 2.0 is less dispersed than an exponential distribution with the same mean. LTdist = lead-time distribution.

†HR = hazard ratio for other-cause death among localized invasive cases relative to the age-matched female population.

These lead-time estimates apply to all cases; the corresponding lead times among screen-detected cases will be higher, because cases that are not screen detected have a lead time of zero.

Under all settings, the average LT most consistent with 31% of cases overdiagnosed markedly exceeded 40 months. However, since the 40-month estimate applies to all invasive cancers detected by screening, including advanced cancers, we need an estimate that pertains only to localized cancers. Noting that approximately 25% of invasive screen-detected cancers are advanced[15] and assuming that the LT among advanced cancers is short (about six months) implies that we should be comparing our results against an estimate of 51 rather than 40 months (since 40 = 0.75×51 + 0.25×6). However, even this value is much lower than the mean lead times from .

Table 2. Frequency of overdiagnosis among localized invasive screen-detected cancers corresponding to given mean lead times under two different settings for the risk of other-cause death and under three different assumptions for the lead-time distribution

Mean lead time among localized invasive breast cancers, y	Frequency of overdiagnosis among localized invasive breast cancers (percent)			
	LThist*: Exponential		LThist*: Weibull (shape = 0.5)	LThist*: Weibull (shape = 2.0)
	HR† = 0.75	HR† = 0.85	HR† = 0.75	HR† = 0.75
2	3.3	4.0	3.3	3.5
4	7.0	8.2	6.9	7.3
6	11.3	13.5	10.6	11.9
8	15.5	17.3	14.1	17.0
10	20.4	21.3	16.2	20.2
12	24.3	24.9	19.5	24.4

*A Weibull distribution with shape 0.5 is more dispersed than an exponential distribution with the same mean. A Weibull distribution with shape 2.0 is less dispersed than an exponential distribution with the same mean. LTdist =

lead-time distribution.

†HR = hazard ratio for other-cause death among localized invasive cases relative to the age-matched female population.

A limitation of our study is that we use standard distributions for the LT. It is possible that some screen-detected cancers would never progress. These cancers will effectively have infinite lead times. The lead-time studies cited[4-9] generally do not explicitly separate these cases from those with a defined, finite LT; rather, these studies specify a single distribution that accommodates longer as well as shorter lead times. Our LT distributions follow these precedents so that our estimated lead times can be compared with the literature.

We conclude that an overdiagnosis rate of 31% among all breast cancer cases in 2008 seems excessive. This may be because of the use of excess incidence, which often yields an overestimate.[1,2] The same reasoning can be applied to examine the plausibility of the estimate of 22% overdiagnosed among invasive cancers in the Canadian breast cancer screening trial.[16] It is commonly believed that excess incidence estimates from clinical trials are a gold standard for estimating overdiagnosis. However, as in population studies, excess incidence estimates from trials can also produce biased results.[17]

References

1. Etzioni R, Gulati R, Mallinger L, et al. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Ann Int Med.* 2013;158(11):831–838.

2. Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen.* 2012;19 Suppl 1:42–56.

3. Bleyer A, Welch HG. Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence. *N Engl J Med.* 2012;367(21):1998–2005.

4. Duffy SW, Chen HH, Tabar L, et al. Estimation of Mean Sojourn Time in Breast-Cancer Screening Using a Markov-Chain Model of Both Entry to and Exit from the Preclinical Detectable Phase. *Stat Med.* 1995;14 (14):1531–1543.

5. Duffy SW, Chen HH, Tabar L, et al. Sojourn time, sensitivity and positive predictive value of mammography screening for breast cancer in women aged 40–49. *Int J Epidemiol.* 1996;25(6):1139–1145.

6. Shen Y, Zelen M. Screening Sensitivity and Sojourn Time From Breast Cancer Early Detection Clinical Trials: Mammograms and Physical Examinations. *J Clin Oncol.* 2001;19(15):3490–3499.

7. Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol.* 1983;118(6):865–886.

8. Chen JS, Prorok PC. Lead time estimation in a controlled screening program. *Am J Epidemiol.* 1983;118 (5):740–751.

9. Etzioni R, Shen Y. Estimating asymptomatic duration in cancer: The AIDS connection. *Stat Med.* 1997;16 (6):627–644.

10. Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. *Breast Cancer Res.* 2013;15(3):R41.

11. Seiffert JE, Shambaugh EM, Kruse M, et al. *SEER Program: Comparative Staging Guide for Cancer.* Available at: http://seer.cancer.gov/archive/manuals/historic/comp_stage1.1.pdf.

12. *Surveillance, Epidemiology, and End Results (SEER) Program* (www.seer.cancer.gov) *SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2012 Sub (1973–2010), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch.* Released April 2013, based on the November 2011 submission.

13. Weiss NS, Rossing MA. Healthy screenee bias in epidemiologic studies of cancer incidence. *Epidemiol.* 1996;7(3):319–322.

14. SEER*Stat Software. *Surveillance Research Program, National Cancer Institute SEER*Stat software Version 8.1.5.* Available at: http://seer.cancer.gov/seerstat/.

15. White E, Miglioretti DL, Yankaskas BC, et al. Biennial versus annual mammography and the risk of late-stage breast cancer. *J Natl Cancer Inst.* 2004;96(24):1832–1839.

16. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomized screening trial. *BMJ.* 2014;348:g366

17. Etzioni R, Gulati R. RE: A model too far. *J Natl Cancer Inst.* 2014;106(4):dju058.

Note

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