The Breast 22 (2013) S73–S76

Contents lists available at SciVerse ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Breast cancer screening: Controversy of impact

Donald A. Berry*,[1]

Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, 1400 Pressler Street, 4-5062 Pickens Academic Tower, Houston, TX 77030-1402, USA

Keywords:
Breast cancer detection
Screening mammography
Biases in screening studies
Randomized screening trials
Modeling population breast cancer mortality

A B S T R A C T

Few medical issues have been as controversial—or as political, at least in the United States—as the role of mammographic screening for breast cancer. The advantages of finding a cancer early seem obvious. Indeed, randomized trials evaluating screening mammography demonstrate a reduction in breast cancer mortality, but the benefits are less than one would hope. Moreover, the randomized trials are themselves subject to criticism, including that they are irrelevant in the modern era because most were conducted before chemotherapy and hormonal therapy became widely used.

In this article I chronicle the evidence and controversies regarding mammographic screening, including attempts to assess the relative contributions of screening and therapy in the substantial decreases in breast cancer mortality that have been observed in many countries over the last 20–25 years. I emphasize the trade-off between harms and benefits depending on the woman's age and other risk factors. I also discuss ways for communicating the associated risks to women who have to decide whether screening (and what screening strategy) is right for them.

Introduction

Few medical questions have been as controversial—or as political, at least in the United States—as the role of mammographic screening for breast cancer. Even breast cancer advocacy groups are at loggerheads. Some arguments on both sides of the question have been more vitriolic than enlightened. An extreme example is a Harvard professor's argumentum ad hominem criticism of the U.S. Preventive Services Task Force as cited in the *Washington Post*: "Tens of thousands of lives are being saved by mammography screening, and these idiots want to do away with it." "It's crazy—unethical, really" [1]. A recent monograph from the opposite perspective is titled *Mammography Screening: Truth, Lies and Controversy* [2].

To many people the benefit of finding a cancer early seems so obvious as to defy the need for the empirical confirmation. Some researchers attempt empirical confirmation by conducting observational studies comparing survival of patients whose cancers were detected by screening versus not. Such analyses are easy to make, whether within or across programs or databases. And they show substantial benefits for screening. But they are fatally flawed. Hundreds if not thousands of such observational studies have been published. These studies are worthless, individually and in total. But believers tout them as evidence.

The principal fatal flaws in these studies are lead-time and length biases [3]. *Lead-time bias* should be easy to recognize and to understand. Suppose a screening mammogram detects a tumor in a 50-year-old woman. Had she not been screened the tumor might have become symptomatic at age 55, say. The lead-time due to screening is 5 years. She would have lived for 5 years longer with her cancer had she been screened, even with no screening benefit.

Length bias is more subtle than lead-time bias, but its impact is even greater. Breast tumors are heterogeneous. Some grow fast and others are indolent. The ones that grow fast have a short *sojourn time*, the period between when it is detectable by mammography but is not yet symptomatic. On the other hand, indolent tumors by definition have long sojourn times. Mammography preferentially detects indolent tumors simply because they are detectable for a longer period. Length bias results because indolent tumors are less likely to recur and they are less likely to be lethal.

Overdiagnosis is an extreme form of length bias, a subject I return to below. Some tumors grow so slowly that they would never be found during the woman's lifetime were it not for screening.

The resolution of lead-time and length biases in comparing cancer survival is to identify and start following individuals before

* Express Mail: 1400 Pressler Street, Box 301402, USA. Tel.: +1 713 794 4141, +1 713 817 5586 (Mobile); fax: +1 713 563 4243, +1 713 563 4242.
E-mail address: dberry@mdanderson.org.
[1] Assistant: Lydia Davis. Tel.: +1 713 794 4142. lbdavis@mdanderson.org.

0960-9776/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.breast.2013.07.013

they have cancer rather than considering individuals after their cancers have been detected. Ideally, patient assignment to be screened or not should be randomized.

Role of women's age in the randomized trials

There have been 10 randomized breast screening trials (although the count depends on which trials are regarded as distinct) [4]. The earliest trial was the Health Insurance Plan (HIP) of New York, initiated in 196x [5]. None of the trials is immune to criticism [6], including that they are not relevant for the modern era of chemotherapy and hormonal therapy. The HIP trial and the Edinburgh trial [7] are sufficiently flawed that I shall not consider them further.

A principal focus of the analyses of the randomized trials has been the age to start screening. The 2009 U.S. Preventive Services Task Force (USPSTF) publication concluded that the "'number needed to invite for screening to extend one woman's life' [is] 1904 for women aged 40–49 years and 1339 for women aged 50–59 years." [4] There is no abrupt change at age 50 in either incidence or lethality of the disease. So these numbers are not constant over their respective intervals. In Fig. 1 I have interpolated within age intervals and extrapolated outside them assuming only breast cancer incidence matters and based on incidence statistics from the U.S. (This figure does not show estimation uncertainty, which is substantial; see below.)

One USPSTF conclusion was that "Screening mammography should not be done routinely for all women age 40–49 years." This conclusion was highly controversial in the U.S. Curiously, the basis of the controversy was never the evidence used but the fact that the task force decided on the cutpoint at age 50. The USPSTF conclusion was essentially the same as that of an NIH Consensus Development Conference [8], largely because the randomized evidence had changed little over the intervening years.

No organization of any repute recommends screening women younger than age 40 at normal risk of breast cancer (although the American Cancer Society once promoted having a "baseline mammogram" at age 35). The USPSTF chose the cutpoint age of 50 by weighing harms and benefits. They conclude that "Women and their doctors should base the decision to start mammography before age 50 years on the risk for breast cancer and preferences about the benefits and harms" [4].

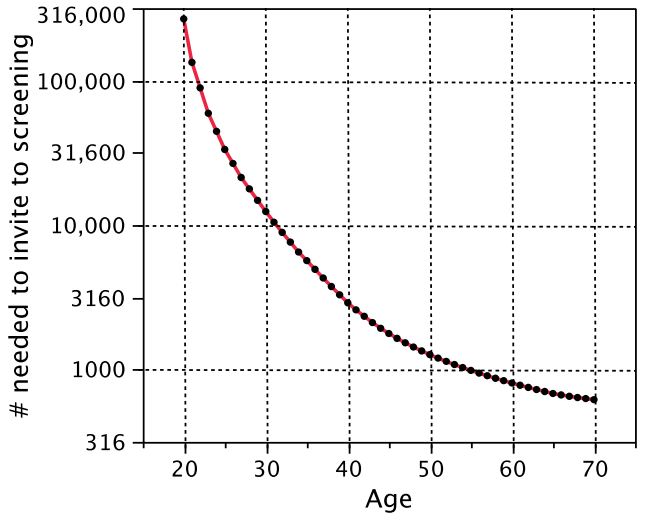

Fig. 1. Number needed to invite to screening to save one life as reported by the USPSTF, but with my interpolations and extrapolations based on U.S. breast cancer incidence statistics. In particular, the USPSTF estimated the number needed to invite for screening to avoid on death over a 20-year period is 1904, 1339, and 377 for women ages 39–49, ages 50–59, and 60–74, respectively. The smooth curve in the plot has these averages over the intervals in question.

The largest and most important set of randomized trials is from Sweden. These trials were updated most recently in 2002 [9]. The Swedish trials utilized a mammogram in the control group. The control mammogram was to occur at the time of the final mammogram in the group invited to screening. The "evaluation model" compares breast cancer mortality based on those cancers detected up until the time of and including at the last mammogram. Unfortunately, the timing of the "last mammogram" slipped, by as much as 14 months [10]. Therefore, Nystrom also provides a "follow-up model," which is not subject to this "evaluation bias." The follow-up model compares all breast cancer deaths in both groups, with follow-up truncated in all trials and groups at the same calendar date. The results of this follow-up model are shown in Table 1.

Table 1 also shows the results of the Canadian NBSS trials [11,12], and the U.K. trial [13].

The rightmost column in Table 1 is the difference between the number of deaths per 1000 women years in control minus the corresponding number in the screened group. The number needed to screen to avoid one death per 1000 woman years is the inverse of the table entry, discussed further below.

Fig. 2 shows the results from Table 1 in graphical form. Fig. 2A shows reduction in breast cancer mortality as it depends on age at entry into a screening program. There is a suggestion of an age effect, especially in the Swedish trials.

Fig. 2B shows the number of deaths avoided per 1000 women years. There are several important aspects of this plot and the fitted spline. One is the similarity of the results across countries within age groups. In particular, despite arguments regarding the anomalous nature of the Canadian trials [14], the results in the Canadian and Swedish trials (when using the follow-up model) are quite comparable. Another important aspect of Fig. 2B is the continuity of the results as a function of age. Still another is the small estimated benefit for women in their 40s and even into their 50s.

In calculating the number needed to invite to screening to avoid one death, the USPSTF referred to a 20-year period. Assuming a risk reduction for screening that is constant over time and persisting over a 20-year period, the analogous number is the inverse of the number indicated in Fig. 2B times 50 (the number of 20-year periods in 1000 women years). Using the spline estimates at ages 45, 55, and 65, these are approximately 1900, 800, and 300, respectively. For comparison, the USPSTF calculations (cf. Fig. 1) for women younger than 50 was 1904 (95% CI, 929–6378), for women

Table 1
Breast cancer mortality of the randomized trials of mammographic screening, as described in the text, with the Swedish trials combined to show results by age group.

Country	Age at entry	Women years (1000s)		Number of deaths		RR	RR 95% CI	Deaths avoided/ 1000 yrs
		IG	CG	IG	CG			
Sweden [8]	40–44	320	281	85	88	0.85	0.64–1.13	0.048
	45–49	377	338	135	128	0.95	0.75–1.21	0.021
	50–54	341	320	144	146	0.92	0.73–1.16	0.034
	55–59	368	357	177	201	0.86	0.70–1.05	0.082
	60–64	260	201	128	129	0.79	0.62–1.01	0.149
	65–69	137	131	84	119	0.68	0.52–0.89	0.295
	70–74	62	59	42	36	1.12	0.73–1.72	−0.067
Canada [10]	40–49	328[a]	328[a]	105	108	0.97	0.74–1.27	0.009
Canada [11]	50–59	216	216	107	105	1.02	0.78–1.33	−0.009
U.K. [12]	40	578	1149	105	251	0.83	0.66–1.04	0.037

RR: relative risk, IG: invited (to screening) group, CG: control group.
[a] Estimated.