

Mammograms and Mortality: How Has the Evidence Evolved?

Amanda E. Kowalski

Abstract

Decades of evidence reveal a complicated relationship between mammograms and mortality. Mammograms may detect deadly cancers early, but they may also lead to the diagnosis and potentially fatal treatment of cancers that would never progress to cause symptoms. I provide a brief history of the evidence on mammograms and mortality, focusing on evidence from clinical trials, and I discuss how this evidence informs mammography guidelines. I then explore the evolution of all-cause mortality relative to breast cancer mortality within an influential clinical trial. I conclude with some responses to the evolving evidence.

Advocates of preventive care emphasize that it saves money and lives. However, preventive care need not save money (as Newhouse discusses in this issue). Preventive care also need not save lives. In this paper, I analyze evolving evidence on the mortality impact of a common form of preventive care: the mammogram, an X-ray picture of the breast.

The rationale for widespread mammography is that early detection of potentially fatal breast cancers enables earlier and more effective treatment. But there is a potential drawback: mammography can detect some early-stage cancers that will never progress to cause symptoms — a phenomenon often referred to as overdiagnosis. In such cases, the emotional, financial, and physical costs of a cancer diagnosis and any subsequent treatments occur without any corresponding health benefit. Because it is hard to tell which women will be harmed by their cancers, there is a tendency to treat all women as if their cancers will be lethal (Mukherjee, 2017). Even if the initial cancer would have never proven life-threatening, exposure to chemotherapy, radiotherapy, and surgery can potentially lead to new conditions,

Amanda E. Kowalski is Gail Wilensky Professor of Applied Economics and Public Policy, University of Michigan, Ann Arbor, Michigan. Her email address is aekowals@umich.edu.

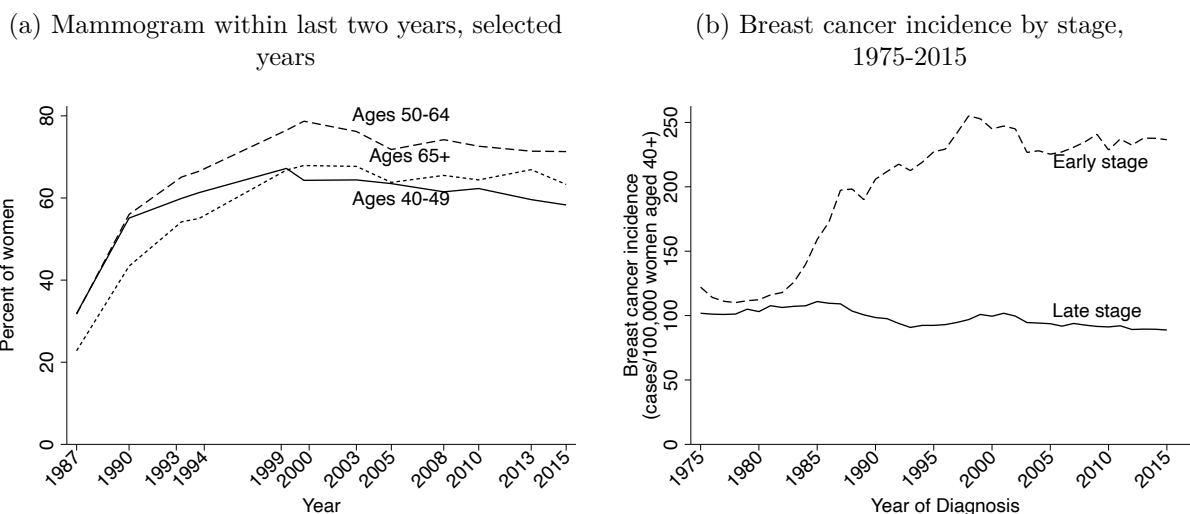
even to new fatal cancers (Praga et al., 2005; Early Breast Cancer Trialists’ Collaborative Group, 2005).

The possibility of overdiagnosis turns out to be central to guidelines for mammography screening. Prior to 2009, the US Preventive Services Task Force (2002) recommended regular mammography screening for asymptomatic women aged 40 and older. In 2009, the task force revised its guidelines in light of the most recent follow-up data available from clinical trials on mammography (US Preventive Services Task Force, 2009).

The task force’s latest guidelines, which are based on a 2016 meta-analysis, reaffirm the 2009 revision. They recommend regular mammography for women aged 50 to 74, but they leave the mammography decision up to individual women in their 40s: “Women who place a higher value on the potential benefit than the potential harms may choose to begin biennial screening between the ages of 40 and 49 years” (Siu, 2016). The task force does not provide guidelines for women older than 74 due to insufficient evidence.

In all age groups, mammography has increased dramatically over time in the United States, as shown in Figure 1a. By 2015, 58.3 percent of women aged 40–49, 71.3 percent of women aged 50–65, and 63.3 percent of women aged 65+ reported receiving a mammogram within the past two years. Mammography for women in their 40s has fallen only slightly and gradually since guidelines changed in 2009. Widespread mammography seems embedded in the US health care system, both as a matter of the acculturation of patients and health care providers, and also as a matter of financial incentives: the aggregate annual cost of mammography has been estimated to be \$2.1 billion just among US women in their 40s with private health insurance (Kunst et al., 2020).

Figure 1: Trends in Mammography and Breast Cancer Incidence in the United States



Source. Panel (a): National Center for Health Statistics (2011, 2019); Panel (b): Author’s calculations from Surveillance, Epidemiology, and End Results Program (SEER) (2020).

Note. Incidence is age-adjusted to the 2000 US population. Following Bleyer and Welch (2012), cancer stage is given by the SEER historic stage A variable. Early-stage cancers are in situ or localized. Late-stage cancers are regional or distant.

The US Preventive Services Task Force identifies overdiagnosis as the most important

harm that mammograms pose. Though false positives can also pose harm, overdiagnosis is a separate phenomenon. According to the task force definition, overdiagnosis refers to “the diagnosis and treatment of noninvasive and invasive breast cancer that would otherwise not have become a threat to their health, or even apparent, during their lifetime” (Siu, 2016).

Overdiagnosis is difficult to identify. If a mammogram catches an early-stage cancer, it is impossible to discern how that cancer would have progressed absent detection, especially for a given individual. However, several types of evidence suggest overdiagnosis of cancer. For example, autopsy studies showing that almost half of older men die with, but not necessarily of, prostate cancer have been important to prostate cancer screening guidelines since the late 1980s (US Preventive Services Task Force, 1989). As summarized by Welch and Black (2010), autopsy studies also suggest overdiagnosis of thyroid and breast cancer, and time series evidence suggests overdiagnosis of thyroid, melanoma, and kidney cancer: as diagnoses have risen dramatically with screening, deaths from these types of cancers have been stable.

Deaths from breast cancer have decreased as mammography has increased, but differences in the decreases across geographic areas suggest overdiagnosis. Autier et al. (2010) find that breast cancer mortality for women under age 50 fell from 1989 to 2006 in all 30 European countries that they examine, but they note that the greatest reductions occurred in countries where screening was less common. Using their results, Gøtzsche and Jørgensen (2013) emphasize that the reduction in breast cancer mortality was 49 percent in Denmark but by only 36 percent in Sweden, despite the availability of screening mammography to half of women in this age group in Sweden versus none in Denmark. In the United States, breast cancer mortality for women under age 50 fell by an intermediate 39 percent over the same period (author’s calculations from Surveillance, Epidemiology, and End Results Program (2019)). Difference-in-difference evidence within Denmark for older women aged 55 to 74 shows steeper declines in breast cancer mortality in regions without screening programs from 1997 to 2006 (Jørgensen et al., 2010). These findings suggest that decreases in breast cancer mortality are due to factors other than screening, such as decreased use of menopausal hormone therapy (Ravdin et al., 2007) and improved treatments.

Bleyer and Welch (2012) provide an approach that quantifies overdiagnosis from screening. The goal of screening is to prevent early-stage cancers from progressing to a late stage. Therefore, an increase in early-stage cancers without a decrease in late-stage cancers of the same magnitude indicates overdiagnosis, under the assumption that late-stage cancers will be detected regardless of screening. Figure 1b shows that the incidence of early-stage breast cancers in the United States has skyrocketed as the incidence of late-stage breast cancers has declined only slightly. Bleyer and Welch (2012) estimate that in 2008 alone, breast cancer was overdiagnosed for 70,000 women, accounting for 31 percent of all breast cancer diagnoses. Conducting the same exercise using the data in Figure 1b, which replicates their main figure to 2008 and extends it to 2015, I find a similar rate of overdiagnosis in the subsequent

seven years.

Evidence of overdiagnosis from randomized controlled trials is the most compelling. Screening detects cancer, so trials should find higher rates of cancer in the intervention arm in the short term. In the long term, if the only impact of screening is to detect cancer early, rates of cancer should converge in the intervention and control arms as women who experience symptoms have their cancers diagnosed. A failure to converge after a reasonable amount of lead time indicates overdiagnosis. Overdiagnosis became apparent in a lung cancer screening trial based on 16 years of follow-up data (Marcus et al., 2006). Since then, overdiagnosis has become apparent in two studies of mammography trials based on at least 15 years of follow-up data, which imply overdiagnosis rates of 5 to 55 percent depending on the subgroup and base rate (Zackrisson et al., 2006; Baines et al., 2016).

Overdiagnosis, though undesirable, might be tolerable if screening reduces mortality. The goal of this paper is to analyze how evidence on mortality has evolved over time, across randomized controlled trials of mammography and within one trial. Preventive care such as mammography aims to preempt harm over time, so it is important to consider evolving evidence. I begin with a brief history of the prominent trials, and I present the latest mortality results. These results are the basis for the US Preventive Services Task Force guidelines on mammography screening as well as the basis for guidelines from other public health authorities in the United States and other countries. In a new empirical exercise, I explore the evolution of all-cause mortality relative to breast cancer mortality within a prominent trial, the Canadian National Breast Screening Study, using 20 years of follow-up data. In the conclusion, I offer some responses to the evolving evidence.

A Brief History of Mammography Trials

The 2016 guidelines for mammography screening from the US Preventive Services Task Force are based on a meta-analysis that focuses on mortality impacts obtained from eight large randomized controlled trials (Nelson et al., 2016). All trials combined include over 600,000 women, and individual trials range in size from approximately 39,000 to 160,000. Overall, the meta-analysis finds no statistically significant reduction in all-cause mortality for women in any age group. Some trials even show imprecise increases in all-cause mortality across all age groups or within an age group. Results focused only on breast cancer mortality are slightly more promising. The meta-analysis finds statistically significant but small reductions in breast cancer mortality for women in their 50s and 60s, and it finds imprecise reductions for women aged 39 to 49 and women aged 70 to 74. In this section, I provide a brief overview of the trials and some related evidence on mammography.

The first randomized controlled trial of mammography screening was established in the United States in 1963. Interest in mammograms was growing at the time because breast cancer was the leading cause of cancer death among US women (National Cancer Institute, 1979). Before mammograms were used to screen asymptomatic women, they were used to

diagnose women with symptoms detected through clinical or self-examination. Large declines in cervical cancer mortality were attributed to pap test screening (Boyes et al., 1973), so there was hope that mammography screening would reduce breast cancer mortality.

Women aged 40–64 who were enrolled in the Health Insurance Plan of Greater New York entered the trial between December 1963 and June 1966. The plan provided medical care on a prepaid basis, with no additional fees for follow-up care. The plan randomly assigned 31,000 women to an intervention arm. Those women were invited for a mammogram and a clinical breast examination, followed by three years of follow-up examinations, which continued even if the women disenrolled from the health insurance plan.

The Health Insurance Plan trial compared the intervention arm to a control arm of 31,000 women who continued to receive care as usual. Preliminary results were published in 1971 based on data through 1969 (Shapiro et al., 1971), which followed almost all women for at least four years and some women for up to six years, depending on their enrollment dates. Of the women invited for screening, about two-thirds appeared for their initial examination, and about two-thirds of those women continued through their third annual examination, resulting in an adherence rate of approximately 45 percent. Of the screened women, biopsies were recommended for 873 women, 545 of which were recommended only on the basis of the clinical exam and not mammography. Of that group, 624 women underwent biopsies, which confirmed breast cancer for 127 of them; for the remaining 497 women, about 80 percent of those who underwent biopsies, the breast cancer diagnosis was a false positive.

Overall, there were 31 breast cancer deaths in the intervention arm and 52 in the control arm. Thus, the “relative risk” of breast cancer mortality in intervention relative to control was about 0.6 ($\approx 31/52$). In absolute terms, there was a less striking reduction of about 7 breast cancer deaths per 10,000 women ($\approx (52-31)/31,000$). Almost all this reduction was for women who died at age 50 and older. Among women who died in their 40s, there were 11 breast cancer deaths in intervention and 12 in control.

The researchers had followed up intensively on all women with confirmed breast cancer cases to identify breast cancer deaths, but they had only completed their process of gathering information on deaths from all causes among 14 percent of women without confirmed breast cancer cases. In the preliminary data, there were 109 deaths from all causes in intervention but only 99 in control (Shapiro et al., 1971). That is, overall deaths were higher among women invited for screening.

The preliminary results on breast cancer deaths from the Health Insurance Plan trial galvanized support for a massive breast screening program in the United States, the Breast Cancer Detection and Demonstration Project. The project did not involve randomization, although it did facilitate some data collection. Between 1973 and 1980, the project recruited 280,000 women to be screened annually for five years via mammograms and clinical breast exams (Cunningham, 1997). The project recruited women as young as 35, and approximately

half of participants were under age 50, despite the lack of evidence in favor of mammography for younger women. Even before any long-term data became available, the American Cancer Society started recommending regular mammograms for women aged 50 and older in 1976; it recommended regular mammograms for women aged 40 and older in 1983 ([American Cancer Society, 2017](#)).

At this point, interest in breast cancer screening programs began to increase around the world, but there was some skepticism given the limited evidence from the Health Insurance Plan Trial. More randomized trials were needed, but given the large numbers of US women who were already receiving mammograms, a follow-up randomized trial would have been practically quite difficult in the United States. A working group convened to review the Breast Cancer Detection Demonstration Project ([Summary Report of the Working Group to Review the National Cancer Institute-American Cancer Society Breast Cancer Detection Demonstration Projects, 1979](#)) and recommended randomized controlled trials to evaluate questions left unanswered by the Health Insurance Plan trial: for example, should the guidelines recommend mammograms for women in their 40s? And what was the relative contribution of mammography compared to breast examination for women in their 50s?

In 1980, researchers initiated the Canadian National Breast Screening Study to evaluate these and other questions ([Miller et al., 1981](#)). Rather than inviting women for screening from a health insurance plan or a population register, this trial conducted a multi-pronged publicity campaign to recruit individual women. Initial media enthusiasm supported recruitment, but over time, high-profile critics fueled public concerns over radiation exposure ([Baines, 1984](#)). Though overdiagnosis was not central to public concerns, the trial protocol acknowledged concerns related to overdiagnosis as the other main potential hazard from mammography ([Miller et al., 1981](#)). The protocol also proposed an approach that has been used to quantify these concerns using data on breast cancer diagnoses from the trial ([Miller et al., 2014](#); [Baines et al., 2016](#)).

The Canadian National Breast Screening Study (CNBSS) enrolled roughly 90,000 women between 1980 and 1985: CNBSS-1 enrolled 50,430 women aged 40 to 49, while CNBSS-2 enrolled 39,405 women aged 50 to 59. These enrollments were in line with power calculations established in the trial protocol ([Miller et al., 1981](#)), informed by the latest breast cancer mortality results from the Health Insurance Plan trial ([Shapiro, 1977](#)). All participants in the CNBSS received a clinical breast exam at enrollment before randomization. Women assigned to the intervention arm received a mammogram during each year of the active study period, which included the enrollment year and the three to four years following enrollment (depending on where and when they enrolled). Women assigned to the control arm received usual care in the community in CNBSS-1 and access to a clinical breast examination during each year of the active study period in CNBSS-2. The trial had the highest adherence rate of all major mammography trials—85 percent ([Nelson et al., 2016](#)). For background

information on the CNBSS, good starting points include a recent book by Pellerin (2019) and academic publications by the investigators, especially those that report the latest results on mortality and breast cancer incidence 25 years after the first participants enrolled (Miller et al., 2014; Baines et al., 2016).

Several other international trials began in the 1970s and 1980s. Table 1 summarizes the main trials. It takes information directly from a table within the meta-analysis that informs the current US Preventive Services Task Force guidelines (Nelson et al., 2016). Although Table 1 contains seven rows, Nelson et al. (2016) count CNBSS-1 and CNBSS-2 separately in their discussion of eight trials. Four trials included in Table 1 began in Sweden, in part because Sweden had population registers that could be used to define a trial population by age within a geographic area. Furthermore, it was straightforward to track outcomes because Sweden had a cause of death registry that began in 1951 and a cancer registry that began in 1958. In 1976, the Malmö Mammographic Screening Trial, later known as MMST I, began in Malmö, Sweden. As shown in Table 1, it included approximately 42,000 women. In the following year, 1977, a trial that was more than twice as large, known as the Swedish Two-County Trial, began in Kopparberg and Östergötland. In 1977, the trial in Malmö closed and began enrolling women in an extension trial with the same protocol, known as MMST II. Other trials began in Stockholm in 1981 and Gothenburg in 1982. The Swedish trials have subsequently been pooled for analysis to increase statistical power (Nyström et al., 1993).

Another trial began in Edinburgh, Scotland in 1978 (Alexander et al., 1999). It recruited approximately 45,000 women from 87 general practices to participate in the first cohort, and it conducted randomization at the practice level. Because of baseline differences between the intervention and control arms, there are substantial concerns about the randomization in the Edinburgh trial. It is therefore not included in the meta-analysis that informs the US Preventive Services Task Force guidelines or in Table 1. Much later, in 1991, another trial began that included participants in Scotland, England, and Wales. That trial is known as the UK Age trial, or simply, the Age trial. Like the Edinburgh trial, the Age trial enrolled women based on their affiliation with general practices, but it conducted randomization at the individual level. An extension of the Age trial, the AgeX trial, began recruiting women aged 47 to 49 and 71 to 73 in 2009. As of late 2018, the AgeX trial has recruited 4 million women, making it over six times larger than the combined size of the trials included in the meta-analysis by Nelson et al. (2016). The trial is powered to detect a 15 percent reduction of breast cancer mortality (Patnick et al., 2018).

Latest Mortality Results across Trials

The meta-analysis by Nelson et al. (2016) that informs the US Preventive Services Task Force mammography guidelines primarily considers the relative risk of breast cancer mortality at the latest available date for each trial. Table 2 reproduces the main results. Within each age group, the table reports results from various studies that provide results for distinct groups

Table 1: Summary of Trial Characteristics

Trial Name	Year Trial Began	Setting and Population	Screening Group; Control Group, n^*	Method of Randomization	Adherence, %
HIP	1963	New York health plan members aged 40-64	30,239; 30,765	Age- and family size-stratified pairs of women were individually randomized by drawing from a list	46
CNBSS-1 and CNBSS-2	1980	Self-selected participants from 15 centers in Canada aged 40-49 (CNBSS-1), and 50-59 (CNBSS-2)	CNBSS-1: 25,214; 25,216. CNBSS-2: 19,711; 19,694	Individual within blocks stratified by center and 5-year age group after CBE	85
MMST I and MMST II	1976-1978	All women aged 43-69 born between 1908 and 1945 living in Malmö, Sweden	MMST I: 21,088; 21,195. MMST II: 9,581; 8,212	Individual, within birth year	70
Swedish Two-County Trial	1977	Women aged 40-70 from Östergötland and Kopparberg counties in Sweden	77,080; 55,985	Clusters, based on geographic units; blocks designed to be demographically homogenous	84
Stockholm	1981	Residents aged 40-64 from southeast greater Stockholm, Sweden	40,318; 19,943	Individual, by day of month; ratio of screening to control group 2:1	81
Gothenburg	1982	All women aged 39-59 born between 1923 and 1944 living in Gothenburg, Sweden	21,650; 29,961	Cluster, based on day of birth for 1923-1935 cohort (18%), by individual for 1936-1944 cohort (82%)	75
Age	1991	Women aged 39-41 from 23 National Health Service breast screening units in England, Scotland, and Wales	53,884; 106,956	Individual, stratified by general practitioner group with random number generation 1991-1992; 1992 onward, randomization via Health Authority computer system	57

CBE = clinical breast examination; CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of New York; MMST = Malmö Mammographic Screening Trial.

* Numbers of participants in screening and control groups vary by publication.

Source. Information taken directly from Nelson et al. (2016) Appendix Table 1.

Note. Some columns from Appendix Table 1 in Nelson et al. (2016) have been omitted, other columns have been included with modified headers and typesetting, and the rows have been reordered to correspond to the narrative in this paper. Although this table contains seven rows, Nelson et al. (2016) count CNBSS-1 and CNBSS-2 separately in their discussion of eight trials.

within the eight main trials. In all age groups, the relative risk of breast cancer mortality is less than one, indicating that the ratio of the breast cancer mortality rate in intervention to control is less than one. The relative risk is only statistically different from one for women aged 50–59 and 60–69.

Nelson et al. (2016) note that the implied reduction in breast cancer mortality is small in absolute terms. Screening 10,000 women aged 50–59 over 10 years prevents 7.7 breast cancer deaths, and screening the same number of women aged 60–69 over 10 years prevents 21.3 (Nelson et al., 2015). The number of breast cancer deaths avoided due to mammography screening seems considerably lower than the number of cause-specific deaths avoided due to screening for other cancers. As a point of comparison, the US Preventive Services Task Force guidelines on colorectal cancer screening report that it prevents 200 to 240 colorectal cancer deaths per 10,000 people aged 50 to 70 (Bibbins-Domingo et al., 2016).

Deaths attributed to a particular type of cancer may not capture the full mortality impact of screening. Cause of death coding can be a subjective enterprise. For example, mammography screening can increase the probability of a breast cancer diagnosis, which can increase the probability that a death is coded as a breast cancer death. Through this mechanism, mammography trial estimates can be biased against showing reductions in breast cancer mortality. However, mammography can also lead to overdiagnosis and other collateral harms, which may or may not be captured in breast cancer mortality. The overall direction of the bias is unclear. Thus, it seems useful to examine all-cause mortality to capture a wider range of potential benefits and harms from mammograms.

The meta-analysis that informs the US Preventive Services Task Force mammography guidelines includes an analysis of all-cause mortality in a technical document (Nelson et al., 2015). Table 3 reproduces the results here. Combined analysis of all trials indicates a relative risk of 0.99 for women aged 39 to 49, 0.97 for women aged 60 to 69, and 0.98 for women aged 70 to 74. None of these relative risks are significantly different from one at the 5 percent level. Furthermore, the relative risk is 1.02 for women aged 50 to 59, indicating net harm. Several individual studies also report relative risks greater than one. All of these relative risks are imprecise and should thus be interpreted with caution. Imprecision aside, relative risks greater than one are striking. Even if a trial shows overdiagnosis in terms of breast cancer incidence, overdiagnosis could still be innocuous in terms of all-cause mortality.

It should also be noted that, for many trials, the mean follow-up for all-cause mortality reported in Table 3 is much shorter than the mean follow-up for breast cancer mortality reported in Table 2. The lack of later follow-up is concerning, given that reductions in mortality due to prevention of late-stage cancers might be expected to happen relatively sooner in the follow-up period, while negative effects of overdiagnosis might manifest relatively later. In one trial, the Canadian National Breast Screening Study, results in terms of all-cause mortality have tended to weaken the case for mammography over time (Miller

Table 2: Effects on Breast Cancer Mortality at Latest Available Date

Author (Year)	Trial Name	Mean Follow-up, years	Relative Risk (95% CI)
Women aged 39 - 49 years			
Habbema et al. (1986)	HIP	14.0	0.75 (0.53 - 1.05)
Miller et al. (2014)	CNBSS-1	21.9	1.04 (0.87 - 1.24)
Nyström et al. (2002)*	MMST I	18.2	0.74 (0.42 - 1.29)
Nyström et al. (2002)*	MMST II	11.2	0.64 (0.39 - 1.06)
Tabár et al. (1995)	Östergötland	12.5	1.02 (0.52 - 1.99)
Tabár et al. (1995)	Kopparberg	12.5	0.73 (0.37 - 1.41)
Nyström et al. (2002)*	Stockholm	14.3	1.52 (0.80 - 2.88)
Bjurstam et al. (2003)	Gothenburg	13.8	0.69 (0.45 - 1.05)
Moss et al. (2015)	Age	17.5	0.93 (0.80 - 1.09)
Overall ($I^2 = 25\%$; $p = 0.230$)			0.92 (0.75 - 1.02)
Women aged 50 - 59 years			
Habbema et al. (1986)	HIP	14.0	0.83 (0.61 - 1.13)
Miller et al. (2014)	CNBSS-2	21.9	0.94 (0.78 - 1.13)
Nyström et al. (2002)*	MMST I	18.1	0.98 (0.75 - 1.29)
Tabár et al. (1995)	Östergötland	12.5	0.85 (0.52 - 1.38)
Tabár et al. (1995)	Kopparberg	12.5	0.48 (0.29 - 0.77)
Nyström et al. (2002)*	Stockholm	13.7	0.56 (0.32 - 0.97)
Bjurstam et al. (2003)	Gothenburg	13.8	0.83 (0.60 - 1.15)
Overall ($I^2 = 38.0\%$; $p = 0.139$)			0.86 (0.68 - 0.97)
Women aged 60 - 69 years			
Habbema et al. (1986)	HIP	14.0	0.85 (0.48 - 1.47)
Nyström et al. (2002)*	MMST I	15.5	0.64 (0.45 - 0.92)
Tabár et al. (1995)	Östergötland	12.5	0.62 (0.43 - 0.91)
Tabár et al. (1995)	Kopparberg	12.5	0.58 (0.35 - 0.96)
Nyström et al. (2002)*	Stockholm	13.1	0.94 (0.46 - 2.02)
Overall ($I^2 = 0.0\%$; $p = 0.739$)			0.67 (0.54 - 0.83)
Women aged 70 - 74 years			
Nyström et al. (2002)*	MMST I	13.6	0.98 (0.15 - 6.60)
Tabár et al. (1995)	Östergötland	12.5	0.82 (0.43 - 1.58)
Tabár et al. (1995)	Kopparberg	12.5	0.76 (0.42 - 1.36)
Overall ($I^2 = 0.0\%$; $p = 0.962$)			0.80 (0.51 - 1.28)

Meta-analysis of trials using the longest follow-up times available. CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of New York; MMST = Malmö Mammographic Screening Trial.

* Used short case accrual.

Source. Information Taken Directly from Nelson et al. (2016) Figure 1.

Note. CI = confidence interval. The columns and panels from Figure 1 in Nelson et al. (2016) have been included with modified headers and typesetting, and the rows have been reordered to correspond to the narrative in this paper. The full version of Figure 1 in Nelson et al. (2016) also contains plots of relative risk and the respective 95 percent CI. I^2 is a measure of inconsistency across studies, which ranges from 0% to 100%.

et al., 1992a,b, 2014). Later in this paper, I will offer some new evidence on the time profile of impacts on breast cancer mortality and all-cause mortality using the data from this trial.

It is hard to understate the controversy surrounding the results from the mammography trials. The idea that finding small, treatable cancers will save lives by stopping them from growing into larger malignant cancers is appealingly simple. Furthermore, the mammography trials began in an era in which sexism and paternalism toward women were much more overt than they are today. Spurred by the Breast Cancer Detection and Demonstration Project, manufacturers of mammography equipment were eager to expand their market. Billboard signs implored, “If you don’t get a mammogram, you need more than your breasts examined” (Pellerin, 2019).

For advocates of mammography, it was thus disappointing when the evidence from the randomized controlled trials was underwhelming. Some began to update their thinking on the value of mammography. Others challenged the methodologies of the trials themselves, particularly whether the randomization had been adequate. The meta-analysis that informs the US Preventive Services Task Force guidelines rated the randomization of all of the trials in Table 1 as “fair” quality but raised various concerns about specific studies in terms like: “Generally effective randomization and comparable groups are assembled initially, but some question remains whether some, although not major, differences occurred in follow-up,” “Important differential loss to follow-up or overall high loss to follow-up; adherence <80%,” “Numbers of participants unclear,” and “Did not maintain comparable groups (includes attrition, crossovers, adherence, contamination).” For comparison, a Cochrane review deemed most of the trials, including the second part of the Malmö trial (MMST II) to be “suboptimally” randomized, but it also deemed the Canadian National Breast Screening Study, the first part of the Malmö trial (MMST I), and the Age trial to be “adequately” randomized (Gøtzsche and Jørgensen, 2013).

Trial Results Inform Mammography Guidelines

The evidence from mammography trials has informed mammography guidelines in many countries. For example, Swedish trials are the only trials in Table 2 that show statistically significant decreases in breast cancer mortality in some age groups. Swedish national guidelines recommend mammograms for women aged 40 to 74 (Ebell et al., 2018).

In contrast, Canadian national guidelines “recommend not screening” with mammography for women aged 40 to 49 (Klarenbach et al., 2018) but “recommend screening with mammography” for women aged 50 to 74. As summarized in Table 2, the most recent mortality results from the Canadian National Breast Screening Study (Miller et al., 2014) show an imprecise increase in breast cancer mortality for women in their 40s and an imprecise decrease in breast cancer mortality for women in their 50s. The most recent breast cancer incidence results indicate overdiagnosis (Baines et al., 2016).

Many other high income countries, including Australia, France, Switzerland, and the

Table 3: Effects on All-Cause Mortality at Latest Available Date

Author (Year)	Trial Name	Mean Follow-up, years	Relative Risk (95% CI)
Women aged 39 - 49 years			
Miller et al. (2002)	CNBSS-1	13.0	1.00 (0.87 - 1.15)
Nyström et al. (2002)	Malmö II	9.1	1.03 (0.89 - 1.20)
Tabár et al. (1989)	Östergötland	7.9	0.93 (0.76 - 1.12)
Tabár et al. (1989)	Kopparberg	7.9	1.33 (1.01 - 1.77)
Frisell et al. (1997)	Stockholm	11.0	1.12 (0.55 - 2.41)
Bjurstam et al. (1997)	Gothenburg	10.0	0.98 (0.86 - 1.12)
Moss et al. (2006)	Age	10.7	0.97 (0.89 - 1.04)
Overall ($I^2 = 0.0\%$; $p = 0.478$)			0.99 (0.94 - 1.06)
Women aged 50 - 59 years			
Miller et al. (2000)	CNBSS-2	13.0	1.06 (0.96 - 1.18)
Tabár et al. (1989)	Östergötland	7.9	0.98 (0.87 - 1.11)
Tabár et al. (1989)	Kopparberg	7.9	1.00 (0.86 - 1.17)
Overall ($I^2 = 0.0\%$; $p = 0.588$)			1.02 (0.94 - 1.10)
Women aged 60 - 69 years			
Tabár et al. (1989)	Östergötland	7.9	0.98 (0.91 - 1.05)
Tabár et al. (1989)	Kopparberg	7.9	0.95 (0.87 - 1.04)
Overall ($I^2 = 0.0\%$; $p = 0.650$)			0.97 (0.90 - 1.04)
Women aged 70 - 74 years			
Tabár et al. (1989)	Östergötland	7.9	0.93 (0.87 - 1.01)
Tabár et al. (1989)	Kopparberg	7.9	1.05 (0.95 - 1.15)
Overall ($I^2 = 72.4\%$; $p = 0.057$)			0.98 (0.86 - 1.14)
Combined ages			
Aron and Prorok (1986)	HIP	10	0.99 (0.93 - 1.05)
Miller et al. (2014)	CNBSS-1 & 2	25	1.02 (0.98 - 1.06)
Nyström et al. (2002)	Malmö I	19.2	0.99 (0.97 - 1.01)
Nyström et al. (2002)	Malmö II	9.1	1.03 (0.89 - 1.20)
Nyström et al. (2002)	Östergötland	17.2	0.98 (0.95 - 1.01)
Nyström et al. (2002)	Stockholm	14.7	0.99 (0.95 - 1.03)
Nyström et al. (2002)	Gothenburg	13.2	0.94 (0.88 - 1.00)
Moss et al. (2006)	Age	10.7	0.97 (0.89 - 1.04)
Overall ($I^2 = 0.0\%$; $p = 0.577$)			0.99 (0.97 - 1.003)

CI = confidence interval; CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of Greater New York.

Source. Information Taken Directly from Nelson et al. (2015) Figures 7 and 8.

Note. Malmö = Malmö Mammographic Screening Trial. The top four panels take information from Nelson et al. (2015) Figure 8, and the bottom panel takes information from Figure 7. The columns and panels from these figures have been included with modified headers and typesetting, and the rows have been reordered to correspond to the narrative in this paper. The full versions of these figures also contain plots of relative risk and the respective 95 percent CI. The ages included for each trial in the “Combined ages” panel can be found in the “Age, year” column of Nelson et al. (2015) Figure 7, which has been omitted here. I^2 is a measure of inconsistency across studies, which ranges from 0% to 100%.

United Kingdom, do not recommend mammography for women in their 40s, and they also do not recommend against it as Canadian guidelines do (Ebell et al., 2018). However, the Swiss Medical Board recommended steps to limit screening programs in 2014 (Biller-Andorno and Jüni, 2014). In 2016, the French Minister of Health released results of an independent review that recommended that the national screening program end or undergo radical reforms (Barratt et al., 2018).

Within the United States, guidelines vary but tend to make selective recommendations similar to those of the US Preventive Services Task Force (CDC, 2020). For women in their 40s, the US Preventive Services Task Force, the American Academy of Family Physicians, and the American College of Physicians all leave the mammography decision up to individual women and their doctors, as does the American Cancer Society for women aged 40 to 44. In contrast, the American College of Radiology recommends regular mammograms for women in their 40s.

Evolution of Mortality Results within a Trial

The medical literature has commonly evaluated randomized controlled trials of mammography by using breast cancer mortality as the primary outcome, and it commonly examines the outcome only at the latest available date. The meta-analysis that informs the US Preventive Services Task Force mammography guidelines follows these practices too. However, breast cancer mortality may or may not capture mortality impacts of mammography that occur through overdiagnosis (and perhaps false positives). Indeed, some in the medical literature argue for a focus on all-cause mortality (Black et al., 2002). In addition, measuring the outcome of a mammography trial at the latest available date does not reveal how benefits and harms of mammography may develop over different timeframes.

In this section, I reanalyze data from the Canadian National Breast Screening Study (CNBSS, 2015), the trial with the longest follow-up length of all trials considered by the Nelson et al. (2016) meta-analysis, with an emphasis on how breast cancer mortality and all-cause mortality evolve over time. Each intermediate result that I examine reflects a fixed follow-up length since enrollment. In contrast, the CNBSS investigators examine results at the latest available calendar date, which is not quite the same thing. Because participants enrolled in the CNBSS over a period of five years, the most recent CNBSS results reflect follow-up lengths that vary from 20 to 25 years across participants. The practice of aggregating information across multiple follow-up lengths is common and reasonable because it preserves balance between intervention and control and incorporates all available information. For example, the meta-analysis that informs the US Preventive Services Task Force guidelines reports an estimate from the latest available date for each trial (Nelson et al., 2016) and reports the mean follow-up length, per Table 2. However, this practice makes trends difficult to interpret.

Mammography trials often reassess results as data accumulates and the latest available

date advances (Miller et al., 1992a,b, 2000, 2002, 2014), and some research in the medical literature has emphasized systematic analysis of cause-specific mortality over time (Miettinen et al., 2002; Hanley, 2005, 2010, 2011; Hanley et al., 2013; Liu et al., 2015). Here, I will show that a systematic assessment of all-cause mortality over time reveals additional information.

As a starting point, I examine breast cancer mortality and all-cause mortality 20 years after enrollment in this study, which is the maximum follow-up length available for all participants. Following the latest mortality results published by the Canadian National Breast Screening Study investigators (Miller et al., 2014), I begin by pooling results for women of all ages. The first row of Table 4 shows that the intervention arm experienced 20 fewer deaths from breast cancer per 100,000 women than the control arm. This imprecise reduction in breast cancer mortality in the intervention arm is consistent with the negative and statistically insignificant point estimate published by the CNBSS investigators at the latest available date (Miller et al., 2014), and it is replicated in the last column of the second panel of Table A.1 in the online Appendix available with the paper at the *JEP* website.

Table 4: Excess Breast Cancer Mortality and All-Cause Mortality Rates in Intervention 20 Years After Enrollment in the Canadian National Breast Screening Study

	(1)	(2)	(1) – (2)
	Intervention	Control	Intervention - Control
Breast Cancer Deaths (per 100,000)	904	924	-20 (65)
All-Cause Deaths (per 100,000)	7,969	7,880	89 (175)
N	44,925	44,910	

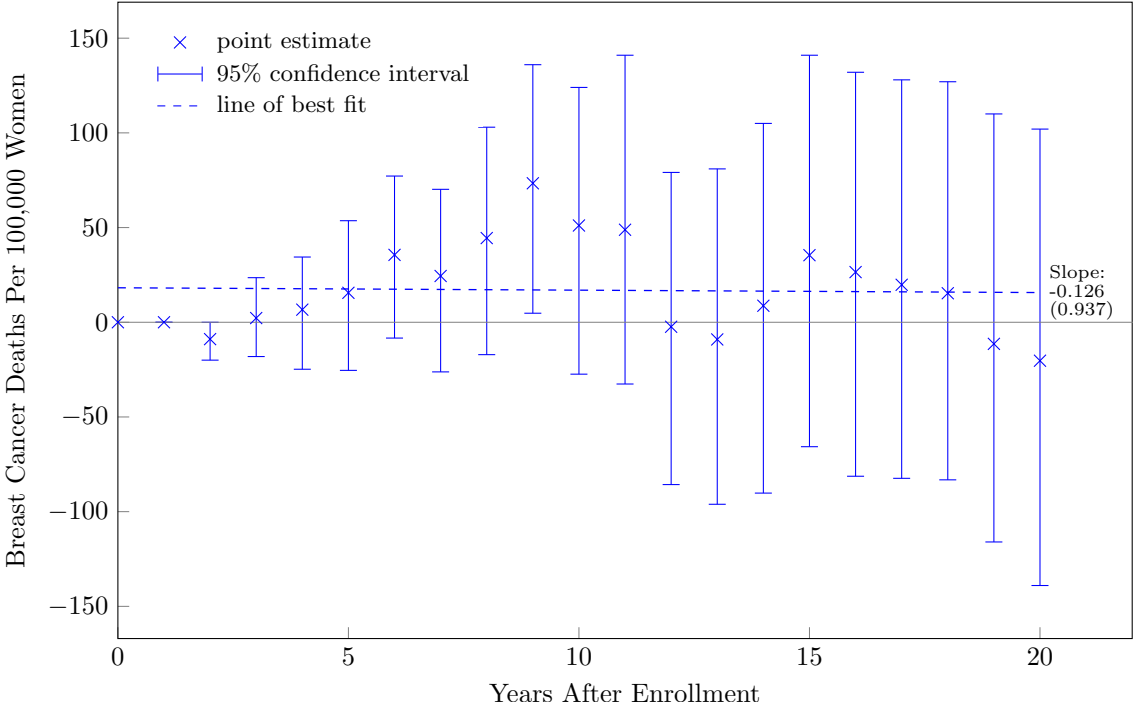
Note. Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. Standard errors in parentheses are calculated as the standard deviation of the point estimates obtained in 200 bootstrap samples. Subjects aged 40–59 at enrollment are included.

I next examine the excess all-cause mortality rate in the intervention arm 20 years after enrollment. The second row of Table 4 shows that the intervention arm experienced 89 more all-cause deaths per 100,000 women than the control arm. Again, this estimate is not statistically different from zero, and it is consistent with the positive but statistically insignificant point estimate published by the Canadian National Breast Screening Study investigators at the latest available date (Miller et al., 2014), replicated in the last column of the second panel of Table A.1 in the online Appendix.

The next step is to look at the evolution of breast cancer mortality over time. I start by constructing point estimates analogous to the point estimate reported in the first row of Table 4 for each intermediate annual follow-up length, starting with the enrollment year as year 0. I plot the results in Figure 2. All but two of the point estimates are statistically indistinguishable from zero. To investigate the existence of a trend, I plot the line of best fit across all point estimates. I obtain the statistical significance of the trend using the

equivalent panel regression to take into account that the point estimates as well as the trend are estimated. The trend line appears flat with a slope that is not statistically different from zero. Based on these results, it is not surprising that the literature, which focuses on breast cancer mortality as the primary outcome at various points in time, does not identify a compelling trend in breast cancer mortality as follow-up length increases.

Figure 2: Trend in Excess Breast Cancer Mortality Rate in Intervention in the Canadian National Breast Screening Study

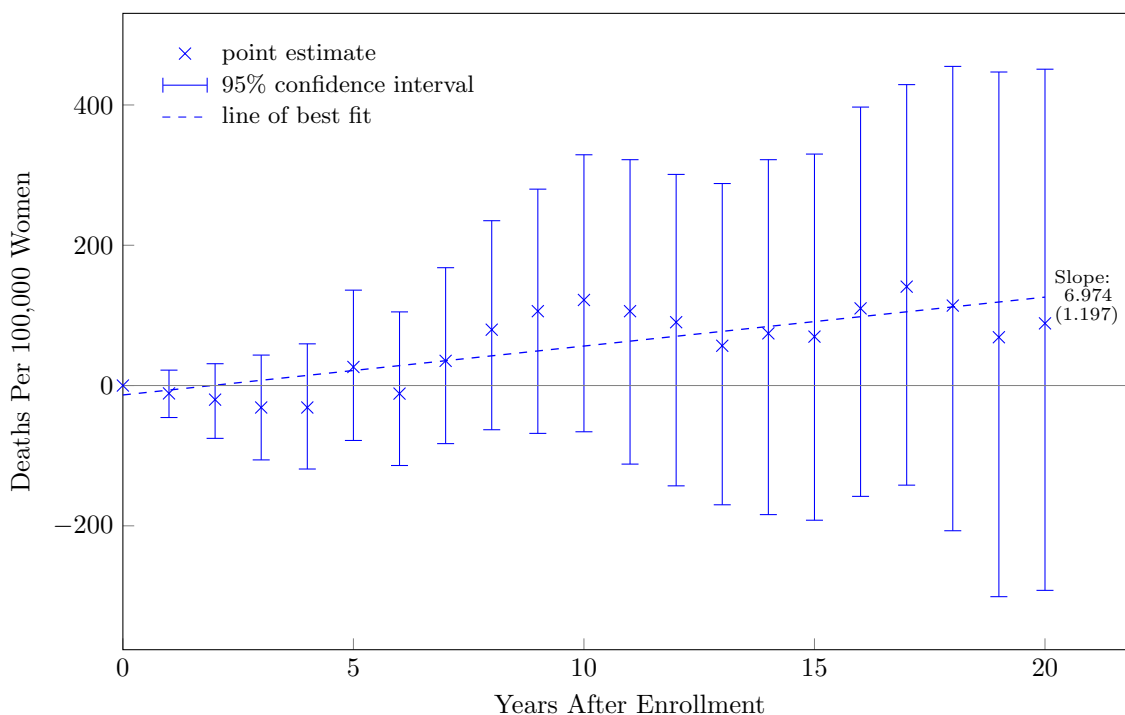


Note. Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. 95 percent confidence intervals in each year after enrollment are obtained from the same 200 bootstrap samples. The standard error on the slope of the line of best fit, in parentheses, is calculated from a panel regression, block bootstrapped by year after enrollment, which takes into account that the point estimates and the line of best fit are estimated. Subjects aged 40–59 at enrollment are included.

I next carry out the same approach with all-cause mortality as the outcome of interest. As shown in Figure 3, through 20 years after enrollment, point estimates are not statistically different from zero at any follow-up length. However, a pattern emerges over time: point estimates in the first four years after enrollment are negative. Year five is the first positive point estimate, and the point estimates are always positive after year six. Although these point estimates are individually imprecise, they are consistent with a hypothesis that life-saving benefits of mammograms outweigh their collateral harms at first, but as more time passes, collateral harms rise.

I plot the line of best fit across all point estimates, which I obtain from an equivalent panel regression, in Figure 3. A pronounced upward-sloping and statistically significant trend is visible in the all-cause mortality results. With each additional year that passes after enrollment, an additional 7 excess deaths per 100,000 women become apparent among

Figure 3: Trend in Excess All-Cause Mortality Rate in Intervention in the Canadian National Breast Screening Study



Note. Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. 95 percent confidence intervals in each year after enrollment are obtained from the same 200 bootstrap samples. The standard error on the slope of the line of best fit, in parentheses, is calculated from a panel regression, block bootstrapped by year after enrollment, which takes into account that the point estimates and the line of best fit are estimated. Subjects aged 40–59 at enrollment are included.

intervention arm participants relative to control arm participants. To put this trend in perspective, annual road traffic deaths in the United States are 12.4 per 100,000 (World Health Organization, 2018).

In Figures A.1 and A.2 in the online Appendix, I show that the positive trend is larger in magnitude for women aged 40–49 at enrollment (a slope of around 8 deaths per 100,000 women per year) than for women aged 50–59 at enrollment (a slope of around 6 deaths per 100,000 women per year). This difference by age group is unsurprising given the weaker evidence on the effect on breast cancer mortality for younger women and the corresponding difference in the US Preventive Services Task Force guidelines for women in their 40s.

There are many mechanisms that could explain the trend in all-cause mortality. For example, breast cancer treatments could have mortality impacts that are not captured by breast cancer mortality. Randomized controlled trials of radiotherapy show that it increases lung and esophageal cancer mortality among women treated for early-stage breast cancer (Early Breast Cancer Trialists’ Collaborative Group, 2005). I add lung and esophageal cancer deaths to breast cancer deaths and estimate the trend. From these three causes combined, I find that an additional 3.6 excess deaths per 100,000 women become apparent in each additional year after enrollment. This trend is statistically significant, and it can explain more than half of the trend in all-cause mortality (results shown in Figure A.3 in the

online Appendix). Perhaps deaths from other types of cancer can explain some of the rest of the difference. However, other cause-specific measures of mortality are subject to similar limitations as breast cancer mortality.

The trend in all-cause mortality is especially striking given that mammography likely increased in the control arm over time. Although the investigators did not collect complete data on mammograms received after the active study period, Canadian breast screening programs began in the late 1980s and 1990s, so it seems likely that some women in the control arm began receiving regular mammography at some point within 20 years of enrollment (Baines et al., 2016). Convergence in mammography behavior should attenuate mortality differences between the intervention and control arms over time—which makes the positive trend in excess all-cause mortality in the intervention arm all the more surprising. Convergence in mammography behavior also implies that the results speak to whether mammography should begin at younger ages, say, in a woman’s 40s rather than in her 50s, in line with the most recent change in the US Preventive Services Task Force guidelines.

These findings should be interpreted with caution. The Canadian National Breast Screening Study began decades ago. In any of the mammography trials, factors such as mammogram technology and breast cancer treatment could have changed over time, so past trends could differ from future trends in ways that are difficult to assess. For instance, technological improvements that allow mammograms to identify smaller tumors could worsen the mortality impact of overdiagnosis if those smaller tumors are less likely to advance to late-stage cancers. In contrast, innovations that make breast cancer treatment milder could alleviate the mortality impact of overdiagnosis.

Despite these challenges, evolving results from mammography trials represent the best available evidence. Examination of the trend in all-cause mortality reveals that the tradeoff between the harms and benefits of mammography has been shifting toward harms over time. It also offers a prediction of how the evidence will continue to evolve, which is useful since organizations such as the US Preventive Services Task Force update guidelines over time.

Responses to Evolving Evidence

One response to the evolving evidence on mortality from mammography trials is to incorporate it into new policies. Section 2713 of the [Affordable Care Act \(2010\)](#) ties health insurance coverage for preventive services to current and future guidelines of the US Preventive Services Task Force. However, it makes an explicit exception for mammography, allowing coverage for annual mammograms for women in their 40s, despite current guidelines.

Another response is to reverse existing policies that are at odds with current guidelines, such as policies that lead to over-utilization of mammography among women older than 74. For example, California state auditors implemented a policy in the late 1980s that required many elderly women enrolled in a particular health plan to receive at least one mammogram. The mean age of the women screened between 1995 and 1997 as a result of

the policy was 81. Unsurprisingly, 17 percent of these women experienced significant burdens from mammography (Walter et al., 2001). This policy was later reversed, but the issue of screening older women persists. One doctor told *Kaiser Health News* that doctors continue to screen older women because of fears of lawsuits, health system bonuses for high screening rates, and because “doing less can be perceived as a lack of caring or as ageism...It can be uncomfortable for a physician to explain why doing less is more” (Szabo, 2017).

Even if it is difficult to limit mammography, medical practice can respond to evolving evidence on mammography by reducing or at least postponing the use of potentially unnecessary treatments, especially treatments for early-stage cancers identified with mammography. *The New York Times* described the results of a recent clinical trial of breast cancer treatment as “good news for women with breast cancer: many don’t need chemo” (as reported in Grady, 2018). The trial found that the addition of chemotherapy to endocrine therapy does not increase invasive disease-free survival for women with certain breast cancer diagnoses, but it has some benefit for women aged 50 or younger whose breast cancers are initially deemed more likely to recur (Sparano et al., 2018).

Yet another response to the evolving evidence on mammography is to produce more targeted evidence. In that vein, it is instructive to learn from responses to experimental evidence in development economics. As Esther Duflo (2020) discusses in her Nobel lecture, initial responses from the microfinance industry were critical of experimental evidence that contradicted the perception of microfinance as a panacea against poverty. However, experimental evidence was never meant to pass a categorical judgment on microfinance; rather, it aimed to uncover contexts in which microfinance may work well. As with microfinance, the underwhelming evidence on benefits of mammography should not motivate a wholesale rejection of the practice, but rather it should motivate research aimed at uncovering the contexts in which mammograms may provide benefits.

Existing and new methods can be used to predict the characteristics of women for whom mammography is most beneficial within trial data. In my ongoing research, I propose an approach to quantify the number of individuals in a trial who would be harmed by an intervention, such as access to mammography, even if the trial does not show harm on average (Kowalski, 2020b). Advances in machine learning can then be applied to determine which women are most likely to benefit. With this knowledge in hand, examination of which women select into mammography (as in Kim and Lee (2017); Einav et al. (2019)) and how these selection patterns translate into heterogeneous treatment effects of mammography (as in Kowalski (2020a)) can help policymakers craft better targeted policies. There has already been careful work that examines the impact of policies on mammography (Mehta et al., 2015; Bitler and Carpenter, 2016; Kadiyala and Strumpf, 2016; Lu and Slusky, 2016; Buchmueller and Goldzahl, 2018; Myerson et al., 2020). Such work can be extended to examine the targeting of mammography in light of evidence on mortality from clinical trials.

Saumya Chatrath, Neil Christy, Tory Do, Simon Essig Aberg, Bailey Flanigan, Pauline Mourot, Srajal Nayak, Dominik Piehlmaier, Ljubica Ristovska, Sukanya Sravasti, and Matthew Tauzer provided excellent research assistance. Timothy Taylor provided exceptionally helpful editing, and Zoey Chopra, Gordon Hanson, Aaron Kesselheim, Enrico Moretti, Heidi Williams, and David Wilson provided helpful comments. I thank Anthony Miller, Teresa To, Cornelia Baines, and Claus Wall for sharing data from the Canadian National Breast Screening Study and for answering questions and providing background information. They have reviewed this manuscript per our data use agreement. All errors are my own. NSF CAREER Award 1350132 and NIA Grant P30-AG12810 provided support. I dedicate my research on breast cancer to Elisa Long.

References

- Affordable Care Act (2010). The Patient Protection and Affordable Care Act, Sec. 2713, Coverage of Preventive Services.
- Alexander, F., T. Anderson, H. Brown, A. Forrest, W. Hepburn, A. Kirkpatrick, B. Muir, R. Prescott, and A. Smith (1999). 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *The Lancet* 353(9168), 1903–1908.
- American Cancer Society (2017). History of ACS recommendations for the early detection of cancer in people without symptoms. <https://www.cancer.org/healthy/find-cancer-early/cancer-screening-guidelines/chronological-history-of-acs-recommendations.html>. Online. Last Revised July 7, 2017. Accessed March 12, 2018.
- Aron, J. and P. Prorok (1986). An analysis of the mortality effect in a breast cancer screening study. *International journal of epidemiology* 15(1), 36–43.
- Autier, P., M. Boniol, C. LaVecchia, L. Vatten, A. Gavin, C. Héry, and M. Heanue (2010). Disparities in breast cancer mortality trends between 30 European countries: Retrospective trend analysis of WHO mortality database. *Bmj* 341, c3620.
- Baines, C. J. (1984). Impediments to recruitment in the Canadian National Breast Screening Study: Response and resolution. *Controlled Clinical Trials* 5(2), 129–140.
- Baines, C. J., T. To, and A. B. Miller (2016). Revised estimates of overdiagnosis from the Canadian National Breast Screening Study. *Preventive medicine* 90, 66–71.
- Barratt, A., K. J. Jørgensen, and P. Autier (2018). Reform of the national screening mammography program in france. *JAMA internal medicine* 178(2), 177–178.
- Bibbins-Domingo, K., D. C. Grossman, S. J. Curry, K. W. Davidson, J. W. Epling, F. A. García, M. W. Gillman, D. M. Harper, A. R. Kemper, A. H. Krist, et al. (2016). Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *JAMA* 315(23), 2564–2575.
- Biller-Andorno, N. and P. Jüni (2014). Abolishing mammography screening programs? a view from the swiss medical board. *New England Journal of Medicine* 370(21), 1965–1967.
- Bitler, M. P. and C. S. Carpenter (2016). Health insurance mandates, mammography, and breast cancer diagnoses. *American Economic Journal: Economic Policy* 8(3), 39–68.
- Bjurstam, N., L. Björneld, S. W. Duffy, T. C. Smith, E. Cahlin, O. Eriksson, L.-O. Hafström, H. Lingaas, J. Mattsson, S. Persson, et al. (1997). The Gothenburg breast screening trial. *Cancer* 80(11), 2091–2099.

- Bjurstam, N., L. Björnelid, J. Warwick, E. Sala, S. W. Duffy, L. Nyström, N. Walker, E. Cahlin, O. Eriksson, L.-O. Hafström, et al. (2003). The Gothenburg breast screening trial. *Cancer* 97(10), 2387–2396.
- Black, W. C., D. A. Haggstrom, and H. Gilbert Welch (2002). All-cause mortality in randomized trials of cancer screening. *Journal of the National Cancer Institute* 94(3), 167–173.
- Bleyer, A. and H. G. Welch (2012). Effect of three decades of screening mammography on breast-cancer incidence. *New England Journal of Medicine* 367(21), 1998–2005.
- Boyes, D., J. Knowelden, and A. Phillips (1973). The evaluation of cancer control measures. *British journal of cancer* 28(2), 105–107.
- Buchmueller, T. C. and L. Goldzahl (2018). The effect of organized breast cancer screening on mammography use: Evidence from france. *Health economics* 27(12), 1963–1980.
- CDC (2020). Breast cancer screening guidelines for women.
- CNBSS (2015). Canadian National Breast Screening Study: Cumulative Data 1980-2005. University of Toronto [distributor].
- Cunningham, M. P. (1997). The breast cancer detection demonstration project 25 years later. *CA: A cancer journal for clinicians* 47(3), 131–133.
- Dufo, E. (2020). Field experiments and the practice of policy. *American Economic Review* 110(7), 1952–73.
- Early Breast Cancer Trialists’ Collaborative Group (2005). Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: An overview of the randomised trials. *The Lancet* 366(9503), 2087–2106.
- Ebell, M. H., T. N. Thai, and K. J. Royalty (2018). Cancer screening recommendations: an international comparison of high income countries. *Public health reviews* 39(1), 7.
- Einav, L., A. Finkelstein, T. Oostrom, A. J. Ostriker, and H. L. Williams (2019). Screening and selection: The case of mammograms. *The American Economic Review*. Forthcoming.
- Frisell, J., E. Lidbrink, L. Hellström, and L.-E. Rutqvist (1997). Followup after 11 years—update of mortality results in the stockholm mammographic screening trial. *Breast cancer research and treatment* 45(3), 263–270.
- Gøtzsche, P. C. and K. J. Jørgensen (2013). Screening for breast cancer with mammography. *Cochrane database of systematic reviews* (6).

- Grady, D. (2018). Good news for women with breast cancer: Many don't need chemo. *New York Times*. Online. Accessed July 27, 2020.
- Habbema, J., G. J. v. Oortmarssen, D. J. van Putten, J. T. Lubbe, and P. J. v. d. Maas (1986). Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. *Journal of the National Cancer Institute* 77(2), 317–320.
- Hanley, J. A. (2005). Analysis of mortality data from cancer screening studies: Looking in the right window. *Epidemiology*, 786–790.
- Hanley, J. A. (2010). Mortality reductions produced by sustained prostate cancer screening have been underestimated. *Journal of Medical Screening* 17(3), 147–151.
- Hanley, J. A. (2011). Measuring mortality reductions in cancer screening trials. *Epidemiologic reviews* 33(1), 36–45.
- Hanley, J. A., M. McGregor, Z. Liu, E. C. Strumpf, and N. Dendukuri (2013). Measuring the mortality impact of breast cancer screening. *Canadian Journal of Public Health* 104(7), e437–e442.
- Jørgensen, K. J., P.-H. Zahl, and P. C. Gøtzsche (2010). Breast cancer mortality in organised mammography screening in denmark: comparative study. *BMJ* 340.
- Kadiyala, S. and E. Strumpf (2016). How effective is population-based cancer screening? Regression discontinuity estimates from the US guideline screening initiation ages. In *Forum for Health Economics and Policy*, Volume 19, pp. 87–139. De Gruyter.
- Kim, H. B. and S.-m. Lee (2017). When public health intervention is not successful: Cost sharing, crowd-out, and selection in Korea's National Cancer Screening Program. *Journal of health economics* 53, 100–116.
- Klarenbach, S., N. Sims-Jones, G. Lewin, H. Singh, G. Thériault, M. Tonelli, M. Doull, S. Courage, A. J. Garcia, and B. D. Thombs (2018). Recommendations on screening for breast cancer in women aged 40–74 years who are not at increased risk for breast cancer. *CMAJ: Canadian Medical Association Journal* 190(49), E1441.
- Kowalski, A. E. (2020a, November). Behavior within a clinical trial and implications for mammography guidelines. Working Paper 25049, National Bureau of Economic Research.
- Kowalski, A. E. (2020b, July). Counting defiers: Examples from health care. *arXiv preprint arXiv:1912.06739*. <https://arxiv.org/abs/1912.06739>.

- Kunst, N., J. B. Long, X. Xu, S. H. Busch, K. A. Kyanko, I. B. Richman, and C. P. Gross (2020, 05). Use and Costs of Breast Cancer Screening for Women in Their 40s in a US Population With Private Insurance. *JAMA Internal Medicine* 180(5), 799–801.
- Liu, Z., J. A. Hanley, O. Saarela, and N. Dendukuri (2015). A conditional approach to measure mortality reductions due to cancer screening. *International Statistical Review* 83(3), 493–510.
- Lu, Y. and D. J. Slusky (2016). The impact of women’s health clinic closures on preventive care. *American Economic Journal: Applied Economics* 8(3), 100–124.
- Marcus, P. M., E. J. Bergstralh, M. H. Zweig, A. Harris, K. P. Offord, and R. S. Fontana (2006). Extended lung cancer incidence follow-up in the mayo lung project and overdiagnosis. *Journal of the National Cancer Institute* 98(11), 748–756.
- Mehta, S. J., D. Polsky, J. Zhu, J. D. Lewis, J. T. Kolstad, G. Loewenstein, and K. G. Volpp (2015). Aca-mandated elimination of cost sharing for preventive screening has had limited early impact. *The American journal of managed care* 21(7), 511.
- Miettinen, O. S., C. I. Henschke, M. W. Pasmantier, J. P. Smith, D. M. Libby, and D. F. Yankelevitz (2002). Mammographic screening: No reliable supporting evidence? *The Lancet* 359(9304), 404–405.
- Miller, A., G. Howe, and C. Wall (1981). The national study of breast cancer screening. *Clinical & Investigative Medicine* 4, 227–258.
- Miller, A. B., C. J. Baines, T. To, and C. Wall (1992a). Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *CMAJ: Canadian Medical Association Journal* 147(10), 1459–1476.
- Miller, A. B., C. J. Baines, T. To, and C. Wall (1992b). Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *CMAJ: Canadian Medical Association Journal* 147(10), 1477–1488.
- Miller, A. B., T. To, C. J. Baines, and C. Wall (2000). Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. *Journal of the National Cancer Institute* 92(18), 1490–1499.
- Miller, A. B., T. To, C. J. Baines, and C. Wall (2002). The Canadian National Breast Screening Study-1: Breast cancer mortality after 11 to 16 years of follow-up: a randomized screening trial of mammography in women age 40 to 49 years. *Annals of internal medicine* 137(5.1), 305–312.

- Miller, A. B., C. Wall, C. J. Baines, P. Sun, T. To, and S. A. Narod (2014). Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: Randomised screening trial. *British Medical Journal* 348.
- Moss, S. M., H. Cuckle, A. Evans, L. Johns, M. Waller, L. Bobrow, et al. (2006). Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *The Lancet* 368(9552), 2053–2060.
- Moss, S. M., C. Wale, R. Smith, A. Evans, H. Cuckle, and S. W. Duffy (2015). Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. *The Lancet Oncology* 16(9), 1123–1132.
- Mukherjee, S. (2017, Sep). Cancer's invasion equation. *The New Yorker*.
- Myerson, R. M., R. D. Tucker-Seeley, D. P. Goldman, and D. N. Lakdawalla (2020). Does medicare coverage improve cancer detection and mortality outcomes? *Journal of Policy Analysis and Management*.
- National Cancer Institute (1979). *A History of cancer control in the United States, 1946-1971*. U.S. National Cancer Institute, Division of Cancer Control and Rehabilitation.
- National Center for Health Statistics (2011). *Health, United States, 2010: With Special Feature on Death and Dying*. Hyattsville, Maryland.
- National Center for Health Statistics (2019). *Health, United States, 2018*. Hyattsville, Maryland.
- Nelson, H. D., A. Cantor, L. Humphrey, R. Fu, M. Pappas, M. Daeges, and J. Griffin (2015). Screening for Breast Cancer: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Technical Report 14-05201-EF-1, AHRQ, Rockville, MD: Agency for Healthcare Research and Quality. <https://www.uspreventiveservicestaskforce.org/uspstf/document/draft-evidence-review-screening-for-breast-cancer/breast-cancer-screening>. Online. Accessed July 8, 2020.
- Nelson, H. D., R. Fu, A. Cantor, M. Pappas, M. Daeges, and L. Humphrey (2016). Effectiveness of breast cancer screening: Systematic review and meta-analysis to update the 2009 US Preventive Services Task Force recommendation. *Annals of internal medicine* 164(4), 244–255.
- Newhouse, J. P. (2020, July). An ounce of prevention. Working Paper 27553, National Bureau of Economic Research.

- Nyström, L., I. Andersson, N. Bjurstam, J. Frisell, B. Nordenskjöld, and L. E. Rutqvist (2002). Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *The Lancet* 359(9310), 909–919.
- Nyström, L., S. Wall, L. Rutqvist, A. Lindgren, M. Lindqvist, S. Ryden, J. Andersson, N. Bjurstam, G. Fagerberg, J. Frisell, et al. (1993). Breast cancer screening with mammography: Overview of Swedish randomised trials. *The Lancet* 341(8851), 973–978.
- Patnick, J., K. Baker, I. Barnes, V. Beral, H. Pan, R. Peto, G. Reeves, M. Richards, and K. Shaw (2018). Nationwide cluster-randomised trial of extending the NHS breast screening age range in England: AgeX trial protocol. <http://www.agex.uk/>. Online. Accessed November 10, 2020.
- Pellerin, R. (2019). *Conspiracy of Hope: The Truth about Breast Cancer Screening*. Goose Lane Editions.
- Praga, C., J. Bergh, J. Bliss, J. Bonnetterre, B. Cesana, R. C. Coombes, P. Fargeot, A. Folin, P. Fumoleau, R. Giuliani, et al. (2005). Risk of acute myeloid leukemia and myelodysplastic syndrome in trials of adjuvant epirubicin for early breast cancer: Correlation with doses of epirubicin and cyclophosphamide. *Journal of clinical oncology* 23(18), 4179–4191.
- Ravdin, P. M., K. A. Cronin, N. Howlader, C. D. Berg, R. T. Chlebowski, E. J. Feuer, B. K. Edwards, and D. A. Berry (2007). The decrease in breast-cancer incidence in 2003 in the United States. *New England Journal of Medicine* 356(16), 1670–1674.
- Shapiro, S. (1977). Evidence on screening for breast cancer from a randomized trial. *Cancer* 39(6), 2772–2782.
- Shapiro, S., P. Strax, and L. Venet (1971). Periodic breast cancer screening in reducing mortality from breast cancer. *Jama* 215(11), 1777–1785.
- Siu, A. L. (2016). Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Annals of internal medicine* 164(4), 279–296.
- Sparano, J. A., R. J. Gray, D. F. Makower, K. I. Pritchard, K. S. Albain, D. F. Hayes, C. E. Geyer Jr, E. C. Dees, M. P. Goetz, J. A. Olson Jr, et al. (2018). Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *New England Journal of Medicine* 379(2), 111–121.
- Summary Report of the Working Group to Review the National Cancer Institute-American Cancer Society Breast Cancer Detection Demonstration Projects (1979). *JNCI: Journal of the National Cancer Institute* 62(3), 647–650.

- Surveillance, Epidemiology, and End Results Program (2019). SEER*Stat Database: Mortality - All COD, Aggregated With State, Total U.S. (1969-2017) (Katrina/Rita Population Adjustment), National Cancer Institute, DCCPS, Surveillance Research Program. Underlying mortality data provided by NCHS.
- Surveillance, Epidemiology, and End Results Program (2020). SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2019 Sub (1975-2017) - Linked To County Attributes - Time Dependent (1990-2017) Income/Rurality, 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program.
- Szabo, L. (2017). Doing more harm than good? Epidemic of screening burdens nation's older patients. *Kaiser Health News*.
- Tabar, L., G. Fagerberg, H.-H. Chen, S. W. Duffy, C. R. Smart, A. Gad, and R. A. Smith (1995). Efficacy of breast cancer screening by age. new results swedish two-county trial. *Cancer* 75(10), 2507–2517.
- Tabar, L., G. Fagerberg, S. W. Duffy, and N. E. Day (1989). The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *Journal of Epidemiology & Community Health* 43(2), 107–114.
- US Preventive Services Task Force (1989). Guide to clinical preventive services: Report of the US Preventive Services Task Force.
- US Preventive Services Task Force (2002). Screening for breast cancer: Recommendations and rationale. *Annals of Internal Medicine* 137(5), 344–346.
- US Preventive Services Task Force (2009). Screening for breast cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine* 151(10), 716–726.
- Walter, L. C., C. Eng, and K. E. Covinsky (2001). Screening mammography for frail older women: What are the burdens? *Journal of general internal medicine* 16(11), 779–784.
- Welch, H. G. and W. C. Black (2010). Overdiagnosis in cancer. *Journal of the National Cancer Institute* 102(9), 605–613.
- World Health Organization (2018). Global status report on road safety 2018. Licence: CC BY-NC-SA 3.0 IGO. http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/. Online.
- Zackrisson, S., I. Andersson, L. Janzon, J. Manjer, and J. P. Garne (2006). Rate of overdiagnosis of breast cancer 15 years after end of malmö mammographic screening trial: follow-up study. *Bmj* 332(7543), 689–692.

Online Appendix

Mammograms and Mortality: How Has the Evidence Evolved?

by Amanda E. Kowalski

Replication of Published Results

The last column of the second panel of Table A.1 replicates published results in [Miller et al. \(2014\)](#), which considers evidence up to the 2005 calendar year, as opposed to 20 years after enrollment. The results are qualitatively similar.

Table A.1: Excess Breast Cancer Mortality and All-Cause Mortality Rates in Intervention 20 Years After Enrollment Versus Up to 2005 Calendar Year in the Canadian National Breast Screening Study

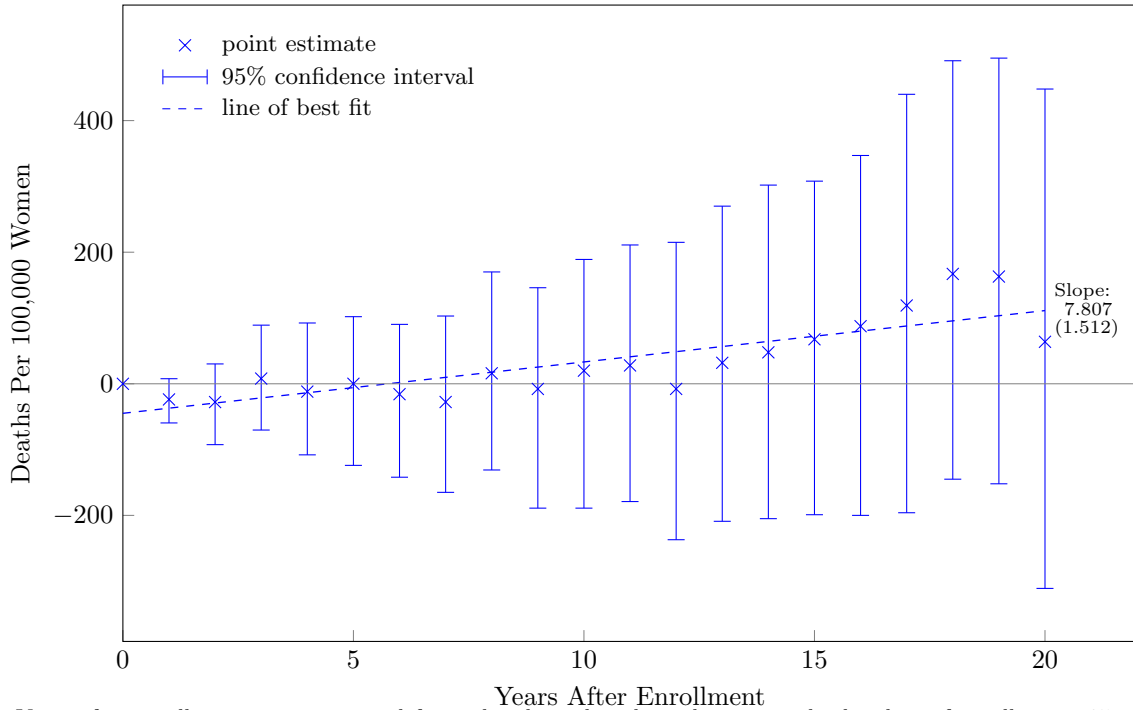
	(1) Intervention	(2) Control	(1) – (2) Intervention - Control	(1)/(2) Relative Risk
20 Years After Enrollment				
Breast Cancer Deaths (per 100,000)	904	924	-20 (65)	0.98 (0.07)
All-Cause Deaths (per 100,000)	7,969	7,880	89 (175)	1.01 (0.02)
Up to 2005 Calendar Year				
Breast Cancer Deaths (per 100,000)	1,113	1,124	-12 (71)	0.99 (0.06)
All-Cause Deaths (per 100,000)	10,660	10,439	221 (195)	1.02 (0.02)
N	44,925	44,910		

Note. Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. Standard errors in parentheses are calculated as the standard deviation of the point estimates obtained in 200 bootstrap samples. Subjects aged 40–59 at enrollment are included. Some differences between statistics might not appear internally consistent because of rounding.

Separate Trends for Women in Their 40s and 50s

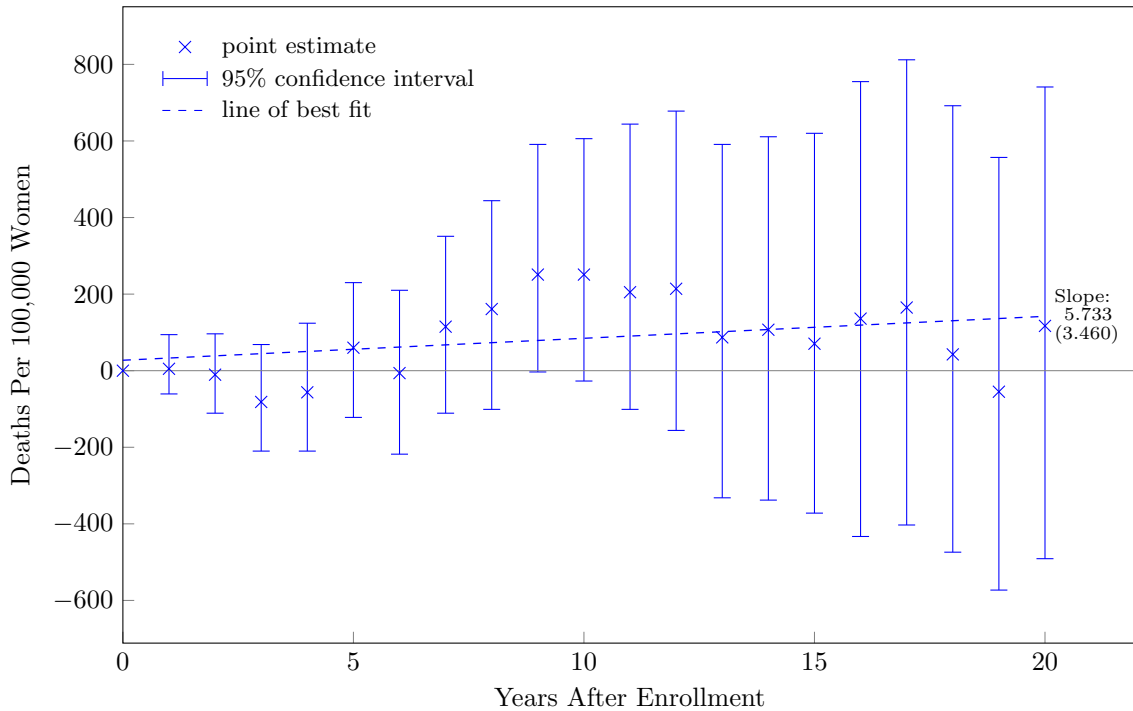
Figures A.1 and A.2 present separate trends for women in their 40s and 50s at enrollment. Despite losses in statistical power that result from dividing the sample, the magnitude of the time trend is similar within each age group. With each additional year of follow-up, mammography assignment yields 8 additional excess deaths per 100,000 women aged 40–49 and 6 additional excess deaths per 100,000 women aged 50–59.

Figure A.1: Trend in Excess All-Cause Mortality Rate in Intervention for Women Aged 40–49 at Enrollment in the Canadian National Breast Screening Study



Note. Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. 95 percent confidence intervals in each year after enrollment are obtained from the same 200 bootstrap samples. The standard error on the slope of the line of best fit is calculated from a panel regression, block bootstrapped by year after enrollment, which takes into account that the point estimates and the line of best fit are estimated. Subjects aged 40–49 at enrollment are included.

Figure A.2: Trend in Excess All-Cause Mortality Rate in Intervention for Women Aged 50–59 at Enrollment in the Canadian National Breast Screening Study

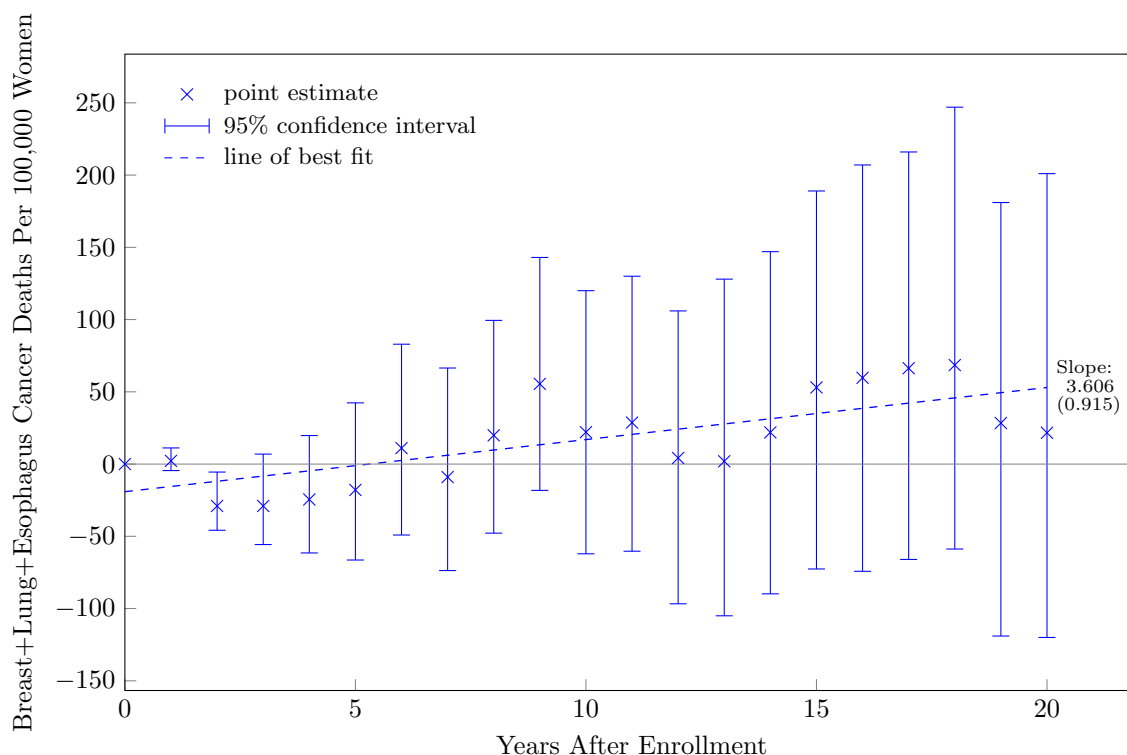


Note. Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. 95 percent confidence intervals in each year after enrollment are obtained from the same 200 bootstrap samples. The standard error on the slope of the line of best fit is calculated from a panel regression, block bootstrapped by year after enrollment, which takes into account that the point estimates and the line of best fit are estimated. Subjects aged 50–59 at enrollment are included.

Trend in Mortality from Breast, Lung, and Esophageal Cancer Combined

Figure A.3 shows that there is a positive, statistically significant trend in excess deaths due to breast, lung, and esophageal cancers combined: with each additional year of follow-up after enrollment, there are an additional 3.6 excess deaths per 100,000 women in intervention relative to control, which explains almost half of the trend in all-cause mortality depicted in Figure 3.

Figure A.3: Trend in Excess Mortality Rate from Breast, Lung, and Esophageal Cancers Combined in Intervention in the Canadian National Breast Screening Study



Note. Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. 95 percent confidence intervals in each year after enrollment are obtained from the same 200 bootstrap samples. Lung cancer mortality is defined using ICD-9 code 162 for subjects who died before 2000 and using ICD-10 codes C33-34 for subjects who died from 2000 onward. Esophageal cancer mortality is defined using ICD-9 code 150 for subjects who died before 2000 and using ICD-10 code C15 for subjects who died from 2000 onward. The standard error on the slope of the line of best fit, in parentheses, is calculated from a panel regression, block bootstrapped by year after enrollment, which takes into account that the point estimates and the line of best fit are estimated. Subjects aged 40–59 at enrollment are included.

References

Miller, A. B., C. Wall, C. J. Baines, P. Sun, T. To, and S. A. Narod (2014). Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: Randomised screening trial. *British Medical Journal* 348.