NBER WORKING PAPER SERIES

BIOLOGY MEETS BEHAVIOR IN A CLINICAL TRIAL: TWO RELATIONSHIPS BETWEEN MORTALITY AND MAMMOGRAM RECEIPT

Amanda E. Kowalski

Working Paper 25049
http://www.nber.org/papers/w25049

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2018

Saumya Chatrath, Tory Do, Bailey Flanigan, Pauline Mourot, Dominik Piehlmaier, Ljubica Ristovska, Sukanya Sravasti, and Matthew Tauzer provided excellent research assistance. 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. I thank Magne Mogstad, Atheendar Venkataramani, and seminar participants at the American Economic Association Annual Meeting, the Canadian Health Economics Study Group, the North American Summer Meetings of the Econometric Society, and Princeton for helpful comments. NSF CAREER Award 1350132 and NIA Grant P30-AG12810 provided support. I dedicate my research on breast cancer to Elisa Long. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Amanda E. Kowalski. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Biology Meets Behavior in a Clinical Trial: Two Relationships Between Mortality and Mammogram Receipt

Amanda E. Kowalski

NBER Working Paper No. 25049
September 2018
JEL No. C18,I1,I12

ABSTRACT

I unite the medical and economics literatures by examining relationships between biology and behavior in a clinical trial. Specifically, I identify relationships between mortality and mammogram receipt using data from the Canadian National Breast Screening Study, an influential clinical trial on mammograms. I find two important relationships. First, I find heterogeneous selection into mammogram receipt: women more likely to receive mammograms are healthier. This relationship follows from a marginal treatment effect (MTE) model that assumes no more than the local average treatment effect (LATE) assumptions. Second, I find treatment effect heterogeneity along the mammogram receipt margin: women more likely to receive mammograms are more likely to be harmed by them. This relationship follows from an ancillary assumption that builds on the first relationship. My findings contribute to the literature concerned about harms from mammography by demonstrating variation across the mammogram receipt margin. This variation poses a challenge for current mammography guidelines for women in their 40s, which unintentionally encourage more mammograms for healthier women who are more likely to be harmed by them.

Amanda E. Kowalski
Department of Economics
University of Michigan
611 Tappan Ave.
Lorch Hall 213
Ann Arbor, MI 48109-1220
and NBER
aekowals@umich.edu

1 Introduction

The U.S. Preventive Service Task Force (USPSTF) revived the debate on mammography when they updated their mammography guidelines in 2009. Although they previously recommended regular mammograms for women in their 40s, the updated guidelines left the mammography decision for women in this age range to individual women and their doctors. The precise USPSTF guidelines for women in their 40s, confirmed in 2016, state: "The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences" (U.S. Preventive Service Task Force, 2017). The empirical relationship between which women benefit from mammograms, based on biology, and which women receive mammograms, based on behavior, is crucial to the impact of these guidelines.

While the medical literature has focused on biology and the economics literature has focused on behavior, I aim to unite both literatures by examining relationships between the two. The medical literature cited by the USPSTF in its guidelines on mammography (Nelson et al., 2016; Moss et al., 2015; Miller et al., 2014; Tabár et al., 2011; Bjurstam et al., 2003; Frisell et al., 1997; Andersson et al., 1988; Shapiro et al., 1982) focuses on health outcomes from large clinical trials but says little about mammogram receipt behavior. In contrast, the economics literature on mammography focuses on mammogram receipt behavior. It sometimes relates mammogram receipt behavior to health outcomes, but it says little about how *variation* in mammogram receipt behavior relates to *variation* in health outcomes (Zanella and Banerjee, 2016; Kadiyala and Strumpf, 2016; Buchmueller and Goldzahl, 2018; Myerson et al., 2018). Kim and Lee (2017) is an exception. Using a regression discontinuity design, they find evidence that women more likely to receive mammograms are healthier, thus identifying a relationship between biology and behavior.

I identify two relationships between biology and behavior using data from a clinical trial and a generalized Roy (1951) model of the marginal treatment effect (MTE) as introduced by Björklund and Moffitt (1987), in the tradition of Heckman and Vytlacil (1999, 2001, 2005), Carneiro et al. (2011), and Brinch et al. (2017). I begin with an MTE model that assumes no more than the local average treatment effect (LATE) assumptions of Angrist and Imbens (1994), as shown by Vytlacil (2002). I model behavior in the first stage and relate it to biology in the second stage. In the first stage, differences in behavior determine whether individuals are always takers, compliers, or never takers, in the terminology of Angrist et al. (1996). The model implies that always takers are the first to receive treatment, followed by compliers, and then never takers. In the second stage, differences in outcomes determine whether biology varies with behavior. By comparing average outcomes across always takers, compliers, and never takers, I demonstrate that it is possible to identify two relationships between biology and behavior in existing clinical trial data. The first relationship identifies

heterogeneous selection into mammography under no ancillary assumptions. The second relationship identifies treatment effect heterogeneity from mammography along the same margin under an ancillary assumption that builds on the first empirical relationship.

I apply the model to data from the Canadian National Breast Screening Study (CNBSS), an extensive trial on mammography cited by the USPSTF in its mammography guidelines. The CNBSS enrolled about 90,000 participants between 1980 and 1985. In the CNBSS, some participants were randomly assigned to an intervention group that received access to annual mammograms during an active study period, consisting of the enrollment year and 3 to 4 years after enrollment. Remaining participants were assigned to the control group. Control group women in their 40s at enrollment received usual care in the community, and control group women in their 50s at enrollment received annual clinical breast examinations during the active study period. The CNBSS data tracks mammogram receipt in each year of the active study period for all participants, including participants in the control group, allowing me to examine behavior through mammogram receipt. Through linkage to the Canadian Mortality Database, the CNBSS data also tracks mortality for all participants through at least 20 years after enrollment, allowing me to examine biology through mortality. Given the controversy surrounding mammography guidelines for women in their 40s, I focus on women aged 40-49 at enrollment, representing 50,430 participants, but I examine robustness using the remaining women aged 50-59 at enrollment.

Applying the MTE model to the CNBSS, I identify two relationships between biology and behavior. First, I find that women who are more likely to receive mammograms are healthier. In terms of the MTE model, this is a finding of heterogeneous selection into treatment, where the "treatment" is mammogram receipt. I identify heterogeneous selection into treatment using a test proposed in the econometric literature by Bertanha and Imbens (2014), Guo et al. (2014), Black et al. (2015) and generalized by Mogstad et al. (2017). This test is also comparable to a test proposed in the insurance literature by Einav et al. (2010). Unlike related tests proposed by Hausman (1978); Heckman (1979); Willis and Rosen (1979); Angrist (2004); Huber (2013), and Brinch et al. (2017), this test does not require any assumptions beyond the LATE assumptions. In Kowalski (2018a,b), I refer to the test as the "untreated outcome test" because it compares the average untreated outcomes of compliers and never takers, and I show that it identifies heterogeneous selection into treatment. In the CNBSS context, heterogeneous selection into treatment identifies a relationship between biology and behavior.

Under an ancillary assumption that builds on the first empirical relationship, I identify a second relationship between biology and behavior. In terms of the MTE model, the second relationship identifies treatment effect heterogeneity along the margin of mammogram receipt. To identify this relationship, I assume weak monotonicity of average untreated out-

comes along the margin of mammogram receipt. This assumption is weaker than related assumptions made by Olsen (1980), Heckman (1979), and Brinch et al. (2017), as discussed by Kline and Walters (2018). Brinch et al. (2017) impose this assumption in conjunction with a corresponding assumption on treated outcomes to test for treatment effect homogeneity. In Kowalski (2016, 2018b), I demonstrate that either assumption is sufficient. Accordingly, in the CNBSS, I only impose one ancillary assumption. The model imposes the assumption of LATE monotonicity in the first stage, as shown by Vytlacil (2002). The ancillary assumption imposes a corresponding weak monotonicity in the second stage, which implies weak monotonicity of average untreated outcomes from always takers to compliers to never takers. In the context of the CNBSS, the ancillary assumption implies that health, measured by mortality in the absence of mammograms, varies monotonically with mammogram receipt. The direction of the monotonicity depends on the first empirical relationship between mammograms and health. In the CNBSS, the ancillary assumption implies that always takers are weakly healthier than compliers because compliers are healthier than never takers.

Covariates collected at baseline provide support for the ancillary assumption. The ancillary assumption implies an upper or lower bound on the average untreated outcome for always takers, which is not observed during the trial. Baseline covariates, which are observed for always takers, can proxy for untreated outcomes. Across several baseline covariates, always takers have higher average socioeconomic status than compliers, who have higher average socioeconomic status than never takers. Therefore, given a positive relationship between socioeconomic status and health without mammograms, health without mammograms should decrease from always takers to compliers to never takers, consistent with the empirical finding and the ancillary assumption. I also find a similar monotonic relationship in baseline health behaviors, providing a potential mechanism.

Applying the ancillary assumption to the CNBSS, I obtain an upper bound on average mortality for always takers without mammograms. Because the treatment effect for always takers is the difference between their mortality with and without mammograms, the ancillary assumption also implies a lower bound on the average treatment effect for always takers. The lower bound on the average treatment effect for always takers is larger than the LATE, the average treatment effect on compliers, and they are statistically different from one another. Therefore, the second relationship that I find between biology and behavior in the CNBSS implies that women who are more likely to receive mammograms are more likely to be harmed by them.

The possibility that harms of mammograms can outweigh benefits is surprising, but an extensive literature considers the possibility (Bleyer and Welch, 2012; Baum, 2013; Miller et al., 2014; Baines et al., 2016; Nelson et al., 2016; Lannin and Wang, 2017). To illustrate a potential mechanism, suppose that two women receive mammograms. Both are diagnosed

with breast cancer, and both indeed have breast cancer. Unbeknownst to the women and their doctors, one woman would die within 20 years in the absence of breast cancer treatment, but the other woman would not because her tumor would grow more slowly. Unable to separate the two, both women receive breast cancer treatment, which has its own mortality risks. Both women die within 20 years. The first woman would have died in the absence of breast cancer treatment, so she is neither harmed nor helped. However, the second women would have survived in the absence of breast cancer treatment, so she is harmed. In this example, the harms of mammograms outweigh the benefits on average.

My findings, which show that health and the net harms from mammography vary along the mammogram receipt margin, pose a challenge for mammography guidelines. The current USPSTF guidelines for women in their 40s leave the mammography decision to individual women and their doctors. My findings imply that under these guidelines, women more likely to receive mammograms are healthier and more likely to be harmed by them. Beyond the mammography context, my findings demonstrate the importance of examining the relationship between biology and behavior in a world that encourages personalized health care. Fortunately, some relationships between biology and behavior can be identified in existing clinical trial data.

In the next section, I begin by replicating previous results from the CNBSS. In Section 3, I apply the MTE model to the CNBSS, and I explain how it relates behavior in the first stage to biology in the second stage. In Section 4, I identify two relationships between biology and behavior using data from the CNBSS. I identify a first relationship between biology and behavior under the model alone. This relationship demonstrates heterogeneous selection into treatment: women more likely to receive mammograms are healthier. Under an ancillary assumption that builds on the first relationship, I identify a second relationship between biology and behavior. This relationship demonstrates treatment effect heterogeneity along the same margin: women more likely to receive mammograms are more likely to be harmed by them. Using covariates collected at baseline, I demonstrate support for the ancillary assumption. I show that my results are robust to a wide variety of alternative specifications in Section 5. I conclude by discussing implications for mammography guidelines and future research in Section 6.

2 Replication of CNBSS Results

A great deal has been written on the CNBSS in the medical literature. Viewing the CNBSS as an influential trial, my focus is not to evaluate the CNBSS itself or previous work on it. Rather, my focus is to extend analysis of the CNBSS to examine relationships between mortality and mammogram receipt. Using CNBSS data, I am able to produce an exact replication of the latest result published by CNBSS investigators in Miller et al. (2014), as

I report in Appendix B. This result shows that access to mammography does not have a statistically significant impact on breast cancer mortality, which is consistent with results published by CNBSS investigators at earlier follow-up lengths (Miller et al., 1992a,b, 1997, 2000, 2002, 2014). This result is are also consistent with other RCT results on mammography considered by the USPSTF in its 2016 mammography guidelines (Nelson et al., 2016).

In the replication that serves as the foundation for my extensions, I depart from the exact replication of Miller et al. (2014) in five ways. These departures facilitate further analysis of relationships between mammogram receipt, but they do not have a material impact on the result. In Appendix C, I demonstrate the robustness of the replication across all five departures. First, for consistency with the economics literature, I report the reduced form difference between the intervention and control groups instead of the relative risk ratio. Second, because breast cancer mortality could be endogenous to mammogram receipt and because collateral harms from mammograms could manifest themselves through causes of death that are not reported as breast cancer, I focus on all-cause mortality, which I refer to as "mortality" for simplicity. Third, for ease of interpretation, I report results at the maximal follow-up length of 20 years after enrollment for all subjects instead of reporting results at a fixed follow-up cutoff. Fourth, because the USPSTF guidelines changed specifically for women in their 40s, I only include women aged 40-49 in my main analysis sample and examine the robustness of my results for women aged 50-59. Fifth, to focus on relationships between mortality and mammogram receipt for women with no known clinical reasons to receive a mammogram before randomization occurred, I exclude women from my analysis sample if they have any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. I examine the robustness of my results to these sample restrictions. My main analysis sample includes 19,505 women.

3 Model

I use an MTE model to identify two relationships between biology and behavior in the CNBSS. I follow the exposition from Kowalski (2018a) closely, making only stylistic changes to the model used by Heckman and Vytlacil (2005) to ensure that the model assumes no more than the LATE assumptions of Angrist and Imbens (1994), as shown by Vytlacil (2002). Applying the model to the CNBSS, I model behavior in the first stage, and I relate biology to behavior in the second stage.

3.1 Behavior: Mammogram Receipt

In the context of the CNBSS, I use "treatment" to refer to mammogram receipt, which I represent with D. I define mammogram receipt D such that $D = 1$ if a participant receives

a mammogram in at least one year after enrollment during the active study period, and I set $D = 0$ otherwise. If mammogram data is missing in any year, I construct D such that the participant did not receive a mammogram in that year.

Let V_T represent potential utility in the treated state, the state with mammogram receipt, and let V_U represent potential utility in the untreated state, the state without mammogram receipt. I relate both potential utilities to realized utility V such that:

$$V = V_U + (V_T - V_U)D. \tag{1}$$

I specify the potential utilities as follows:

$$V_T = \mu_T(Z, X) + \nu_T \tag{2}$$

$$V_U = \mu_U(Z, X) + \nu_U, \tag{3}$$

where $\mu_T(\cdot)$ and $\mu_U(\cdot)$ are unspecified functions, X is an optional vector of observed covariates, Z is an observed binary instrument, and ν_T and ν_U are unobserved terms with unspecified distributions. In the CNBSS, the instrument represents random assignment to the intervention group such that $Z = 1$ for intervention group participants and $Z = 0$ for control group participants. I assume:

A.1. (First Stage Independence) The random variable $\nu_U - \nu_T$ is independent of Z conditional on X, which implies that $F(\nu_U - \nu_T \mid X)$, denoted as U_D, is independent of Z conditional on X.

A.2. (First Stage Technical Assumption) The cumulative distribution function of $\nu_U - \nu_T$ conditional on X, which I denote with F, is continuous and strictly increasing.

These assumptions imply the following equation for mammogram receipt conditional on random assignment:

$$D = 1\{U_D \leq \mathrm{P}(D = 1 \mid Z = z, X)\}, \tag{4}$$

where $U_D = F(\nu_U - \nu_T \mid X)$. I show for completeness in Appendix A.1 that this equation follows from the statement that participants receive mammograms if and only if their potential treated utility V_T exceeds their potential untreated utility V_U. Equivalently, under the mammogram receipt equation (4), participants receive mammograms if their values of U_D are less than the threshold $\mathrm{P}(D = 1 \mid Z = z, X)$. As I show for completeness in Appendix A.2, the model implies that U_D is distributed uniformly between 0 and 1. I interpret U_D as the "unobserved net cost of treatment," where the "net cost" is equal to the cost minus the benefit. In the CNBSS context, participants with the lowest unobserved net cost of mammogram receipt receive mammograms first.

There are two special cases of the mammogram receipt equation (4) for intervention and control group participants:

$$D = 1\{U_D \leq p_{CX}\} \qquad \text{where } p_{CX} = P\big(D = 1 \mid Z = 0, X\big), \tag{5}$$

$$D = 1\{U_D \leq p_{IX}\} \qquad \text{where } p_{IX} = P\big(D = 1 \mid Z = 1, X\big), \tag{6}$$

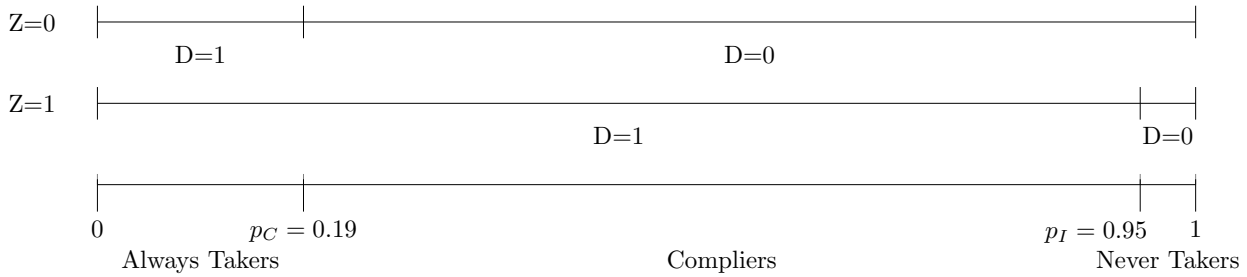
where the probabilities p_{CX} and p_{IX} can be estimated in the control group ($Z = 0$) and the intervention group ($Z = 1$), respectively. It is always possible to rename the intervention group as the control group and vice versa to satisfy $p_{CX} \leq p_{IX}$. I therefore proceed with $p_{CX} \leq p_{IX}$. I make the following assumptions, which are verifiable:

A.3. (First Stage Relevance) $\mu_T(Z, X) - \mu_U(Z, X)$ is a nondegenerate random variable conditional on X.

A.4. (First Stage Mammogram Receipt Differs from Random Assignment with Positive Probability) $0 < \mathrm{P}(D = 1 \mid Z = z, X) < 1$.

Under these assumptions, I partition the mammogram receipt margin U_D into distinct ranges. I depict the ranges in Figure 1. The top line depicts the ranges for control group participants. In my main analysis sample from the CNBSS, 19% of control group participants receive mammograms, so $p_C = 0.19$, where I suppress X to indicate that p_C represents an average in the full sample, not in a sample conditional on X. By (5), control group participants that receive mammograms have $0 \leq U_D \leq 0.19$. These participants must be always takers. The middle line of Figure 1 depicts ranges of U_D for intervention group participants. In my main analysis sample from the CNBSS, 95% of intervention group participants receive mammograms, so $p_I = 0.95$. By (6), intervention group participants that do not receive mammograms have $0.95 < U_D \leq 1$. These participants must be never takers. I depict U_D for participants in the control and intervention group on same axis in the bottom line of Figure 1. Participants in the middle range ($0.19 < U_D \leq 0.95$) receive mammograms if and only if they are in the intervention group, so they must be compliers. The depiction in the bottom line of Figure 1 is consistent with the ordering from always takers to compliers to never takers originally shown by Vytlacil (2002). In the CNBSS, this ordering identifies variation in behavior along the mammogram receipt margin: always takers receive mammograms first, followed by compliers, followed by never takers.

Figure 1: Ranges of U_D for Always Takers, Compliers, and Never Takers



U_D : unobserved net cost of treatment

3.2 Biology: Mortality

I relate mammogram receipt D to mortality Y as follows:

$$Y = Y_U + (Y_T - Y_U)D. \tag{7}$$

I specify treated mortality Y_T and untreated mortality Y_U such that:

$$Y_T = g_T(X, U_D, \gamma_T) \tag{8}$$

$$Y_U = g_U(X, U_D, \gamma_U), \tag{9}$$

where $g_T(\cdot)$ and $g_U(\cdot)$ are unspecified functions that need not be additively separable in their observed an unobserved components, unlike the potential utility functions in (2) and (3). X is the same optional vector of observed covariates from the first stage of the model, U_D is the unobserved net cost of treatment from the first stage of the model, and γ_T and γ_U are unobserved terms with unspecified distributions. I assume:

A.5. (Second Stage Independence) The random vector (U_D, γ_T) and the random vector (U_D, γ_U) are independent of Z conditional on X.

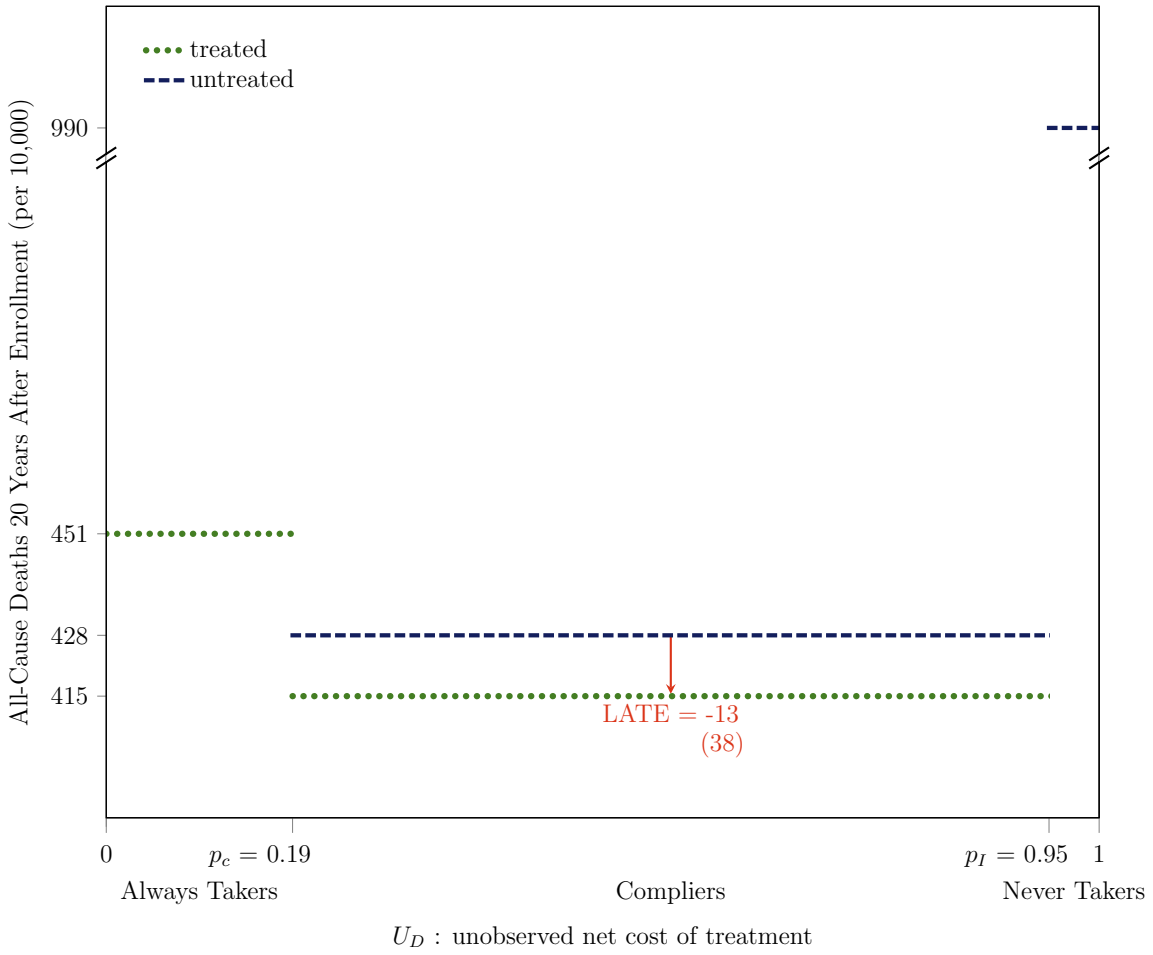
Under this final assumption, the model is equivalent to the LATE assumptions.

3.3 Biology and Behavior: Mortality and Mammogram Receipt

The model relates behavior to biology, as I show graphically in Figure 2. The horizontal axis depicts behavior via the mammogram receipt margin U_D, and the vertical axis depicts biology via mortality. Over the relevant ranges of the horizontal axis, I depict average outcomes that I obtain using the model, as I detail algebraically in Appendix A.3 and graphically in Appendix D. Imbens and Rubin (1997), Katz et al. (2001), Abadie (2002), and Abadie (2003) show how to obtain the same average outcomes using the LATE assumptions.

In Figure 2, I use dotted lines to represent average treated outcomes. As shown, 20 years after enrollment, always takers ($0 \leq U_D \leq p_C$) experienced 451 deaths per 10,000

Figure 2: Biology and Behavior:
Average Mortality for Always Takers, Compliers, and Never Takers
Under the LATE Assumptions



Note. Bootstrapped standard errors in parentheses. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer.

participants, while treated compliers ($p_C \leq U_D \leq p_I$) experienced 415 deaths per 10,000 participants. I use dashed lines to represent average untreated outcomes. Over the same time period, untreated compliers ($p_C \leq U_D \leq p_I$) experienced 428 deaths per 10,000 participants and never takers ($p_I \leq U_D \leq 1$) experienced 990 deaths per 10,000 participants. As I emphasize in Figure 2, average untreated outcomes are not observed for always takers, and average treated outcomes are not observed for never takers.

As originally shown by Imbens and Rubin (1997), the LATE is equal to the difference between the average treated and untreated outcomes for compliers. In Figure 2, I depict the

LATE with an arrow. The LATE is not statistically different from zero, but its magnitude indicates that compliers experienced an average decrease of 13 deaths per 10,000 participants. The LATE is also equal to the reduced form reported in Table A.2 divided by the first stage (-10 / (0.95-0.19) = -13). The depiction in Figure 2 makes clear that the LATE represents the average treatment effect on compliers but that always and never takers make up a substantial fraction of the sample, leaving room for the possibility of selection and treatment effect heterogeneity.

Within the model, I characterize selection and treatment effect heterogeneity along the entire margin of mammogram receipt using functions from the MTE literature (see Carneiro and Lee, 2009; Brinch et al., 2017):

$$\text{Marginal Untreated Outcome (MUO):} \quad \text{MUO}(x,p) = \text{E}\left[Y_U \mid X = x, U_D = p\right] \tag{10}$$

$$\text{Marginal Treated Outcome (MTO):} \quad \text{MTO}(x,p) = \text{E}\left[Y_T \mid X = x, U_D = p\right] \tag{11}$$

$$\text{Marginal Treatment Effect (MTE):} \quad \text{MTE}(x,p) = \text{E}\left[Y_T - Y_U \mid X = x, U_D = p\right] \tag{12}$$

where x is a realization of the covariate vector X and p is a realization of the unobserved net cost of treatment U_D. In the context of the CNBSS, these functions relate biology to behavior along the entire mammogram receipt margin U_D. In Kowalski (2016, 2018a,b), I refer to the first function as the "marginal treated outcome (MUO)" function, and I refer to the second function as the "marginal untreated outcome (MTO)" function. The difference between the MTO function and the MUO function yields the "marginal treatment effect (MTE)" function of Heckman and Vytlacil (1999, 2001, 2005).

In Kowalski (2018a), I show that the MUO function characterizes selection heterogeneity along U_D. As I show in Appendix A.4, a definition of selection heterogeneity used in the econometrics literature can be obtained as a weighted integral of the MUO function. Relative to definitions used in the literature, the MUO function is more general, and it does not depend on the fraction of individuals assigned to the intervention group, which is a feature of the experimental design. Intuitively, variation in average untreated outcomes across U_D can only be due to selection heterogeneity; it cannot be due to treatment effect heterogeneity because only treated outcomes can reflect treatment effects. In the CNBSS, the MUO function characterizes how selection on mortality changes along the mammogram receipt margin. Therefore, it characterizes a first relationship between biology and behavior in the CNBSS.

In Kowalski (2018a), I show that the MTO function characterizes the sum of selection and treatment effect heterogeneity along U_D. In contrast to average untreated outcomes, which can only reflect selection heterogeneity, average treated outcomes can reflect selection heterogeneity, treatment effect heterogeneity, or both. It is tempting to think that treated outcomes and untreated outcomes can be interchanged without consequence, but the treat-

ment effect is defined as the treated outcome minus the untreated outcome, not the untreated outcome minus the treated outcome. Therefore, the treatment effect has magnitude *and* direction. Renaming the untreated outcome as the treated outcome and vice versa would change the direction of the treatment effect, illustrating why there is a material distinction between treated and untreated outcomes.

The MTE function characterizes treatment effect heterogeneity along U_D. As I show in Kowalski (2018a) and in Appendix A.4, a definition of treatment effect heterogeneity used in the econometrics literature can be obtained as a weighted integral of the MTE function. In the CNBSS, the MTE function characterizes how the impact of mammography on mortality changes along the mammogram receipt margin. Therefore, it characterizes a second relationship between biology and behavior in the CNBSS.

4 Results: Two Relationships Between Mortality and Mammogram Receipt

Applying MTE methods to the CNBSS, I identify two main relationships between biology and behavior. First, under the model that assumes no more than the LATE assumptions, I find selection heterogeneity: women who are more likely to receive mammograms are healthier. Second, under an ancillary assumption, I find treatment effect heterogeneity along the margin of mammogram receipt: women more likely to receive mammograms are more likely to be harmed by them.

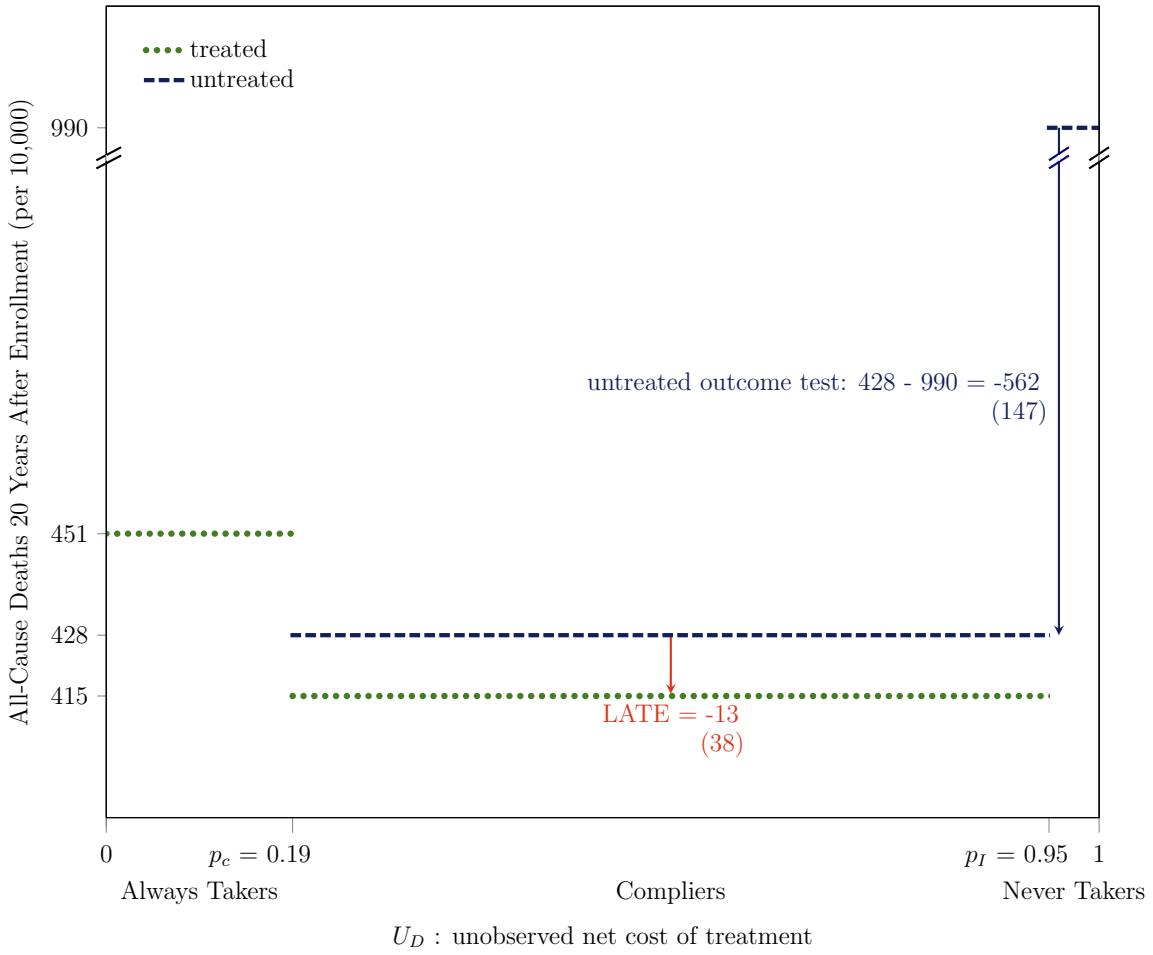
4.1 Selection Heterogeneity on Mortality Along the Margin of Mammogram Receipt

I test for selection homogeneity using the following test statistic, which gives the difference in average mortality without mammograms between compliers ($p_C < U_D \leq p_I$) and never takers ($p_I < U_D \leq 1$):

$$\mathrm{E}[Y_U \mid p_C < U_D \leq p_I] - \mathrm{E}[Y_U \mid p_I < U_D \leq 1] = \int_0^1 (\omega(p, p_C, p_I) - \omega(p, p_I, 1)) \, \mathrm{MUO}(p) \, \mathrm{d}p, \tag{13}$$

where $\omega(p, p_L, p_H) = 1\{p_L \leq p < p_H\}/(p_H - p_L)$. The test of the null hypothesis that this test statistic is equal to zero is equivalent to or similar to tests proposed by Bertanha and Imbens (2014), Guo et al. (2014), and Black et al. (2015), which are generalized by Mogstad et al. (2017). It is also comparable to a test proposed in the insurance literature by Einav et al. (2010). In Kowalski (2018a,b), I refer to the test as the "untreated outcome test" because it compares average untreated outcomes, and I show that it identifies a special case of selection heterogeneity. Intuitively, because never takers and compliers without mammograms

Figure 3: Untreated Outcome Test Shows Negative Selection Heterogeneity
Under the LATE Assumptions



Note. Bootstrapped standard errors in parentheses. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer.

are untreated, a difference in their average outcomes can only reflect selection heterogeneity. A negative test statistic indicates negative selection heterogeneity, and a positive test statistic indicates positive selection heterogeneity. Using algebra, (13) demonstrates that the untreated outcome test identifies a special case of selection heterogeneity because the test statistic can be obtained as a weighted integral of the MUO function, which characterizes selection heterogeneity over the entire selection margin.

Applying the untreated outcome test to data from the CNBSS, I find that never takers experienced 562 more deaths per 10,000 participants than untreated compliers, as depicted in Figure 3. The standard error of 147 indicates that the difference is statistically different

from zero. Therefore, I can reject selection homogeneity. The sign of the untreated outcome test statistic indicates that women more likely to receive mammograms are less likely to die from any cause 20 years after enrollment. Using mortality as a measure of health, I find the first of two relationships between biology and behavior in the CNBSS: women more likely to receive mammograms are healthier. This selection heterogeneity is consistent with the results of Kim and Lee (2017), who find evidence that women more likely to receive mammograms are healthier using a regression discontinuity design.

4.2 Treatment Effect Heterogeneity on Mortality Along the Margin of Mammogram Receipt

To test for treatment effect heterogeneity, I impose the following ancillary assumption:

M.1. (Weak Monotonicity of the MUO Function) For all p_1, $p_2 \in [0,1]$ such that $p_1 < p_2$: $\mathrm{E}[Y_U \mid U_D = p_1] \leq \mathrm{E}[Y_U \mid U_D = p_2]$ or $[Y_U \mid U_D = p_1] \geq \mathrm{E}[Y_U \mid U_D = p_2]$.

Brinch et al. (2017) impose this assumption in conjunction with an analogous assumption on the MTO function to test for treatment effect homogeneity. In Kowalski (2016, 2018b), I demonstrate that either assumption is sufficient. Weak monotonicity of the MUO function implies weak monotonicity of average untreated outcomes from always takers, to compliers, to never takers. Accordingly, in the CNBSS, I only impose Assumption M.1, which assumes weak monotonicity of the MUO function. In Appendix E, I present alternative weak monotonicity assumptions on the MTO and MTE functions, and I discuss why I do not impose them in the CNBSS. While the model imposes LATE monotonicity in the first stage, as shown by Vytlacil (2002), these assumptions impose corresponding weak monotonicities in the second stage.

In the context of the CNBSS, Assumption M.1 implies that average health, measured by mortality in the absence of mammograms, varies monotonically from always takers to compliers to never takers. The direction of the monotonicity depends on the first empirical relationship between mammogram receipt and health. In the CNBSS, compliers are healthier than never takers on average, so the ancillary assumption implies that always takers are weakly healthier than compliers on average. As illustrated in Figure 4, Assumption M.1 implies that always takers without mammograms would face no more than 428 deaths per 10,000 participants.

Under Assumption M.1, I test the null hypothesis of treatment effect homogeneity using the following decision rule, which has an outcome that is equal to 1 if the test rejects

treatment effect homogeneity and 0 otherwise:

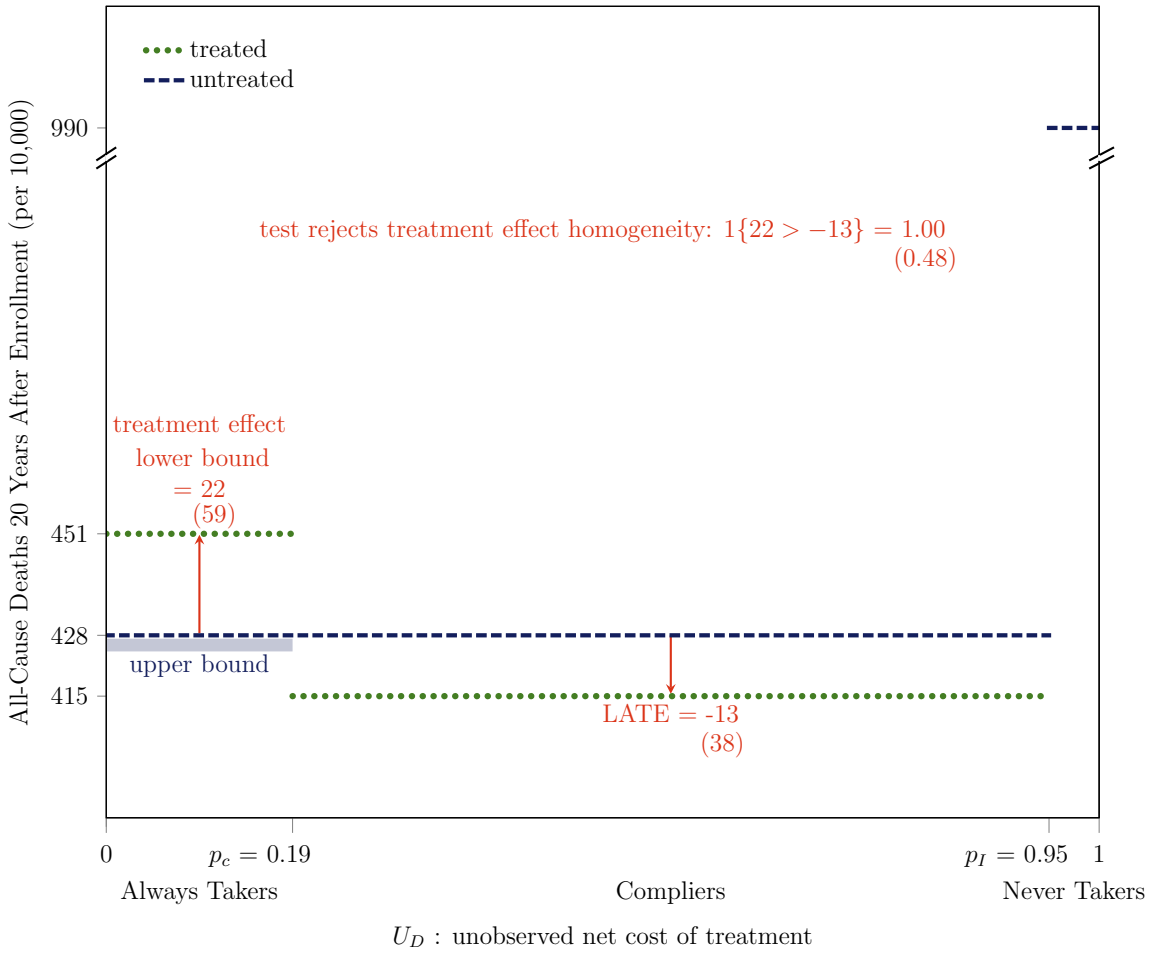
$$1\left\{\begin{array}{l}\mathrm{E}[Y_T|0 \leq U_D < p_C] - \mathrm{E}[Y_U|p_C < U_D \leq p_I] > \mathrm{E}[Y_T - Y_U|p_C \leq U_D \leq p_I] \\ \qquad\qquad\qquad\qquad \text{if } \mathrm{E}[Y_U \mid p_C < U_D \leq p_I] - \mathrm{E}[Y_U \mid p_I < U_D \leq 1] \leq 0, \\ \mathrm{E}[Y_T|0 \leq U_D < p_C] - \mathrm{E}[Y_U|p_C < U_D \leq p_I] < \mathrm{E}[Y_T - Y_U|p_C < U_D \leq p_I] \\ \qquad\qquad\qquad\qquad \text{if } \mathrm{E}[Y_U \mid p_C < U_D \leq p_I] - \mathrm{E}[Y_U \mid p_I < U_D \leq 1] > 0. \end{array}\right\} \tag{14}$$

As shown, this decision rule has two cases, which depend on the sign of the untreated outcome test statistic. If the untreated outcome test statistic is negative, then untreated compliers die at a lower rate than never takers. In this case, under Assumption M.1, average mortality for untreated compliers represents an *upper* bound on the average untreated mortality of always takers. I can therefore derive a *lower* bound on the average treatment effect of always takers that I can compare to the LATE to evaluate treatment effect homogeneity. Because the lower bound on the average treatment effect for always takers is strictly greater than the LATE, the average treatment effect for always takers cannot be equal to the LATE. However, if the untreated outcome test statistic were positive, then average mortality for untreated compliers would represent a *lower* bound on the average untreated outcome of always takers. Hence, I would derive an *upper* bound on the average treatment effect of always takers that must be strictly less than the LATE to reject treatment effect homogeneity.

Applying the test to the CNBSS, I reject treatment effect homogeneity. As reported in Figure 4, under Assumption M.1, I derive a lower bound on the average treatment effect of always takers that indicates that always takers experience at least an additional 22 deaths per 10,000 participants when they receive mammograms. The lower bound on the average treatment effect for always takers is strictly greater than the LATE, so always takers face a strictly greater average treatment effect than compliers. Therefore, the decision rule yields a value of one, and I can reject treatment effect homogeneity in the CNBSS as reported in column (4) of Table 2. The standard error of 0.48 shows that average treatment effects for always takers and compliers are statistically different from one another. Furthermore, the direction of heterogeneity in the average treatment effect indicates that women more likely to receive mammograms are more likely to be harmed by mammograms.

This finding seems surprising at first. However, anecdotal evidence from a clinical nurse suggest a potential mechanism: "I never, though, had a patient whose worry about those side effects came close to her worry about the disease. Being preoccupied with saving ones life produces a myopia, in which other worries unrelated to ones possibly imminent death fall away." (Brown, 2017). Healthier women might be more susceptible to this type of myopia because breast cancer represents a larger shock to their health. Given their fear of the disease, they might undertake more aggressive surgeries and treatments. Consequently,

Figure 4: Test Rejects Treatment Effect Homogeneity
Under the Ancillary Assumption of Weak Monotonicity of the MUO Function
and the LATE Assumptions



Note. Bootstrapped standard errors in parentheses. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.

they could be more likely to experience collateral harms from mammography.

4.2.1 Support for the Ancillary Assumption Using Baseline Covariates

I investigate support for the ancillary assumption in the CNBSS using covariates collected at baseline. The ancillary assumption implies an upper or lower bound on the average untreated outcome for always takers, which is not observed during the trial. Baseline covariates, which are observed for always takers, can proxy for their untreated outcomes.

I begin by examining variation in baseline covariates related to socioeconomic status

because socioeconomic status is known to be inversely correlated with mortality (Pappas et al., 1993; Cutler and Lleras-Muney, 2010; National Center for Health Statistics, 2012). Specifically, I compare average characteristics at baseline across always takers, compliers, and never takers in Table 1. I detail how I obtain average characteristics for compliers in Appendix A.5. Covariates related to socioeconomic status show a general pattern of monotonic variation from always takers to compliers to never takers, with always takers having the highest socioeconomic status. Furthermore, covariates related to health behaviors such as smoking status, body mass index, and mammograms prior to enrollment suggest a monotonic relationship across always takers, compliers, and never takers, where always takers exhibit the best health behaviors. Therefore, variation in baseline socioeconomic status and health behavior supports the assumption that average health decreases from always takers to compliers to never takers.

Table 1: Baseline Summary Statistics for Always Takers, Compliers, and Never Takers

	Means			Difference in Means	
	(1) Always Takers	(2) Compliers	(3) Never Takers	(1)-(2)	(2)-(3)
Baseline Socioeconomic Status					
University, trade or business school	0.50	0.46	0.39	0.04	0.08
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
In work force	0.65	0.64	0.65	0.02	-0.02
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Age at first birth	24.28	23.98	23.57	0.30	0.41
	(0.12)	(0.05)	(0.21)	(0.14)	(0.22)
No live birth	0.16	0.15	0.13	0.01	0.01
	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
Married	0.80	0.81	0.75	-0.01	0.06
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Husband in work force / alive	0.81	0.81	0.76	-0.00	0.05
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Baseline Health Behavior					
Non-Smoker	0.78	0.75	0.63	0.03	0.12
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Body Mass Index	23.87	24.42	24.48	-0.56	-0.06
	(0.10)	(0.05)	(0.21)	(0.12)	(0.22)
Used oral contraception	0.74	0.71	0.67	0.03	0.04
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Used estrogen	0.13	0.13	0.15	-0.00	-0.02
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Mammograms prior to enrollment	0.23	0.13	0.13	0.10	-0.00
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Practiced breast self examination	0.47	0.44	0.38	0.03	0.06
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)

Note. Bootstrapped standard errors in parentheses. Missing values correspond to redacted numbers in accordance with Data Use Agreement. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Baseline breast-related covariates are not reported here because they are all zero based on the sample restriction. Some differences between statistics might not appear internally consistent because of rounding.

5 Robustness

I examine the robustness of the two empirical relationships that I find between biology and behavior along many dimensions. Specifically, I examine alternative subsamples, alternative definitions of mammogram receipt, and alternative outcomes. Although I focus on women aged 40-49 at enrollment, I also examine results for women aged 50-59 at enrollment.

To facilitate comparisons with the main specification, I present tables that summarize the main results from Figures 2, 3, and 4, starting with Table 2. Column (1) reports the LATE. Column (2) reports the untreated outcome test statistic, which identifies the first relationship between biology and behavior: selection heterogeneity. When the untreated outcome test statistic is negative and statistically different from zero, this relationship shows that women more likely to receive mammograms are healthier. Column (4) reports the outcome of the decision rule in (14), which identifies the second relationship between biology and behavior: treatment effect heterogeneity. When the untreated outcome test statistic is negative, as it is in the main specification and almost all reported alternative specifications, column (3) gives a lower bound on the average treatment effect for always takers. If the lower bound in column (3) is greater than the LATE in column (1), then the average treatment effect on always takers must exceed the treatment effect on compliers. In this case, a rejection of treatment effect heterogeneity in column (4) indicates that women more likely to receive mammograms are more likely to be harmed by them.

5.1 Alternative Subsamples

5.1.1 Alternative Subsample Based on Breast-Related Covariates at Baseline

I investigate robustness of my results to the exclusion of participants with any nonzero breast-related covariates at baseline. To do so, I examine the subsample that only includes participants removed by this restriction and the full sample without this restriction. I begin by comparing baseline covariates across always takers, compliers, and never takers in the two samples in Tables A.5 and A.6 of Appendix F. Even in the sample of women with any nonzero breast-related covariates at baseline, covariates that measure socioeconomic status and health behavior lend support to the assumption that women more likely to receive mammograms are healthier. The breast-related covariates themselves suggest the opposite relationship. However, the breast-related covariates could be endogenous to breast cancer screening that occurred before the CNBSS began, especially since always takers report prior mammograms and breast self examination at a meaningfully higher rate than compliers. Nonetheless, the exclusion of women with any baseline breast-related covariates from the main analysis sample is conservative.

In Table 2, I summarize results in the alternative samples that include women with any nonzero breast-related covariates. In Appendix G, I report the full results necessary

to construct analogous versions of Figures 1-2 for the alternative samples. As shown, the untreated outcome test is negative and the test rejects treatment effect heterogeneity in both alternative samples, so the main results are robust.

Table 2: Alternative Samples and Outcomes

	N	(1) Local Average Treatment Effect LATE	(2) Untreated Outcome Test	(3) Always Taker Treatment Effect Lower Bound	(4) Test Rejects Treatment Effect Homogeneity
Main Specification					
Main specification	19,505	-13	-562	22	1.00
		(38)	(147)	(59)	(0.48)
Alternative Subsample Based on Breast-Related Covariates					
Any nonzero breast-related covariates	30,925	27	-759	60	1.00
		(40)	(135)	(39)	(0.47)
Both zero and nonzero breast-related covariates	50,430	9	-672	53	1.00
		(27)	(103)	(31)	(0.34)
Alternative Outcomes					
Breast cancer mortality	19,505	-12	-43	30	1.00
		(13)	(47)	(25)	(0.43)
Breast cancer incidence	19,505	58	-301	206	1.00
		(34)	(119)	(65)	(0.17)
Alternative Sample Based on Age Group					
Participants Aged 50-59 at Enrollment	17,210	67	-1,140	-149	0.00
		(60)	(226)	(103)	(0.20)

Note. Bootstrapped standard errors in parentheses. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. Baseline breast-related covariates refer to the following: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.

5.1.2 Alternative Subsamples Based on Enrollment Year and Center

I examine robustness of my findings in subsamples based on enrollment year in Appendix H and CNBSS center in Appendix I. In all but two of the 21 subsamples, women more likely to receive mammograms are healthier. This relationship is statistically different from zero in most subsamples, even though sample sizes shrink dramatically. Furthermore, in all but one of the subsamples by enrollment year in which I find the first relationship, I also find the second relationship: I reject treatment effect homogeneity such that women more likely to receive mammograms are more likely to be harmed by them. The subsamples by CNBSS center are generally smaller than the subsamples by enrollment year, but I still find the second relationship in many of them.

5.2 Alternative Definitions of Mammogram Receipt

In the CNBSS, I define mammogram receipt D such that $D = 1$ if a participant receives a mammogram in at least one year after enrollment during the active study period and I set $D = 0$ otherwise. If mammogram data is missing for a given participant in a given year, I construct D such that the participant did not receive a mammogram in that year. In Tables 3

and Appendix J, I consider narrower and broader definitions of mammogram receipt. Under the narrowest definition, participants must receive a mammogram in *all* active study period years after enrollment to be considered "treated." Under the broadest definition, participants must receive a mammogram *or* be missing mammogram data in any active study period year after enrollment to be considered "treated." The narrowest and broadest definitions are arguably too extreme, so it is notable that all reported specifications yield point estimates consistent with the first relationship from the main specification: women more likely to receive mammograms are healthier. Because this relationship holds, the reported results do not contradict the second relationship from the main specification, even in the two reported specifications in which the test does not reject treatment effect heterogeneity. To contradict the second relationship, implying that women more likely to receive mammograms are *less* likely to be harmed by them, the untreated outcome test would have to be positive and the test reported in the last column would have to reject treatment effect heterogeneity.

Table 3: Alternative Definitions of Mammogram Receipt

	N	(1) Local Average Treatment Effect LATE	(2) Untreated Outcome Test	(3) Always Taker Treatment Effect Lower Bound	(4) Test Rejects Treatment Effect Homogeneity
Main Specification					
Mammogram in at least one year after enrollment during the active study period, missing in year = no mammogram in year					
Main specification	19,505	-13	-562	22	1.00
		(38)	(147)	(59)	(0.48)
Narrower Definitions of Mammogram Receipt					
Mammogram in more than one year after enrollment during the active study period, missing in year = no mammogram in year					
At least two active study period years	19,505	-12	-465	-27	0.00
		(35)	(106)	(77)	(0.49)
At least three active study period years	19,505	-12	-420	56	1.00
		(36)	(94)	(145)	(0.48)
All active study period years	19,505	-15	-225	-135	0.00
		(42)	(75)	(138)	(0.37)
Broader Definitions of Mammogram Receipt					
Mammogram in at least one year after enrollment during the active study period					
Missing in year = mammogram in year	19,505	-24	-776	103	1.00
		(69)	(835)	(43)	(0.43)

Note. Bootstrapped standard errors in parentheses. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt. In the main specification, mammogram receipt is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.

5.3 Alternative Outcomes

5.3.1 Breast Cancer Mortality and Breast Cancer Incidence

For comparison with the literature, I examine breast cancer mortality and breast cancer incidence as alternatives to all-cause mortality. As reported in Table 2 and Appendix G, the results corroborate the main results. Measuring health in terms of either breast cancer outcome, women more likely to receive mammograms are healthier and more likely to have their health harmed by mammograms.

5.3.2 Mortality at Alternative Follow-up Lengths

I investigate the robustness of the main mortality results, which measure mortality 20 years after enrollment, to all earlier annual follow-up lengths in Table 4 and Appendix K. At all earlier follow-up lengths, the untreated outcome test statistic is negative, consistent with the first relationship that I find in the main specification: women more likely to receive mammograms are healthier. Furthermore, the test rejects treatment effect heterogeneity at some early follow-up lengths and at all follow-up lengths starting 15 years after enrollment, consistent with the second relationship that I find in the main specification: women more likely to receive mammograms are more likely to be harmed by them. The pattern of the results suggests that collateral harms from mammograms emerge over time.

5.4 Alternative Sample: Participants Aged 50-59 at Enrollment

Even though I focus on women aged 40-49 at enrollment because the change in the 2009 USPSTF recommendations affected this age group specifically, I examine the robustness of my results when focusing on women aged 50-59 at enrollment at the bottom of Table 2. As shown, selection heterogeneity goes in the same direction regardless of the age group at enrollment: women more likely to receive mammograms are significantly healthier. I cannot reject treatment effect homogeneity for women aged 50-59 at enrollment, but the results are still consistent with the results for women aged 40-49 at enrollment. The lower bound on the average treatment effect for always takers does not rule out values above the LATE, so it is possible that even women aged 50-59 women are more likely to receive mammograms are more likely to be harmed by them.

6 Conclusion

The success of public health interventions depends crucially on relationships between biology and behavior. Using an MTE model, I show that relationships between biology and behavior can be identified within a clinical trial. I apply this model to the CNBSS, an influential and extensive trial on mammography, and I identify two key relationships between mortality and mammogram receipt. First, under an MTE model that assumes no more than the LATE assumptions, I find that women more likely to receive mammograms are healthier. This

Table 4: Results for Mortality at Alternative Follow-Up Lengths

Years Since Enrollment	N	(1) Local Average Treatment Effect LATE	(2) Untreated Outcome Test	(3) Always Taker Treatment Effect Lower Bound	(4) Test Rjects Treatment Effect Homogeneity
Main specification: 20	19,505	-13	-562	22	1.00
		(38)	(147)	(59)	(0.48)
19	19,505	-13	-485	50	1.00
		(37)	(142)	(58)	(0.40)
18	19,505	-8	-492	54	1.00
		(35)	(139)	(56)	(0.41)
17	19,505	-8	-456	18	1.00
		(33)	(135)	(50)	(0.48)
16	19,505	-16	-471	15	1.00
		(31)	(134)	(46)	(0.47)
15	19,505	-15	-480	-11	1.00
		(31)	(131)	(42)	(0.50)
14	19,505	-21	-396	-38	0.00
		(30)	(121)	(38)	(0.45)
13	19,505	-24	-365	-30	0.00
		(28)	(115)	(36)	(0.49)
12	19,505	-27	-334	-23	1.00
		(27)	(106)	(32)	(0.50)
11	19,505	-10	-351	-30	0.00
		(25)	(105)	(28)	(0.42)
10	19,505	-15	-306	-41	0.00
		(23)	(97)	(25)	(0.37)
9	19,505	-12	-314	-35	0.00
		(20)	(97)	(21)	(0.36)
8	19,505	-2	-340	-14	0.00
		(18)	(97)	(21)	(0.44)
7	19,505	-6	-351	-15	0.00
		(17)	(97)	(18)	(0.46)
6	19,505	-5	-317	-24	0.00
		(15)	(93)	(16)	(0.33)
5	19,505	-5	-269	-12	0.00
		(13)	(86)	(15)	(0.45)
4	19,505	-9	-218	-3	1.00
		(11)	(77)	(14)	(0.49)
3	19,505	-6	-209	-3	1.00
		(9)	(76)	(11)	(0.50)
2	19,505	-3	-194	-3	1.00
		(9)	(67)	(9)	(0.50)
1	19,505	-5	-55	-5	0.00
		(5)	(40)	(5)	(0.00)

Note. Bootstrapped standard errors in parentheses. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.

relationship reflects heterogeneous selection. Second, under an ancillary assumption that builds on the first relationship, I find variation in the impact of mammography on mortality along the same mammogram receipt margin. This relationship reflects treatment effect heterogeneity from mammography. Putting both relationships together, women more likely to receive mammograms are healthier and are more likely to be harmed with mammograms. These relationships pose a challenge for current mammography guidelines for women in their 40s, which leave the mammogram receipt decision to individual women and their doctors. My results imply that the guidelines unintentionally encourage more mammograms for healthier women who are more likely to be harmed by them.

My results, though surprising at first, are consistent with a growing body of literature on potentially collateral harms on mammography. One potential explanation for my results is that healthier women are more likely to receive mammograms and that upon a breast cancer diagnosis, they are more likely to undertake more aggressive breast cancer treatments. Analysis of covariates collected at baseline lends support to this explanation by showing that women more likely to receive mammograms have higher socioeconomic status and better health behaviors. They might consequently be more responsive to a breast cancer diagnosis.

The main limitation of my analysis is that the active study period of the CNBSS took place in the 1980s, and medical progress since then might have altered the risks and benefits of mammography. However, changes in environment are an inherent limitation of long-term analysis. My results support the need for further evidence on mammography in the current environment.

My main contribution is to unite the medical and economics literatures by demonstrating an approach to examine relationships between biology and behavior in existing clinical trial data. I also illustrate how these relationships can inform public health recommendations. By taking into account relationships between biology and behavior, public health recommendations can target treatment to individuals most likely to benefit from them, and the combination of economics and medicine can advance progress toward personalized health care.

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American statistical Association 97*(457), 284–292.

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics 113*(2), 231–263.

Andersson, I., K. Aspegren, L. Janzon, T. Landberg, K. Lindholm, F. Linell, O. Ljungberg,

J. Ranstam, and B. Sigfusson (1988). Mammographic screening and mortality from breast cancer: the malmö mammographic screening trial. *Bmj 297*(6654), 943–948.

Angrist, J. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica 66*(2), 249–288.

Angrist, J. D. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal 114*(494), C52–C83.

Angrist, J. D. and G. W. Imbens (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

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.

Baum, M. (2013). Harms from breast cancer screening outweigh benefits if death caused by treatment is included. *Bmj 346*(jan23 1).

Bertanha, M. and G. W. Imbens (2014, December). External validity in fuzzy regression discontinuity designs. Working Paper 20773, National Bureau of Economic Research.

Björklund, A. and R. Moffitt (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, 42–49.

Bjurstam, N., L. Björneld, 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, D. A., J. Joo, R. LaLonde, J. A. Smith, and E. J. Taylor (2015). Simple tests for selection bias: Learning more from instrumental variables.

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.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond late with a discrete instrument. *Journal of Political Economy 125*(4), 000–000.

Brown, T. (2017, October). Breast cancer is serious. pink is not. *New York Times*.

Buchmueller, T. C. and L. Goldzahl (2018). The effect of organized breast cancer screening on mammography use: Evidence from france. Technical report, National Bureau of Economic Research.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011). Estimating marginal returns to education. *The American economic review 101*(6), 2754–2781.

Carneiro, P. and S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*(2), 191–208.

Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.

Cutler, D. M. and A. Lleras-Muney (2010). Understanding differences in health behaviors by education. *Journal of health economics 29*(1), 1–28.

Einav, L., A. Finkelstein, and M. R. Cullen (2010). Estimating welfare in insurance markets using variation in prices. *The Quarterly Journal of Economics 125*(3), 877.

Frisell, J., E. Lidbrink, L. Hellström, and L.-E. Rutqvist (1997). Follow up after 11 years–update of mortality results in the stockholm mammographic screening trial. *Breast cancer research and treatment 45*(3), 263–270.

Guo, Z., J. Cheng, S. A. Lorch, and D. S. Small (2014). Using an instrumental variable to test for unmeasured confounding. *Statistics in medicine 33*(20), 3528–3546.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 1251–1271.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–162.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998). Characterizing selection bias using experimental data. *Econometrica 66*(5), 1017–1098.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2001). Local instrumental variables. In C. Hsiao, K. Morimune, and J. L. Powell (Eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, pp. 1–46. Cambridge University Press.

Huber, M. (2013). A simple test for the ignorability of non-compliance in experiments. *Economics Letters 120*(3), 389–391.

Imbens, G. W. and D. B. Rubin (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies 64*(4), 555–574.

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.

Katz, L. F., J. R. Kling, J. B. Liebman, et al. (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics 116*(2), 607–654.

Kim, H. B. and S. 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.

Kline, P. M. and C. R. Walters (2018, March). On heckits, late, and numerical equivalence. Working Paper XX, National Bureau of Economic Research.

Kowalski, A. (2016, June). Doing more when you're running late: Applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments. Working Paper 22362, National Bureau of Economic Research.

Kowalski, A. (2018a, May). Extrapolation using selection and moral hazard heterogeneity from within the oregon health insurance experiment. Working Paper 24647, National Bureau of Economic Research.

Kowalski, A. (2018b, July). How to examine external validity within an experiment. Working Paper 24834, National Bureau of Economic Research.

Lannin, D. R. and S. Wang (2017). Are small breast cancers good because they are small or small because they are good? *New England Journal of Medicine 376*(23), 2286–91.

Manski, C. F. (1997). Monotone treatment response. *Econometrica: Journal of the Econometric Society*, 1311–1334.

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 (1997). The canadian national breast screening study: update on breast cancer mortality. *JNCI Monographs 1997*(22), 37–41.

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_Part_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. *Bmj 348*, g366.

Mogstad, M., A. Santos, and A. Torgovitsky (2017, July). Using instrumental variables for inference about policy relevant treatment effects. Working Paper 23568, National Bureau of Economic Research.

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.

Myerson, R. M., D. Lakdawalla, L. D. Colantonio, M. Safford, and D. Meltzer (2018). Effects of expanding health screening on treatment-what should we expect? what can we learn? Technical report, National Bureau of Economic Research.

National Center for Health Statistics (2012). Health, United States, 2011: With special feature on socioeconomic status and health.

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 effectiveness of breast cancer screening. *Annals of internal medicine 164*(4), 244–255.

Nelson, H. D., M. Pappas, A. Cantor, J. Griffin, M. Daeges, and L. Humphrey (2016). Harms of breast cancer screening: systematic review to update the 2009 us preventive services task force recommendationharms of breast cancer screening. *Annals of internal medicine 164*(4), 256–267.

Olsen, R. J. (1980). A least squares correction for selectivity bias. *Econometrica: Journal of the Econometric Society*, 1815–1820.

Pappas, G., S. Queen, W. Hadden, and G. Fisher (1993). The increasing disparity in mortality between socioeconomic groups in the united states, 1960 and 1986. *New England journal of medicine 329*(2), 103–109.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers 3*(2), 135–146.

Shapiro, S., W. Venet, P. Strax, L. Venet, and R. Roeser (1982). Ten-to fourteen-year effect of screening on breast cancer mortality 2. *Journal of the National Cancer Institute 69*(2), 349–355.

Tabár, L., B. Vitak, T. H.-H. Chen, A. M.-F. Yen, A. Cohen, T. Tot, S. Y.-H. Chiu, S. L.-S. Chen, J. C.-Y. Fann, J. Rosell, et al. (2011). Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology 260*(3), 658–663.

U.S. Preventive Service Task Force (2017). Grade definitions. u.s. preventive services task force. november 2017. https://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions. Online. Accessed June 8, 2018.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

Willis, R. J. and S. Rosen (1979). Education and self-selection. *Journal of political Economy 87*(5, Part 2), S7–S36.

Zanella, G. and R. Banerjee (2016). Experiencing breast cancer at the workplace. *Journal of Public Economics 134*, 53–66.

Online Appendix

Biology Meets Behavior in a Clinical Trial: Two Relationships Between Mortality and Mammogram Receipt

by Amanda E. Kowalski

Contents

A Proofs and Derivations

A.1 Proof of the mammogram receipt equation

Mammogram receipt D is given by

$$\begin{aligned}
D &= 1\{0 \leq V_T - V_U\} \\
&= 1\{0 \leq \mu_T(z,X) + \nu_T - \mu_U(z,X) - \nu_U\} \\
&= 1\{\nu_U - \nu_T \leq \mu_T(z,X) - \mu_U(z,X)\} \\
&= 1\{F(\nu_U - \nu_T) \leq F(\mu_T(z,X) - \mu_U(z,X))\} && (F \text{ increasing under A.2}) \\
&= 1\{U_D \leq F(\mu_T(z,X) - \mu_U(z,X))\} && (U_D = F(\nu_U - \nu_T \mid X) \text{ by definition}) \\
&= 1\{U_D \leq \mathrm{P}(D = 1 \mid Z = z, X)\},
\end{aligned}$$

where the last equality follows from

$$\begin{aligned}
F(\mu_T(z,X) - \mu_U(z,X) \mid X) &= \mathrm{P}(\nu_U - \nu_T \leq \mu_T(z,X) - \mu_U(z,X) \mid X) \\
&= \mathrm{P}(\nu_U - \nu_T \leq \mu_T(Z,X) - \mu_U(Z,X) \mid Z = z, X) \\
&\qquad ((\nu_U - \nu_T) \perp Z \mid X \text{ by A.1}) \\
&= \mathrm{P}(0 \leq \mu_T(Z,X) + \nu_T - \mu_U(Z,X) - \nu_U \mid Z = z, X) \\
&= \mathrm{P}(0 \leq V_T - V_U \mid Z = z, X) \\
&= \mathrm{P}(D = 1 \mid Z = z, X). \qquad \blacksquare
\end{aligned}$$

A.2 Proof that U_D is uniformly distributed between 0 and 1

The uniform distribution of U_D between 0 and 1 is not a separate assumption of the model, but due to the "probability integral transformation". It shows that the cumulative distribution function of any random variable $\widetilde{\nu} = \nu_T - \nu_U$ applied to itself must be distributed uniformly between 0 and 1 (for example, see Casella and Berger (2002, page 54)):

A random variable Y is distributed uniformly between 0 and 1 if and only if $F_Y(x) = x$ for $0 \leq x \leq 1$. Therefore, U_D is distributed uniformly between 0 and 1:

$$\begin{aligned}
F_{U_D}(u) &= P(U_D \leq u) \\
&= P(F(\nu_U - \nu_T) \leq u) \\
&= P(\nu_U - \nu_T \leq F^{-1}(u)) && (F \text{ increasing under A.2}) \\
&= F(F^{-1}(u)) = u. && (F \text{ continuous under A.2})
\end{aligned}$$

$\blacksquare$

A.3 Derivation of average outcomes

Suppressing the covariate vector X for simplicity, the expected value of Y_T for control group participants receiving mammograms, the always takers, is:

$$\begin{aligned}
\mathrm{E}[Y \mid D=1, Z=0] &= \mathrm{E}[Y_U + D(Y_T - Y_U) \mid D=1, Z=0] && \text{(by (7))} \\
&= \mathrm{E}[Y_T \mid D=1, Z=0] \\
&= \mathrm{E}[Y_T \mid 0 \leq U_D \leq p_C, Z=0] \quad \text{(by (6), where } p_C = P(D=1|Z=0)) \\
&= \mathrm{E}[g_T(U_D, \gamma_T) \mid 0 \leq U_D \leq p_C, Z=0] && \text{(by (8))} \\
&= \mathrm{E}[g_T(U_D, \gamma_T) \mid 0 \leq U_D \leq p_C] && (Z \perp (U_D, \gamma_T) \text{ by (A.5)}) \\
&= \mathrm{E}[Y_T \mid 0 \leq U_D \leq p_C].
\end{aligned}$$

I report the value that I obtain in my main analysis sample of the CNBSS, 451 deaths per 10,000 women, over the relevant range for always takers ($0 \leq U_D \leq p_C$) in Figure A.1 in Appendix D. Following similar calculations, I obtain the expected value of Y_T for intervention group participants receiving mammograms $\mathrm{E}[Y_T \mid 0 \leq U_D \leq p_I] = \mathrm{E}[Y \mid D=1, Z=1]$. In the CNBSS, this value is equal to 422 deaths per 10,000 women. These two values allow me to calculate the expected value of Y_T for compliers receiving mammograms such that:

$$\mathrm{E}[Y_T \mid p_C < U_D \leq p_I] = \frac{p_I}{p_I - p_C}\mathrm{E}[Y_T \mid 0 \leq U_D \leq p_I] - \frac{p_C}{p_I - p_C}\mathrm{E}[Y_T \mid 0 \leq U_D \leq p_C],$$

In the CNBSS, this value is equal to 415 deaths per 10,000 women, as shown in Figure 2 and in lighter shading in Figure A.1.

Turning to average untreated outcomes, I calculate the expected value of Y_U for control group participants without mammograms $\mathrm{E}[Y_U \mid p_C < U_D \leq 1] = \mathrm{E}[Y \mid D=0, Z=0]$, which is equal to 463 deaths per 10,000 women. I also calculate the expected value of Y_U for intervention group patients without mammograms $\mathrm{E}[Y_U \mid p_I < U_D \leq 1] = \mathrm{E}[Y \mid D=0, Z=1]$. This yields the expected untreated outcome for never takers, which is equal to 990 deaths per 10,000 women. Using these two values, I calculate the expected value of Y_U for compliers without mammograms as follows:

$$\mathrm{E}[Y_U \mid p_C < U_D \leq p_I] = \frac{1-p_C}{p_I - p_C}\mathrm{E}[Y_U \mid p_C < U_D \leq 1] - \frac{1-p_I}{p_I - p_C}\mathrm{E}[Y_U \mid p_I < U_D \leq 1].$$

In the CNBSS, this value is equal to 428 deaths per 10,000 women, as shown in as shown in Figure 2 and in lighter shading in Figure A.1.

A.4 Derivation of selection and treatment effect definitions from the literature

Definitions of selection and treatment effect heterogeneity from the econometrics literature, notably used by Angrist (1998) and Heckman et al. (1998), are as follows:

$$\text{Selection Heterogeneity:} \qquad \mathrm{E}[Y_U \mid D = 1] - \mathrm{E}[Y_U \mid D = 0] \tag{15}$$

$$\text{Treatment Effect Heterogeneity:} \qquad \mathrm{E}[Y_T - Y_U \mid D = 1] - \mathrm{E}[Y_T - Y_U \mid D = 0]. \tag{16}$$

These definitions can be obtained as weighted integrals of the MUO and MTE functions. For example, I can express (15) as a weighted integral of the MUO function:

$$\begin{aligned} = \int_0^1 \Big[& \mathrm{P}(Z = 0 \mid D = 1)\, \omega(p, 0, p_c) + \mathrm{P}(Z = 1 \mid D = 1)\, \omega(p, 0, p_I) \\ & - \mathrm{P}(Z = 0 \mid D = 0)\, \omega(p, p_c, 1) - \mathrm{P}(Z = 1 \mid D = 0)\, \omega(p, p_I, 1) \Big] \, \mathrm{MUO}(p)\, \mathrm{d}p \end{aligned} \tag{17}$$

with weights $\omega(p, p_L, p_H) = 1\{p_L \leq p < p_H\}/(p_H - p_L)$. As shown, this weighted integral depends on the probability of assignment to the intervention group, which is a feature of the experimental design. Under this definition, selection heterogeneity is not identified without further assumptions because the average untreated outcome for always takers is not observed.

A.5 Derivation of average characteristics for compliers

While average *outcomes* for compliers should depend on random assignment, average characteristics at baseline should not. Similarly to Imbens and Rubin (1997), Katz et al. (2001), and Abadie (2002, 2003), I obtain average characteristics at baseline for compliers by weighting average characteristics for compliers assigned to the control and intervention group by their respective probabilities such that:

$$\mathrm{E}[X \mid p_C < U_D \leq p_I] = \mathrm{P}(Z = 1)\mathrm{E}[X \mid p_C < U_D \leq p_I] + \mathrm{P}(Z = 0)\mathrm{E}[X \mid p_C < U_D \leq p_I].$$

B Exact Replication of Miller et al. (2014)

Table A.1: Results Up to 2005 Calendar Year for Participants Aged 40-59 at Enrollment

	(1) Intervention	(2) Control	(3) Relative Risk Ratio (1)/(2)	(4) Reduced Form (1)-(2)
Breast Cancer Deaths (per 10,000)	111	112	0.99 (0.06)	-1 (7)
All-Cause Deaths (per 10,000)	1,066	1,044	1.02 (0.02)	22 (20)
N	44,925	44,910		

Note. Bootstrapped standard errors in parentheses. Women aged 40-59 at enrollment are included.

C Robustness of Replication on Main Analysis Sample

Table A.2: Results for Participants with No Nonzero Breast-Related Covariates at Baseline 20 Years after Enrollment

	(1) Intervention	(2) Control	(3) Relative Risk Ratio (1)/(2)	(4) Reduced Form (1)-(2)
Age Group 40-49 at Enrollment (Main Analysis Sample)				
Breast Cancer Deaths (per 10,000)	57	66	0.87	-9
			(0.15)	(10)
All-Cause Deaths (per 10,000)	451	461	0.98	-10
			(0.06)	(29)
N	9,806	9,699		
Age Group 50-59 at Enrollment				
Breast Cancer Deaths (per 10,000)	76	79	0.96	-3
			(0.18)	(14)
All-Cause Deaths (per 10,000)	1,221	1,166	1.05	55
			(0.04)	(49)
N	8,521	8,689		
Age Group 40-59 at Enrollment				
Breast Cancer Deaths (per 10,000)	66	72	0.91	-6
			(0.11)	(9)
All-Cause Deaths (per 10,000)	809	794	1.02	15
			(0.04)	(28)
N	18,327	18,388		

Note. Bootstrapped standard errors in parentheses. Participants with any nonzero values of the following breast-related covariates at baseline are excluded: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer.

Table A.3: Results for Participants with Any Nonzero Breast-Related Covariates at Baseline
20 Years after Enrollment

	(1) Intervention	(2) Control	(3) Relative Risk Ratio (1)/(2)	(4) Reduced Form (1)-(2)
Age Group 40-49 at Enrollment				
Breast Cancer Deaths (per 10,000)	101	92	1.09	8
			(0.12)	(11)
All-Cause Deaths (per 10,000)	519	502	1.03	17
			(0.05)	(26)
N	15,408	15,517		
Age Group 50-59 at Enrollment				
Breast Cancer Deaths (per 10,000)	116	126	0.92	-10
			(0.11)	(15)
All-Cause Deaths (per 10,000)	1,160	1,181	0.98	-21
			(0.04)	(44)
N	11,190	11,005		
Age Group 40-59 at Enrollment				
Breast Cancer Deaths (per 10,000)	107	106	1.01	1
			(0.08)	(8)
All-Cause Deaths (per 10,000)	789	784	1.01	5
			(0.03)	(24)
N	26,598	26,522		

Note. Bootstrapped standard errors in parentheses. Only participants with any nonzero values of the following breast-related covariates at baseline are included: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer.

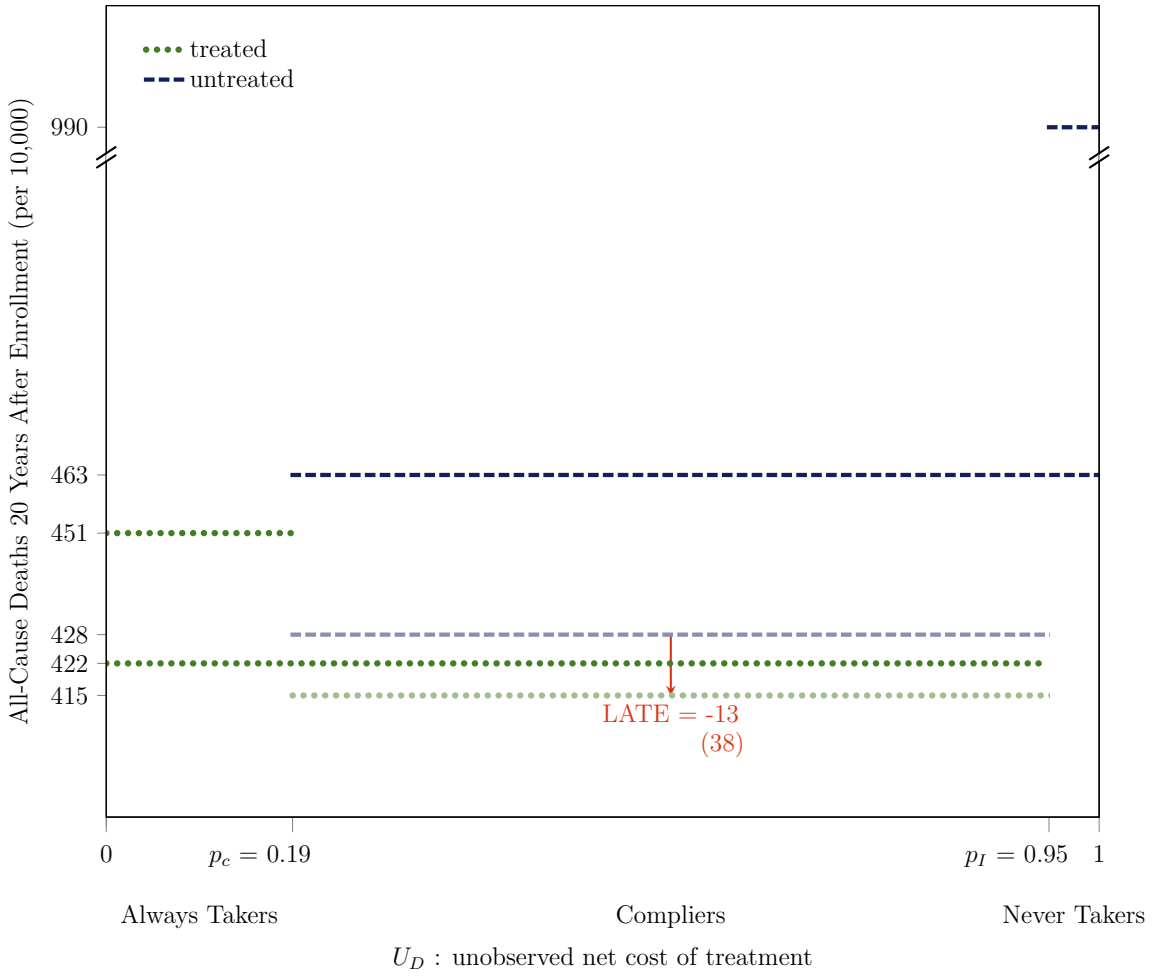
Table A.4: Results for All Participants
20 Years after Enrollment

	(1) Intervention	(2) Control	(3) Relative Risk Ratio (1)/(2)	(4) Reduced Form (1)-(2)
Age Group 40-49 at Enrollment				
Breast Cancer Deaths (per 10,000)	84	82	1.02 (0.10)	2 (8)
All-Cause Deaths (per 10,000)	493	486	1.01 (0.04)	6 (19)
N	25,214	25,216		
Age Group 50-59 at Enrollment				
Breast Cancer Deaths (per 10,000)	99	106	0.94 (0.08)	-7 (9)
All-Cause Deaths (per 10,000)	1,186	1,174	1.01 (0.03)	12 (35)
N	19,711	19,694		
Age Group 40-59 at Enrollment				
Breast Cancer Deaths (per 10,000)	90	92	0.98 (0.07)	-2 (7)
All-Cause Deaths (per 10,000)	797	788	1.01 (0.02)	9 (17)
N	44,925	44,910		

Note. Bootstrapped standard errors in parentheses. All participants are included. Some differences between statistics might not appear internally consistent because of rounding.

D Graphical Derivation of Average Outcomes

Figure A.1: Observed Mortality Averages



Note. Bootstrapped standard errors in parentheses. Calculated averages for compliers are reported in shaded colors. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer.

E Alternative Ancillary Weak Monotonicity Assumptions

Following Brinch et al. (2017) and Kowalski (2016, 2018b), I impose the following ancillary assumption on the MUO function:

M.1. (Weak Monotonicity of the MUO function) For all p_1, $p_2 \in [0,1]$ such that $p_1 < p_2$: $\mathrm{E}[Y_U \mid U_D = p_1] \leq \mathrm{E}[Y_U \mid U_D = p_2]$ or $[Y_U \mid U_D = p_1] \geq \mathrm{E}[Y_U \mid U_D = p_2]$.

Brinch et al. (2017) and Kowalski (2016, 2018b) also impose the following analogous assumption on the MTO function:

M.2. (Weak Monotonicity of the MTO function) For all p_1, $p_2 \in [0,1]$ such that $p_1 < p_2$: $\mathrm{E}[Y_T \mid U_D = p_1] \leq \mathrm{E}[Y_T \mid U_D = p_2]$ or $[Y_T \mid U_D = p_1] \geq \mathrm{E}[Y_T \mid U_D = p_2]$.

I could also impose the following analogous assumption on the MTE function, which is similar to the Manski (1997) assumption of monotone treatment response:

M.3. (Weak Monotonicity of the MTE function) For all p_1, $p_2 \in [0,1]$ such that $p_1 < p_2$: $\mathrm{E}[Y_T - Y_U \mid U_D = p_1] \leq \mathrm{E}[Y_T - Y_U \mid U_D = p_2]$ or $[Y_T - Y_U \mid U_D = p_1] \geq \mathrm{E}[Y_T - Y_U \mid U_D = p_2]$.

As I show in Kowalski (2016, 2018b), just as Assumption M.1 implies an upper or lower bound on the average treatment effect for always takers, Assumption M.2 implies an upper or lower bound on the average treatment effect for never takers. Either assumption is sufficient to test treatment effect homogeneity, and in practice both assumptions always yield the same result.

In the CNBSS, I impose Assumption M.1, but I do not impose Assumption M.2 because it is harder to defend. Assumption M.1 is an assumption on untreated outcomes. In the CNBSS, baseline covariates proxy for untreated outcomes. As I discuss in Section 4.2.1, examination of baseline covariates provides support for Assumption M.1. In contrast, Assumption M.2 is an assumption on treated outcomes. Treated outcomes reflect selection and treatment effects, so it is harder to proxy for them with baseline covariates available in the CNBSS, making Assumption M.2 harder to defend.

Rather than making an assumption on the combined influence of selection and treatment effects via Assumption M.2, it is more transparent to make an assumption on treatment effects directly via Assumption M.3. However, the imposition of Assumption M.3 alone directly *assumes* treatment effect heterogeneity, so it does not facilitate a useful empirical test of treatment effect homogeneity. Furthermore, the imposition of Assumption M.3 alone does not inform the direction of the treatment effect heterogeneity based on empirical quantities.

It could be productive to impose Assumption M.3 in conjunction with Assumption M.1. In the CNBSS, under Assumption M.1, the average treatment effect for always takers is

larger than the LATE, as shown in Figure 4. Therefore, the additional imposition of Assumption M.3 implies that the average treatment effect for never takers is weakly smaller than the LATE. The LATE is positive but not statistically different from zero, therefore providing suggestive evidence that the net harm from mammograms is positive for compliers. Under Assumptions M.2 and M.3, the average net harm from mammograms could be positive or negative for never takers. Because this result is effectively inconclusive in the CNBSS, it does not seem worthwhile to further impose Assumption M.3 to obtain it.

F Baseline Summary Statistics for Alternative Samples

Table A.5: Baseline Summary Statistics for Participants With Any Nonzero Breast-Related Covariates

	Means			Difference in Means	
	(1) Always Takers	(2) Compliers	(3) Never Takers	(1)-(2)	(2)-(3)
Baseline Socioeconomic Status					
University, trade or business school	0.51	0.45	0.36	0.06	0.09
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
In work force	0.65	0.63	0.56	0.02	0.07
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Age at first birth	24.11	24.02	23.47	0.09	0.55
	(0.07)	(0.04)	(0.19)	(0.10)	(0.20)
No live birth	0.17	0.15	0.13	0.02	0.02
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Married	0.80	0.81	0.74	-0.01	0.07
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Husband in work force / alive	0.81	0.81	0.74	0.00	0.07
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Baseline Health Behavior					
Non-Smoker	0.75	0.73	0.58	0.02	0.16
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Body Mass Index	23.72	24.19	24.11	-0.47	0.08
	(0.06)	(0.04)	(0.21)	(0.08)	(0.22)
Used oral contraception	0.72	0.70	0.66	0.01	0.05
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Used estrogen	0.14	0.15	0.16	-0.01	-0.01
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Mammograms prior to enrollment	0.45	0.30	0.29	0.15	0.01
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Practiced breast self examination	0.59	0.52	0.45	0.07	0.08
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Baseline Breast-Related Covariates					
Breast cancer in family	0.52	0.50	0.43	0.02	0.08
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Any other breast disease	0.35	0.23	0.24	0.12	-0.00
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Patient reported symptoms	0.41	0.37	0.46	0.04	-0.10
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Referred for review by nurse	0.28	0.21	0.32	0.07	-0.11
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Abnormality found by nurse	0.41	0.37	0.44	0.04	-0.06
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Ever told has breast cancer (%)	.	0.22	.	-0.10	-0.41
		(0.04)		(0.07)	(0.33)

Note. Bootstrapped standard errors in parentheses. Missing values correspond to redacted numbers in accordance with Data Use Agreement. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. Only women aged 40-49 with any nonzero values of the following breast-related covariates at baseline are included: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.

Table A.6: Baseline Summary Statistics for All Participants

	Means			Difference in Means	
	(1) Always Takers	(2) Compliers	(3) Never Takers	(1)-(2)	(2)-(3)
Baseline Socioeconomic Status					
University, trade or business school	0.51	0.45	0.37	0.06	0.08
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
In work force	0.65	0.63	0.60	0.02	0.03
	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
Age at first birth	24.16	24.00	23.51	0.15	0.49
	(0.06)	(0.03)	(0.13)	(0.08)	(0.15)
No live birth	0.16	0.15	0.13	0.02	0.01
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Married	0.80	0.81	0.74	-0.01	0.07
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Husband in work force / alive	0.81	0.81	0.75	-0.00	0.06
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Baseline Health Behavior					
Non-Smoker	0.76	0.74	0.60	0.02	0.14
	(0.00)	(0.00)	(0.02)	(0.01)	(0.02)
Body Mass Index	23.76	24.29	24.27	-0.53	0.02
	(0.05)	(0.03)	(0.14)	(0.07)	(0.15)
Used oral contraception	0.72	0.71	0.66	0.02	0.04
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Used estrogen	0.14	0.14	0.15	-0.00	-0.01
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Mammograms prior to enrollment	0.39	0.23	0.22	0.16	0.01
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Practiced breast self examination	0.56	0.49	0.42	0.07	0.07
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)
Baseline Breast-Related Covariates					
Breast cancer in family	0.38	0.29	0.24	0.09	0.05
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Any other breast disease	0.26	0.13	0.13	0.12	0.00
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Patient reported symptoms	0.30	0.21	0.26	0.09	-0.05
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Referred for review by nurse	0.20	0.12	0.18	0.09	-0.06
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Abnormality found by nurse	0.30	0.21	0.25	0.09	-0.03
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Ever told has breast cancer (%)	.	0.12	.	-0.04	-0.23
		(0.02)		(0.05)	(0.18)

Note. Bootstrapped standard errors in parentheses. Missing values correspond to redacted numbers in accordance with Data Use Agreement. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. All women aged 40-49 are included. Some differences between statistics might not appear internally consistent because of rounding.

G Full Results: Alternative Samples and Outcomes

Table A.7: Full Results: Alternative Samples and Outcomes

					Treated		Untreated			Treatment Effect	
					(1)	(2)	(3)	(4)	(3)-(4)	(1)-(3)	(5)
	N	p_C	p_I	Local Average Treatment Effect LATE	Always Takers	Compliers	Compliers	Never Takers	Untreated Outcome Test	Always Takers Lower Bound	Test Rejects Treatment effect Homogeneity
Main Specification											
Main specification	19,505	0.19	0.95	-13	451	415	428	990	-562	22	1.00
				(38)	(49)	(26)	(29)	(135)	(147)	(59)	(0.48)
Alternative Subsample Based on Breast-Related Covariates											
Any nonzero breast-related covariates	30,925	0.32	0.96	27	511	479	452	1,211	-759	60	1.00
				(40)	(30)	(33)	(24)	(123)	(135)	(39)	(0.47)
Both zero and nonzero breast-related covariates	50,430	0.27	0.96	9	495	451	442	1,114	-672	53	1.00
				(27)	(25)	(21)	(18)	(95)	(103)	(31)	(0.34)
Alternative Outcomes											
Breast cancer mortality	19,505	0.19	0.95	-12	88	47	58	.	-43	30	1.00
				(13)	(23)	(10)	(8)		(47)	(25)	(0.43)
Breast cancer incidence	19,505	0.19	0.95	58	571	424	366	667	-301	206	1.00
				(34)	(57)	(30)	(25)	(110)	(119)	(65)	(0.17)
Alternative Sample Based on Age Group											
Participants Aged 50-59 at Enrollment	17,210	0.14	0.95	67	982	1,198	1,131	2,271	-1,140	-149	0.00
				(60)	(94)	(44)	(42)	(210)	(226)	(103)	(0.20)

Note. Bootstrapped standard errors in parentheses. Missing values correspond to redacted numbers in accordance with Data Use Agreement. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. Baseline breast-related covariates refer to the following: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.

H Full Results by Enrollment Year

Table A.8: Full Results by Calendar Year of Enrollment

					Treated		Untreated			Treatment Effect	
					(1)	(2)	(3)	(4)	(3)-(4)	(1)-(3)	(5)
Enrollment Year	N	p_C	p_I	Local Average Treatment Effect LATE	Always Takers	Compliers	Compliers	Never Takers	Untreated Outcome Test	Always Takers Lower Bound*	Test Rejects Treatment Effect Homogeneity
1980	1,110	0.15	0.89	-39	.	418	458	.	-362	152	1.00
				(188)		(130)	(134)		(431)	(315)	(0.47)
1981	2,535	0.15	0.93	-98	.	425	523	.	-500	9	1.00
				(113)		(82)	(79)		(346)	(167)	(0.46)
1982	3,747	0.18	0.94	-20	448	370	390	.	-464	58	1.00
				(83)	(112)	(64)	(59)		(297)	(131)	(0.48)
1983	5,143	0.19	0.95	-69	445	432	501	.	-292	-56	1.00
				(79)	(90)	(56)	(54)		(267)	(106)	(0.50)
1984	6,238	0.20	0.97	73	442	417	344	1,579	-1,235	98	1.00
				(65)	(82)	(53)	(42)	(347)	(364)	(93)	(0.50)
1985	732	0.25	0.98	17	.	472	.	.	456	-228	1.00
				(192)		(137)			(127)	(211)	(0.32)

Note. Bootstrapped standard errors in parentheses. Missing values correspond to redacted numbers in accordance with Data Use Agreement. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. *In the reported specification in which the untreated outcome test statistic is positive, this column gives an *upper* bound on the average treatment effect for always takers. Some differences between statistics might not appear internally consistent because of rounding.

I Full Results by CNBSS Center

Table A.9: Full Results by CNBSS Center

					Treated		Untreated			Treatment Effect	
					(1)	(2)	(3)	(4)	(3)-(4)	(1)-(3)	(5)
Center	N	p_C	p_I	Local Average Treatment Effect LATE	Always Takers	Compliers	Compliers	Never Takers	Untreated Outcome Test	Always Takers Lower Bound*	Test Rejects Treatment Effect Homogeneity
1	4,060	0.17	0.92	-39	.	397	436	881	-444	-90	0.00
				(90)		(58)	(67)	(216)	(253)	(126)	(0.48)
2	1,003	0.11	0.96	-207	.	212	419	.	-1,009	713	1.00
				(154)		(103)	(108)		(747)	(451)	(0.32)
3	1,727	0.23	0.91	-140	.	423	563	.	-323	-104	1.00
				(143)		(99)	(111)		(405)	(199)	(0.50)
4	947	0.11	0.93	18	.	421	403	.	-168	-3	0.00
				(157)		(113)	(111)		(397)	(300)	(0.50)
5	2,739	0.15	0.95	-201	800	338	539	.	-127	261	1.00
				(108)	(203)	(79)	(74)		(289)	(217)	(0.48)
6	2,444	0.28	0.98	242	.	490	248	.	-2,360	52	0.00
				(105)		(85)	(61)		(866)	(108)	(0.27)
7	772	0.18	0.96	85	.	410	.	.	-300	103	1.00
				(185)		(139)			(647)	(293)	(0.50)
8	546	0.25	0.96	75	.	497	.	.	-487	324	1.00
				(278)		(230)			(820)	(356)	(0.50)
9	1,029	0.25	0.96	-69	.	297	367	.	-938	21	1.00
				(164)		(124)	(118)		(732)	(217)	(0.49)
10	1,388	0.16	0.97	76	.	510	434	.	-1,740	112	1.00
				(159)		(115)	(109)		(948)	(247)	(0.50)
11	1,092	0.10	0.96	204	.	620	416	.	-84	-52	0.00
				(165)		(127)	(101)		(513)	(270)	(0.50)
12	769	0.35	1.00	-84	.	356	.	-	-	-	-
				(198)		(178)					
13	366	0.19	0.95	-97	.	.	.	.	-448	-377	0.00
				(321)					(1,258)	(365)	(0.50)
14	279	0.12	1.00	238	.	.	.	-	-	-	-
				(304)							
15	344	0.15	0.99	-40	.	.	.	.	367	-367	1.00
				(238)					(167)	(170)	(0.14)

Note. Bootstrapped standard errors in parentheses. Missing values correspond to redacted numbers in accordance with Data Use Agreement. Dashes represent statistics that are not applicable because the given subsample does not have any never takers. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. *In the reported specification in which the untreated outcome test statistic is positive, this column gives an *upper* bound on the average treatment effect for always takers. Some differences between statistics might not appear internally consistent because of rounding. Some differences between statistics might not appear internally consistent because of rounding.

J Full Results: Alternative Definitions of Mammogram Receipt

Table A.10: Full Results for Alternative Definitions of Mammogram Receipt

					Treated		Untreated			Treatment Effect	
				Local Average Treatment Effect	(1)	(2)	(3)	(4)	(3)-(4)	(1)-(3)	(5)
	N	p_C	p_I	LATE	Always Takers	Compliers	Compliers	Never Takers	Untreated Outcome Test	Always Takers Lower Bound	Test Rejects Treatment effect Homogeneity
Main Specification											
Mammogram in at least one year after enrollment during the active study period, missing in year= no mammogram in year											
Main specification	19,505	0.19	0.95	-13	451	415	428	990	-562	22	1.00
				(38)	(49)	(26)	(29)	(135)	(147)	(59)	(0.48)
Narrower Definitions of Mammogram Receipt											
Mammogram in more than one year after enrollment during the active study period, missing in year = no mammogram in year											
At least two active study period years	19,505	0.08	0.91	-12	392	407	420	884	-465	-27	0.00
				(35)	(72)	(21)	(28)	(93)	(106)	(77)	(0.49)
At least three active study period years	19,505	0.03	0.84	-12	.	379	391	811	-420	56	1.00
				(36)		(21)	(31)	(75)	(94)	(145)	(0.48)
All active study period years	19,505	0.02	0.70	-15	.	382	396	622	-225	-135	0.00
				(42)		(21)	(38)	(47)	(75)	(138)	(0.37)
Broader Definitions of Mammogram Receipt											
Mammogram in at least one year after enrollment during the active study period											
Missing in year = mammogram in year	19,505	0.58	1.00	-24	503	376	400	.	-776	103	1.00
				(69)	(30)	(63)	(32)		(835)	(43)	(0.43)

Note. Bootstrapped standard errors in parentheses. All-cause deaths are measured 20 years after enrollment for all participants, based on the exact calendar date of enrollment. The treatment is mammogram receipt. In the main specification, mammogram receipt is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.

K Full Results: Mortality at Alternative Follow-Up Lengths

Table A.11: Full Results for Mortality at Alternative Follow-Up Lengths

					Treated		Untreated			Treatment Effect	
					(1)	(2)	(3)	(4)	(3)-(4)	(1)-(3)	(5)
Years Since Enrollment	N	p_C	p_I	Local Average Treatment Effect LATE	Always Takers	Compliers	Compliers	Never Takers	Untreated Outcome Test	Always Takers Lower Bound	Test Rejects Treatment Effect Homogeneity
Main specification: 20	19,505	0.19	0.95	-13	451	415	428	990	-562	22	1.00
				(38)	(49)	(26)	(29)	(135)	(147)	(59)	(0.48)
19	19,505	0.19	0.95	-13	434	371	384	869	-485	50	1.00
				(37)	(48)	(24)	(28)	(131)	(142)	(58)	(0.40)
18	19,505	0.19	0.95	-8	390	328	336	828	-492	54	1.00
				(35)	(48)	(24)	(26)	(128)	(139)	(56)	(0.41)
17	19,505	0.19	0.95	-8	330	304	312	768	-456	18	1.00
				(33)	(43)	(23)	(25)	(126)	(135)	(50)	(0.48)
16	19,505	0.19	0.95	-16	291	260	277	747	-471	15	1.00
				(31)	(40)	(22)	(23)	(125)	(134)	(46)	(0.47)
15	19,505	0.19	0.95	-15	236	232	247	727	-480	-11	1.00
				(31)	(36)	(21)	(23)	(121)	(131)	(42)	(0.50)
14	19,505	0.19	0.95	-21	192	210	231	626	-396	-38	0.00
				(30)	(33)	(20)	(22)	(112)	(121)	(38)	(0.45)
13	19,505	0.19	0.95	-24	170	176	201	566	-365	-30	0.00
				(28)	(31)	(19)	(20)	(106)	(115)	(36)	(0.49)
12	19,505	0.19	0.95	-27	148	144	171	505	-334	-23	1.00
				(27)	(28)	(18)	(19)	(98)	(106)	(32)	(0.50)
11	19,505	0.19	0.95	-10	104	124	134	485	-351	-30	0.00
				(25)	(23)	(16)	(17)	(98)	(105)	(28)	(0.42)
10	19,505	0.19	0.95	-15	77	103	118	424	-306	-41	0.00
				(23)	(21)	(14)	(16)	(91)	(97)	(25)	(0.37)
9	19,505	0.19	0.95	-12	.	77	90	404	-314	-35	0.00
				(20)		(12)	(15)	(90)	(97)	(21)	(0.36)
8	19,505	0.19	0.95	-2	.	61	64	404	-340	-14	0.00
				(18)		(11)	(14)	(90)	(97)	(21)	(0.44)
7	19,505	0.19	0.95	-6	.	47	53	404	-351	-15	0.00
				(17)		(10)	(13)	(90)	(97)	(18)	(0.46)
6	19,505	0.19	0.95	-5	.	41	46	364	-317	-24	0.00
				(15)		(9)	(12)	(86)	(93)	(16)	(0.33)
5	19,505	0.19	0.95	-5	.	29	34	303	-269	-12	0.00
				(13)		(7)	(11)	(80)	(86)	(15)	(0.45)
4	19,505	0.19	0.95	-9	.	16	25	.	-218	-3	1.00
				(11)		(6)	(9)		(77)	(14)	(0.49)
3	19,505	0.19	0.95	-6	.	.	14	.	-209	-3	1.00
				(9)			(8)		(76)	(11)	(0.50)
2	19,505	0.19	0.95	-3	.	.	8	.	-194	-3	1.00
				(9)			(8)		(67)	(9)	(0.50)
1	19,505	0.19	0.95	-5	.	.	.	.	-55	-5	0.00
				(5)					(40)	(5)	(0.00)

Note. Bootstrapped standard errors in parentheses. The treatment is mammogram receipt, which is equal to one if a participant receives a mammogram in at least one year after enrollment during the active study period. Missing mammogram data in any year is set to no mammogram in that year. The sample includes women aged 40-49 at enrollment, excluding women with any nonzero values of the following breast-related covariates at baseline: breast cancer in family; any other breast disease; patient reported symptoms; referred for review by nurse; abnormality found by nurse; ever told has breast cancer. Some differences between statistics might not appear internally consistent because of rounding.