ONLINE APPENDIX
Screening and Selection: The Case of Mammograms

by Einav, Finkelstein, Oostrom, Ostriker, and Williams

A Coding mammograms and outcomes in claims data

We follow Segel et al. (2017) in coding the incidence of screening mammograms (hereafter "mammograms") and the results of those mammograms in the HCCI claims data.

We code a woman as having a screening mammogram on a given date if she has a claim with ICD-9 procedure code V76.12 or CPT codes 77057 or G0202 on that date, but no claims for any other mammogram within the previous 12 months and no prior claims for breast cancer treatment.[1] Previous work has documented that claims-based measures of mammogram rates tend to be lower than mammogram rates in self-reported survey data. For example, Freeman et al. (2002) document this pattern in Medicare data, and Cronin et al. (2009) document similar evidence in a study of Vermont women. Consistent with these studies, Appendix Figure A.1 documents the age profile of the annual screening mammogram rate, as measured by both the Behavioral Risk Factor Surveillance System (BRFSS) survey and the algorithm described above using the HCCI claims data. Between ages 39 and 41, the mammogram rate jumps by approximately the same amount—25 percentage points—by both measures, but the survey data describe mammogram rates as being approximately 10 to 20 percentage points higher than the claims data rate at all ages. Of course, the samples are not perfectly comparable, as the BRFSS sample is of all women with health insurance (public or private) from 2002-2012, while our final HCCI sample consists of women privately insured by Aetna, Humana or United from 2009-2011.

We code the outcome of a screening mammogram as negative if there are no subsequent claims for either follow-up testing or breast cancer treatment within the next 12 months. We code the outcome as a false positive if there is at least one claim for follow-up testing in the following three months (i.e. a subsequent mammogram, a breast biopsy, a breast ultrasound, or other radiologic breast testing) in the following three months, but no claims for breast cancer treatment in the next 12 months. We code the outcome of a mammogram as true positive if, within 12 months following the mammogram, there is at least one claim for breast cancer. We consider a woman to have a subsequent mammogram if she has a claim with ICD-9 procedure code V76.12 or CPT codes 77057 or G0202. A woman has a breast biopsy if she has a claim with ICD-9 procedure code 85.11, 85.12, 85.20, or 85.21 or CPT codes 19100, 19101, or 19120. A breast ultrasound is coded with ICD-9 procedure code 88.73 or CPT code 76645. Radiologic breast testing is

[1] Segel et al. (2017) focused on data from 2003-2004, so used the CPT code 76092. In 2007 this code was replaced by 77057. In addition, Hubbard et al. (2015) identify CPT code G0202 as indicating a screening mammogram claim. Segel et al. (2017) provide codes for "other" (non-screening) mammograms, which we omit.

coded with ICD-9 procedure code 87.35, 87.36, 87.73, or 88.85 or CPT codes 76003, 77002, 76095, 77031, 76086, 76087, 76088, 77053, 77054, 76355, 76360, 76362, 77011, 77012, 77013, 76098, 76100, 76101, 76102, 76120, 76125, 76140, 76150, 76350, or 76365. Breast cancer is coded with ICD-9 procedure code 233.0, V103.0, or 174.0 through 174.9 or CPT code 19160, 19162, 19180, 19200, 19220, 19240, 19301, 19303, 19305, 19307, 38740, or 38745. The codes used to identify these claims are provided in Appendix Table A.1, along with their references.

The linked SEER-Medicare data allow us to cross validate this claims-based coding process against cancer diagnoses in the cancer registry. The results are very encouraging. Appendix Tables A.2 and A.3 describe the concordance of true positive mammograms as coded using this algorithm with actual diagnoses as recorded in the SEER-Medicare data. For those who were diagnosed with breast cancer and had a mammogram in the year of diagnosis, 99.6% of mammograms were coded as true positive using our algorithm. Meanwhile, 93% of mammograms for women who were never diagnosed with breast cancer were negative, while 5.8% were false positives. Most women with true positive mammograms were diagnosed with breast cancer in the year of or the year following the mammogram, while 84% of those without true positive mammograms were never diagnosed and a further 14% were not diagnosed until more than 1 year after the mammogram (2% were diagnosed in the year following the mammogram, but none in the year of the mammogram).

B Non-cancer characteristics of compliers, always-takers, and never-takers

We examine the non-cancer characteristics of those who do and do not receive mammograms, by age. To do so, we draw on two data sources: the HCCI data on privately insured women from 2009-2011, and data from the Behavioral Risk Factor Surveillance System Survey (BRFSS) for even years 2000-2012, restricted to women with health insurance (public or private). The HCCI data allow us to code whether or not the woman had a Pap test or a flu shot in the prior year, as well as total health care spending in the prior year and the number of emergency room (ER) visits in the prior year. The BRFSS data allow us to observe (contemporaneous) measures of flu shots and Pap tests, as well as other health behaviors. Specifically, we follow Chetty et al. (2016) and measure whether the woman currently smokes, has BMI greater than 30, and exercises in the past month, and we follow Kowalski (2019) and measure mean BMI, whether the woman uses birth control, and whether the woman uses oral contraception. We also include several additional health behaviors: whether the woman always wears a seat belt, her number of alcoholic drinks in the prior month, and whether she has a personal doctor. Finally, we measure some basic demographics of the women including income, education, employment, marital status, and whether she lives in an urban, suburban, or rural area.

Appendix Figures A.3, A.4, and A.5 show age-specific mean characteristics for the overall population and the subset of women who receive mammograms. The figures also super-impose results from a simple linear regression discontinuity estimate at age 40 for those who get mammograms. Specifically, we estimate

(A.1) $y_i = \gamma_0 + \gamma_1 \tilde{a}_i + \gamma_2 I(\tilde{a}_i > 0) + \gamma_3 \tilde{a}_i I(\tilde{a}_i > 0) + \gamma_4 m_i + \gamma_5 \tilde{a}_i m_i + \gamma_6 m_i I(\tilde{a}_i > 0) + \gamma_7 \tilde{a}_i m_i I(\tilde{a}_i > 0) + \varepsilon_i,$

where y_i is a health behavior or demographic characteristic for woman i, $\tilde{a}_i = a_i - 40$ is woman i's age in years relative to age 40, $I(\tilde{a}_i > 0)$ is an indicator for whether woman i's age is greater than 40, m_i is an indicator for whether woman i received a mammogram, and ε_i is a normally distributed error term. We estimate the regression on women aged 35-50, excluding women aged 40 from the regression. The regression allows for the outcome y_i to vary linearly in age, and for a level change and a slope change at age 40. It also allows these patterns to flexibly vary between those who receive a mammogram and those who do not.

We use the resulting estimates to implement the Abadie (2002; 2003) approach to characterizing compliers and never-takers. For example, women who obtained a mammogram without a recommendation provide estimates of characteristics for always-takers. Since women who obtain a mammogram with the recommendation at age 40 include both compliers and always-takers, we can use estimates of the always-taker sample and the share of compliers to back out the characteristics of compliers. Similarly, women who do not obtain a mammogram after age 40 provide estimates of characteristics for never-takers. As in equation (2) in the main text, we assume that there are no defiers; that is, we assume that there are no women who would obtain a mammogram without the recommendation at age 39 but would not get screened with the recommendation at age 40.

More formally, we first calculate the population fraction of always-takers, never-takers, and compliers, by estimating the following regression

(A.2) $$m_i = \eta_0 + \eta_1 \tilde{a}_i + \eta_2 I(\tilde{a}_i > 0) + \eta_3 \tilde{a}_i I(\tilde{a}_i > 0) + \zeta_i,$$

where all variables are defined as in equation (A.1), and ζ_i is a normally distributed error term. The population fraction of always-takers at age 40 is then given by the fraction of women who receive a mammogram without the recommendation, $\pi_{AT} = \hat{\eta}_0$, and the population fraction of compliers is given by the change in the fraction of women who receive a mammogram at age 40, $\pi_C = \hat{\eta}_2$. Finally, given our assumptions, the population fraction of never-takers is given by $\pi_{NT} = 1 - \pi_C - \pi_{AT}$.

With the population shares of each group, we can now estimate the group-specific mean for each characteristic. Let $f_g(y)$ denote the mean for characteristic y at age 40 for each group $g \in \{AT, NT, C\}$, where AT are the always-takers, NT are the never-takers, and C are the compliers. We then compute $f_g(y)$ at age 40 using the above estimates. For example, for always-takers, we use the predicted values from equation (A.1) by setting $m_i = 1$, $\tilde{a}_i = 0$, $I(\tilde{a}_i > 0) = 0$ to obtain $f_{AT}(y) = \hat{\gamma}_0 + \hat{\gamma}_4$. Similarly, for never-takers, we set $m_i = 0$, $\tilde{a}_i = 0$, $I(\tilde{a}_i > 0) = 0$ to obtain$f_{NT}(y) = \hat{\gamma}_0$. Computing the mean of y for compliers, $f_C(y)$, requires some arithmetic. Women who receive a mammogram with the recommendation are either always-takers or compliers. Therefore, the mean of y for women who received a mammogram with a recommendation, denoted by $f_T(y)$, is equal to $\frac{\pi_{AT}}{\pi_{AT}+\pi_C} f_{AT}(y) + \frac{\pi_C}{\pi_{AT}+\pi_C} f_C(y)$. We can estimate $f_T(y)$ from equa-

tion (A.1) by setting $m_i = 1$, $\tilde{a}_i = 0$, $I(\tilde{a}_i > 0) = 1$ to obtain $f_{AT}(y) = \hat{\gamma}_0 + \hat{\gamma}_2 + \hat{\gamma}_4 + \hat{\gamma}_6$. We then compute $f_C(y) = [f_T(y)(\pi_{AT} + \pi_C) - \pi_{AT} f_{AT}(y)]/\pi_C$.

In Figure 4 in the main text we present the results from this exercise by reporting the ratios of characteristics for compliers relative to never-takers, $f_C(y)/f_{NT}(y)$, and for compliers relative to always-takers, $f_C(y)/f_{AT}(y)$. Standard errors are constructed using a bootstrap with 100 repetitions clustered at the age level.

C Clinical model: the Erasmus model

We use the Erasmus model to generate estimates of the underlying onset rate by age of cancer and cancer type, as well as the evolution of (untreated) cancers. We adjust the model to better match certain key moments of the SEER data. This (modified) Erasmus data, together with assumed parameters from the mammogram decision model (specifically, equations (1) and (2) in the main text) and the observed policy recommendation (40 and above), generates an age-specific share of women who are screened, as well as the tumor characteristics (in-situ and invasive rates) conditional on getting screened, which we then attempt to match by method of moments to the observed data on the age-specific share of women who are screened and the tumor characteristics conditional on getting screened.

As described in the main text, the Erasmus model is one of seven models developed for the Cancer Intervention and Surveillance Modeling Network (CISNET) as part of a project decomposing breast cancer mortality reductions from 1975-2000 into effects from the dissemination of mammography versus the development of advanced treatment techniques (Clarke et al. 2006). Each of the groups participating in the project wrote a model of breast cancer incidence and mortality in the US over this time period and then compared the mortality rates under scenarios with and without mammography and advanced treatment. For convenience, we focus on one of these models, the Erasmus model (Tan et al., 2006).

In what follows we describe our implementation of the Erasmus model. This implementation directly follows Tan et al. (2006), with all the assumptions we describe being theirs. We then describe the calibration changes we make to the model based on some of our own external data and assumptions.

C.1 Model details

Tumor incidence

The model allows us to simulate a cohort of women i, each with a year of birth b_i and a year of death from other causes d_i which is randomly determined and dependent on the year of birth. Specifically, it assumes that in each year y the probability that a person born in year b (such that $y \geq b$) dies of causes other than breast cancer is Q_y^b. A woman's year of death is defined as the lesser of 110 and the first year in which a random draw from a uniform distribution on [0,1] falls below Q_y^b. It assumes that no woman dies from other causes before age 30.

The model further assumes that there exists a probability C_b that any woman from cohort b will get cancer before age 85. It defines age $a_y^b = y - b$ as the age in year y of a woman born in year b and assumes

that for every cohort b and year y such that $20 \leq a_y^b \leq 85$ there exists S_a, the probability that a woman experiences tumor onset at age a conditional on eventually getting cancer. For each woman i with any cancer, we can therefore construct the year of tumor onset t_i as the lesser of the year in which she turns 85 and the first year in which a random draw from a uniform distribution on [0, 1] falls below S_{y-b_i}.

Tumor type and in-situ characteristics

At onset, cancer type is defined to be either an invasive tumor or one of three types of non-invasive tumors. Invasive tumors are assigned a minimum size and other tumor characteristics (as described in Appendix Table A.4) at onset and immediately begin growing. Non-invasive tumors are also known as ductal carcinoma in-situ (DCIS), which we refer to in the text as in-situ, and they can be one of three types: (a) DCIS-regressive tumors eventually disappear without causing any harm; (b) DCIS-invasive tumors eventually transform into a harmful invasive tumor but do no harm in the meantime; and (c) DCIS-clinical tumors do no harm but are eventually clinically detected. The model assumes that the outcome of each DCIS tumor (regression, invasion, or detection) occurs w_i years after onset, where w_i is generated by random draws from an exponential distribution with mean W. None of the three types of DCIS tumors can be clinically detected during the duration of this "dwell time," but they can be screen-detected with a screening-year-specific probability E_y if screening occurs. The type of tumor is defined at onset subject to age-specific probabilities I_a (invasive), V_a (DCIS-invasive), R_a (DCIS-regressive), and C_a (DCIS-clinical) such that $I_a + V_a + R_a + C_a = 1$. Values for these and other Erasmus parameters are given in Appendix Table A.5.

For DCIS tumors that become invasive, onset of invasive disease is defined as the moment when the tumor size reaches the minimum value of the screening threshold diameter; this threshold varies with the woman's age as well as over time (to reflect improvements in screening technology). The dwell time for DCIS tumors was calibrated in the MISCAN breast cancer model based on the duration from onset of DCIS to the 1975 screening threshold diameter.

Invasive tumor characteristics

The model assumes that the fundamental characteristic of invasive tumors is their year-dependent size s_i^y. For all invasive tumors, it defines s_i^0 (the size in the year of onset) to be equal to 0.01 cm. It is assumed that all invasive tumors grow exponentially. Tumor size in year y is therefore given by $s_i^0 (1+g_i)^y$ where g_i is the individual-specific growth rate (drawn from a lognormal distribution at tumor onset). It further assumes that diagnosis depends on tumor size and the individual's "screen detection diameter" r_i^{ay} (drawn at the time of screening from an age- and detection-year-specific Weibull distribution) and "clinical diagnosis diameter" c_i (log normally distributed and set at tumor onset). If the woman undergoes screening, the tumor can be detected if $s_i^y > r_i$. Alternatively, if the tumor grows so large that $s_i^y > c_i$, the woman will certainly detect it due to the appearance of clinical symptoms. Tumor size also determines mortality: if a woman diagnoses her tumor before it reaches its "fatal diameter" f_i (drawn at onset from a year-specific Weibull distribution), she will receive treatment and survive, but if not, she will die regardless of treatment.

The model defines for each invasive tumor the length of time the woman will survive after the tumor reaches its fatal diameter, called the "survival duration since fatal diameter" and denoted u_i (log normally

distributed). It assumes that if the tumor has not been clinically detected by the time 0.9*u_i years have passed since the fatal diameter was reached, it will be clinically detected due to distant metastases at that time.

Finally, it assumes that the growth rate g_i, clinical diagnosis diameter c_i, and survival duration u_i are correlated with coefficients ρ_{gc}, ρ_{gu}, and ρ_{cu}. The variables described in this section ($s_i^y, r_i^{ay}, f_i\ g_i, c_i, u_i$) , combined with the woman's age and the year of initiation, fully specify the course of the disease for an invasive tumor, subject to potential screening regimens.

C.2 Parameterizing the Erasmus model

We begin by choosing certain population-specific parameters required as inputs for the Erasmus model: the other-cause death probability, the overall tumor incidence, and the tumor incidence by age. As in Tan et al. (2006), the other-cause death probability follows the approach of Rosenberg (2006). However, we adjusted the tumor incidence parameters (overall cohort incidence and quadratic incidence by age) that are given in Tan et al. (2006) in order to match the SEER data's share of diagnoses that are in-situ and invasive for those under 40 and over 40. After establishing these population-specific parameters, we simulate individual life histories under a no-screening assumption, and use the tumor sizes and types to determine the population cancer rate by age.

Other-cause death probability

Following Rosenberg (2006), we computed probability of death due to other causes as the difference between the all-cause mortality and breast cancer specific mortality. We obtained all-cause mortality for ages 0-110 and years 1933-2010 from the Human Mortality Database. Using breast cancer death totals from the National Center for Health Statistics and female population totals from the Human Mortality Database, we calculated breast-cancer-specific mortality for ages 0-110 and years 1959-2010. To impute values for previous years, we assumed that the age-specific breast cancer mortality rate in any year before 1958 was equal to the rate in 1958. We combined these data to calculate non-breast-cancer mortality rates for all years between 1933 and 2010.

Age profile of cancer incidence

The Erasmus model provided a CDF of tumor incidence in 5-year increments, implying a step function of yearly incidence that produces spikes in tumor onset within a cohort every 5 years (see Appendix Table A.6, first column reproduced from Tan et al. (2006), based on estimates of US population in 1975). We constructed a smoothed CDF of tumor incidence by fitting to the Erasmus CDF using a constrained polynomial (quadratic) fit: $y = ax^2 + bx + c$. We fitted a, b, c, the start age x_{start} (at which the CDF should be zero), and the end age x_{end} (at which the CDF should be one). Restrictions included:

$$ax_{start}^2 + bx_{start} + c = 0$$

$$ax_{end}^2 + bx_{end} + c = 1$$

$$2ax_{start} + b \geq 0$$

The values that minimize the error $\sum(\hat{y} - y)^2$ across each of the fourteen ages in Appendix Table A.6 are $x_{start} = 24$, $x_{end} = 85$, $a = 0.000268$, $b = -0.01282$, $c = 0.15327$. We assume that the incidence before age 24 is 0. The fit is shown in Appendix Figure A.7.

Adjusting cancer incidence rates

Tan et al. (2006) calculate cumulative tumor incidence by birth cohort based on observed (i.e. diagnosed) incidence in the US from 1975-1979. Implicitly, this assumes that all tumors are diagnosed. It will therefore miss any undiagnosed tumors. Not surprisingly, therefore, when we use the original Erasmus parameters and our calibrated screening policy described below, the model substantially under-predicts observed diagnoses. To rectify this, we allowed the cohort tumor incidence to vary with a multiplicative shift α which uniformly affects each cohort's tumor incidence.

We calibrate α as follows. We define the parameters $\theta = (\alpha, pscrn_{inv}, pscrn_{dcis})$ where $pscrn_{inv}$ is the probability of a mammogram conditional on having an invasive tumor and $pscrn_{insitu}$ is the probability of a mammogram conditional on having an in-situ tumor. We then estimate θ by maximum likelihood. Specifically, we maximize the log likelihood of observing SEER tumor types (1973-2013, for women 25-34). The model's original incidence and the incidence multiplicatively shifted by α are plotted against the SEER diagnosis rates in Appendix Figure A.8. We also plot the model's diagnosis rates with no screening; with the multiplicative shift this roughly matches with the SEER diagnosis levels.

C.3 Visual representation and results from Erasmus model: underlying cancer rate

The first panel of Appendix Figure A.9 visualizes the Erasmus model using a flow chart. The second panel shows example sequences of progression for each of the four types of tumors, in the absence of screening. The first two rows show the progress of DCIS-regressive and DCIS-clinical tumors, which are harmless and differ only in their behaviors at the end of their dwell time: DCIS-regressive tumors disappear, while DCIS-clinical tumors are detected clinically, for example at a routine physical exam. If these tumors are screened, they will be diagnosed with a probability equal to the "sensitivity" as described in Appendix Table A.5. Likewise, before it switches to its invasive phase, the DCIS-invasive tumor can also be detected by a screening mammogram in the same way. After it becomes invasive, the DCIS-invasive tumor (row 3) and the invasive tumor (row 4) can only be detected if the size exceeds the year- and age-specific screening diameter of the year in which it is screened. If a woman's tumor is screened (or clinically diagnosed) before it reaches the fatal diameter, her life is saved, but if not, she will eventually die, regardless of detection or treatment in later years. In most cases, when a woman's tumor reaches the fatal diameter without being diagnosed, she will be clinically diagnosed before death. The flow chart omits deaths due to other causes.

Appendix Figure A.10 plots the share of women in each of five categories when the Erasmus model is calculated with no screening. The calculation is based on birth cohorts from 1950-1975, and focuses on women aged 30-50 in 2000-2005. At any given age, the share of women with detectable invasive or DCIS cancer is substantially smaller than the share of women who have already been diagnosed clinically, indicating that there is a small window of time during which a cancer can be screened before it is clinically detected.

Using the calibrated other-cause death probabilities and incidence rates, we solve the Erasmus model assuming that there is no screening for birth years 1950-1975. We restrict to years 2000-2005 and ages 30-50, producing a set of individual life-histories that can be categorized in every year as dead due to breast cancer, dead due to other causes, clinically diagnosed, currently undiagnosed invasive cancer, currently undiagnosed DCIS, or no cancer. (We consider invasive cancer that is too small to be detectable, and regressed DCIS tumors, to be the same as "no cancer.")

We take the "population cancer rate" at each age, or the share of women who have a tumor by a certain age, from the Erasmus model. The Erasmus model assumes that cancers can only be detected by mammogram once they have reached a certain size, so we assume the screening diameter is 1.09 cm—the average screen-detectable size in the Erasmus model—and count the share of women with detectable invasive cancer as the share of women with tumors above that size in the Erasmus model. We also count 80% of the women with DCIS tumors, under the assumption in the Erasmus model that about 80% of technically "detectable" in-situ tumors will be detected in any given year. We do not count DCIS-regressive tumors after they have regressed, and after a DCIS-invasive tumor has transitioned to an invasive tumor we determine its detectability based on the rules for invasive cancers.[2]

D Estimation of mammogram model

We estimate our model of mammogram demand by method of moments. The moments are generated from the Erasmus model combined with our model of screening decisions. We first use the Erasmus model to generate cancer incidence and tumor growth under a no-screening assumption, as described above. Specifically, we simulate a panel of ten million women born between 1910 and 1974. We start at age 20 and model cancer incidence and tumor growth using the Erasmus model, assuming no screening. We use the tumor sizes and types to determine the population cancer rate by age.

Then, for a given set of parameters $\alpha^o, \gamma^o, \delta^o, \alpha^r, \gamma^r, \delta^r$, we apply the mammogram decision model (by age and recommendation status) to the cancer rate age profile from the Erasmus model to generate the main moments by age.

Although the model is static, it does have a dynamic element in it, as we calculate the model-generated moments only for the women who were not diagnosed with cancer in previous years, or those who did not

[2]Note that this leads to an unintuitive model behavior in which DCIS tumors are detectable at smaller sizes than invasive tumors. In the Erasmus model, invasive tumors are initialized at 0.01 mm and are not considered screen-detectable (by us) until they reach 1.09 cm. DCIS-invasive tumors are initialized at the screening threshold of the year and age in which they become invasive. Since this is sometimes smaller than 1.09 (1.09 is just the average of the distribution of screening thresholds in 2010), the model could simulate a DCIS-invasive tumor which is detectable for several years, then becomes undetectable, then becomes detectable again.

die (from breast cancer or other causes) prior to the given age. To do this, we must make an assumption about what fraction of clinically-diagnosed women under the no-screening assumption overlaps with the screen-diagnosed population when the mammogram decision model is applied. One extreme would be to assume that there is no overlap (perfect negative correlation between clinical and screen diagnosis), so that if 0.01 of the population were clinically-diagnosed under the no-screening assumption, and 0.02 of the population were screen-diagnosed for a given set of parameters, a total of 0.03 of the women would be diagnosed with cancer. We chose to make the other assumption, that there was perfect positive correlation between clinical and screen diagnosis. In this case, if 0.01 of the population were clinically-diagnosed and 0.02 were screen-diagnosed, only 0.02 of the women would be diagnosed with cancer. This likely produces an underestimate of the effects of screening, because it minimizes the number of women who are diagnosed each year.

With this simulated population of women, an assumed value of parameters associated with the mammogram decisions with and without recommendation (equations (1) and (2) in the main text) and the observed policy recommendation (40 and above), the model generates an age-specific share of women who are screened, and the tumor characteristics (in-situ and invasive rates), conditional on getting screened.

As mentioned in the main text (footnote 17), the in-situ rate moment differs from Figure 3a in the main text. Figure 3a shows the in-situ rate of all diagnosed cancers that appear in the SEER database, but the moment we match with the model is the in-situ rate of *screen-detected* cancers. Cancers that are clinically diagnosed are highly unlikely to be in-situ, so the SEER value likely underestimates the true value of share in-situ for screening mammogram-diagnosed cancers. We adjust the SEER moment at each age using Bayes' rule:

$$P(M) * P(insitu|M) + (1 - P(M)) * P(insitu| \sim M) = P(insitu),$$

where M is the event that a diagnosed tumor was screen-detected. We assume that $P(M)$, the share of diagnoses detected by screening mammogram, is 0.2 for ages 35-39 and 0.34 for ages 40-49 (following Roth et al. (2011)). We assume that $P(insitu| \sim M)$ = 0.08, following Ernster et al. (2002). $P(insitu)$ is given by the SEER moments in Figure 3a, allowing us to back out $P(insitu|M)$, our object of interest, which is the moment we actually match. The results for $P(insitu|M)$ for ages 40-49 range from 52% to 55%, which is in line with the 56% reported in this age group by Ernster et al. (2002).

With our 48 moments in hand (16 moments for each of three types), we then search for the parameters that minimize the (weighted) distance between these generated moments and the observed moments. We apply a linear weight that decreases on each side of age 40, so that the weight on moments associated with ages 39 and 41 is 10/11 of the weight on the age 40 moment, the weight on moments associated with ages 38 and 42 is 9/11 of the weight on the age 40 moment, and so on. To achieve a reasonable fit, we also weight the moments by the inverse of their standard deviation. We chose 2,000 random starting values in the parameter space defined as follows:

$$\alpha^o \in [-10, 10], \gamma^o \in [-0.2, 0.2], \delta^o_{insitu} \in [-2, 2], \delta^o_{invasive} \in [-2, 2]$$
$$\alpha^r \in [-2, 2], \gamma^r \in [-0.2, 0.2], \delta^r_{insitu} \in [-2, 2], \delta^r_{invasive} \in [-2, 2]$$

and applied the Nelder-Mead algorithm to each of these starting vectors. We then iteratively applied the Nelder-Mead algorithm to the best starting value to further minimize the objective function.

E Counterfactual simulations of mammogram model

Our counterfactuals analyze the impact of changing the recommendation age as well as the selection response. In both cases, we first model the underlying onset rate of cancer and the evolution of cancers using the Erasmus model described in Section III.B and Appendix C. Since we are interested in analyzing the impact of potential future recommendation changes, we apply the most recent year's value of any time-varying parameters of the Erasmus model. In practice, this means we use the breast-cancer-specific and non-breast-cancer mortality for 2010, the scale parameter for fatal diameter β_F^y from 1975 (see Appendix Table A.5), the screening sensitivity E_y from 2000, the screening diameter scale parameters from 2000 (see Appendix Table A.7), and the tumor incidence for the 1970 cohort (see Appendix Table A.8). We simulate this model for 10 million women's life histories, and in particular from ages 35-50.

We then apply the screening decision as described in Section III.B for each women and year. The baseline model uses the parameter values given in Table 2 in the main text, with the recommendation applied starting at age 40. We change the age of the recommendation in Table 3 and the selection parameters δ^r in Table 4.

In all counterfactuals that retain the age-40 recommendation (i.e. the ones that aim to isolate a counterfactual selection responses), we specify that the age-specific mammogram rates must be the same as in the baseline specification, while the type of women who respond to the recommendation is allowed to change. This allows the counterfactuals to consider only differences in selection, not levels. After imposing the counterfactual selection coefficients, we add an age-specific constant so that the age-specific mammogram rates are unchanged relative to the baseline. In all counterfactuals that both use the age-45 recommendation (i.e. the ones that consider a counterfactual policy recommendation) and impose alternative selection patterns, we make a similar adjustment so that the age-specific mammogram rates match those produced by the age-45 counterfactual with the baseline estimated selection parameters. The screening decisions along with underlying natural history in the Erasmus model determine whether a given mammogram screen results in a negative test, a false positive, or true positive based on the cancer type of the screened woman.

The Erasmus model parameters also reveal whether a mammogram detects a cancer early enough to prevent breast-cancer-related morality. If an invasive tumor is detected before it reaches the fatal diameter (see Appendix Table A.5 on Erasmus parameters), the person survives to die of natural causes. If the invasive tumor is detected after the tumor is larger than the fatal diameter, the person dies of breast cancer after some stochastic period determined by survival duration parameters (see Appendix Table A.5 on Erasmus parameters). Breast cancer related mortality is driven by invasive tumors; in-situ tumors are only fatal if they progress to an invasive tumor.

To estimate total spending under different counterfactuals, we first calculated in the HCCI data, the total age-specific spending in the twelve months following no mammogram, a negative mammogram, a false positive mammogram, and a true positive mammogram. At each age, each simulated woman falls into

one of these categories. We add up the spending for a given woman across ages 35-50 based on her relevant mammogram outcomes in each year. For example, suppose a woman had a true positive mammogram at age 42, and no mammograms at any other age. We would add the average spending in the HCCI data for women with no mammograms for ages 35-41, the average spending for a woman in the twelve months following a true positive mammogram at age 42, and the average spending for women with no mammograms at ages 43-50. Note that the screening decision only applies to women who are alive and have never been diagnosed with breast cancer; once a women receives a true positive diagnosis she is no longer screened.

F Sensitivity analysis

We explore the robustness of our estimates to changing features of our clinical model. In particular, we focus on statistics that can be compared with other sources, such as the share of in-situ tumors that become invasive, and the share of tumors that are non-malignant. We undertake three types of sensitivity analysis. First, we consider assumptions about the underlying incidence rate of cancer. As discussed in Section III.B, in our baseline analysis we adjusted upward the original Erasmus estimates of the underlying incidence rate of cancer to match the US population, rather than the combination of Swedish and US data on which it was originally calibrated (see Appendix C); in our first sensitivity analysis, we undo this adjustment and use the original Erasmus incidence assumptions.

Second, we consider assumptions about the share of in-situ tumors that will become invasive if not treated. The Erasmus model implies that almost two-thirds of in-situ tumors will become invasive if not treated; a review of the literature suggests that this is on the high end of model estimates, which range from 14% to 60% (Burstein et al., 2004). We therefore examine two specifications that test sensitivity to decreasing the share of in-situ tumors that become invasive. Specifically, the Erasmus model assumes that 62.5% of in-situ tumors will become invasive, while alternative estimates suggest that the fraction of DCIS tumors that would become invasive is 14% (Eusebi et al., 1994) or 28% (Page et al., 1982). In these checks, we also shift the tumor type distribution to match these estimates at age 40. This sensitivity proportionally reduces the share of DCIS-invasive tumors at all ages, and proportionally increases the share of tumors that are DCIS-regressive and DCIS-clinical at all ages. The share of invasive tumors remains the same.

Finally, we consider assumptions about the share of tumors that are non-malignant. Non-malignant refers to tumors that have no potential to be invasive and therefore would never result in a breast cancer mortality. Specifically, in our natural history model, recall that there are invasive tumors as well as three types of non-invasive tumors (also known as DCIS or in-situ): DCIS-regressive, DCIS-clinical and DCIS-invasive. The invasive and DCIS-invasive tumors are referred to as "malignant" due to their potential to cause harm, while the DCIS-regressive and DCIS-clinical tumors will never become invasive and are therefore referred to as "non-malignant." The Erasmus model's parameters (see Appendix Table A.9) imply that 3-9% of all tumors are non-malignant.[3] In contrast, estimates of over-diagnosis, or the diagnosis of a cancer that would not harm a woman in her lifetime, vary from <5% to >30% (American Cancer Society, 2017).

[3]The share of cancer that is in situ of any kind (DCIS-clinical, DCIS-regressive, or DCIS-invasive) with no screening is approximately 15% at age 35 and 9% at age 50 (see Appendix Figure A.6). The age gradient is because some of the DCIS invasive becomes invasive and some of the DCIS regressive regresses.

Compared to other models, the Erasmus model seems to have a low estimate of non-malignancy, or equivalently a high estimate of the share of cancer that is invasive or could become invasive. Therefore, each of our sensitivity analysis decreases the amount of invasive or potentially invasive tumors. Specifically, in an alternate natural history model (Fryback et al., 2006), the share of tumors with "non-malignant potential" was 42%. Alternate estimates of over-diagnosis are provided by three trials in which women in the control group were not invited to be screened at the end of the active trial period. In a meta-analysis, estimates of the excess incidence was 19% when expressed as a proportion of the cancers diagnosed during the active screening period (Marmot et al., 2013). We therefore increase the share of non-malignant tumors from approximately 8% at age 40 to (separately) 19% and 42% at age 40. In each of these sensitivity analyses, we increase the share of DCIS-regressive and DCIS-clinical at all ages in a proportional shift so that the share of non-malignant tumors at age 40 is either 19% or 42%. We separately decrease the share of tumors that are invasive or DCIS-invasive by a proportional shift so that the total tumor types sum to 100%.

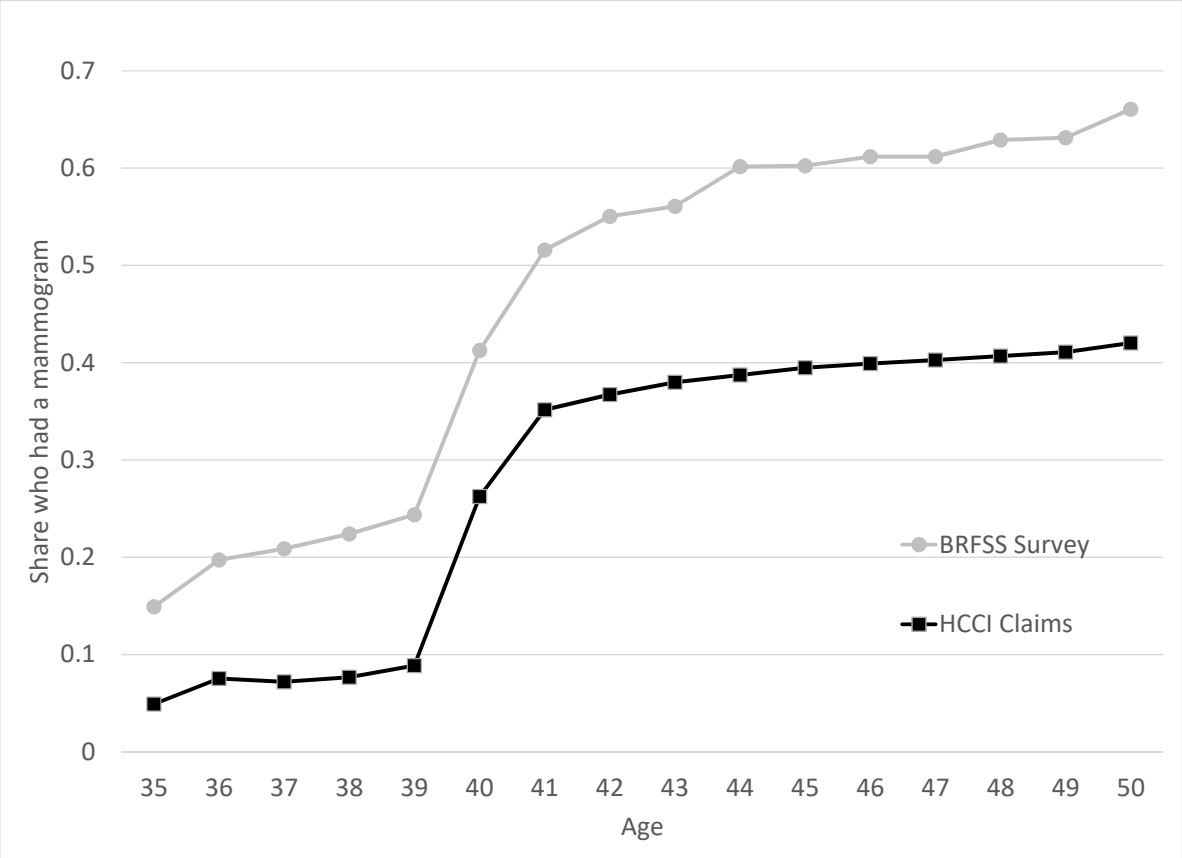
For each sensitivity analysis, we first reproduce the Erasmus model natural history with the appropriate adjustments. We then re-estimate the mammogram decision model using the same data moments (see Figure 5 in the main text) and the women simulated using the revised natural history model. To construct counterfactuals, we apply the new parameter estimates to the revised natural history model. Qualitatively, we can anticipate the impact of these changes: reducing the overall incidence of cancer, reducing the share in-situ that will transition to invasive, and increasing the share non-malignant all serve to make screening less effective, and therefore delaying screening becomes less consequential.

The results are summarized in Appendix Table A.10. As we emphasized in the main text, the details of the model are critical for the quantitative results, and indeed the mortality levels vary considerably compared to the baseline model in all specifications. In addition, the mortality cost of delaying the recommendation falls. This occurs for two reasons. First, conditional on the same mammogram decision model estimates, screening is less effective with fewer malignant tumors. Therefore, delaying screening is less costly. In addition, changing the share of tumors that are non-malignant affects the estimation of $\delta^r_{in-situ}$ as shown in Appendix Table A.11. In these sensitivity checks, $\delta^r_{in-situ}$ is lower than in the baseline estimates. This occurs because the natural history model now has more in-situ tumors. One of the moments we match is the share of in-situ tumors among diagnoses. In order to observe the same amount of in-situ diagnoses with more underlying in-situ tumors, we must be screening fewer in-situ women and more invasive women. The magnitude of this selection change depends on the magnitude of the change in the sensitivity specification; the last specification is the most aggressive in increasing the in-situ tumor share. Since the women who chose to get screened due to the recommendation now have fewer in-situ tumors (which could potentially become invasive), screening is less effective as well.

More importantly, we also examine how these sensitivity analyses affect our selection results, and here we find that the qualitative conclusions are quite robust. In all cases except one (the incidence shift, reported in row (1) of Appendix Table A.10), moving from the estimated selection to no selection (or consistent selection) has a large (relative) effect on the number of women who die by age 50. The intuition is as in the baseline model. Under the estimated selection, the women who select into the recommendation are healthier and less likely to have invasive or in-situ cancer. Therefore, the cost of delaying the recommendation (in

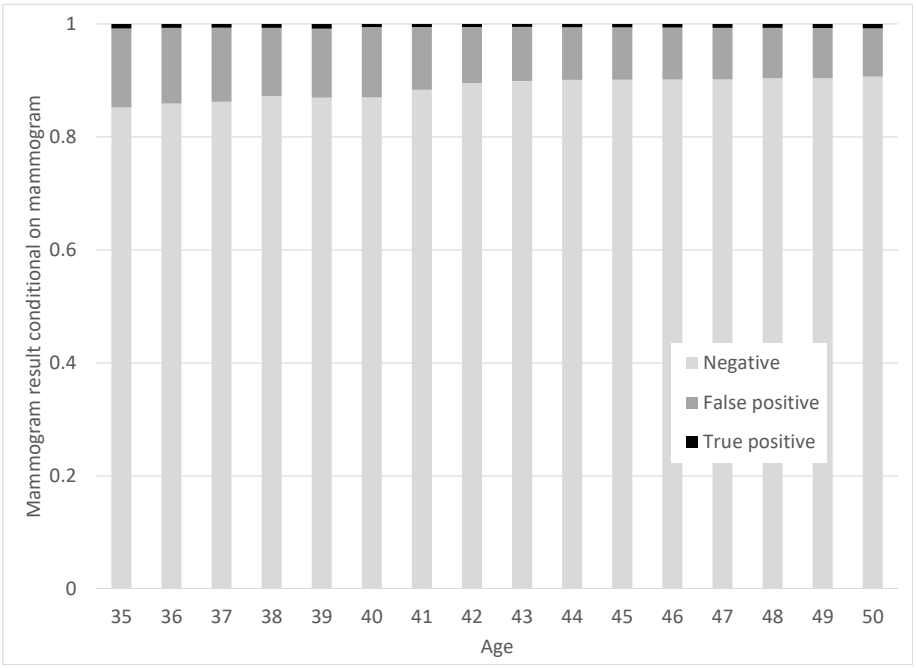
terms of lives lost) is low. If there were no selection, the women who responded to the recommendation would be more likely to have cancer than in the estimated selection specification. Thus, delaying the recommendation would have a higher cost in terms of an increase in deaths. Finally, if there were consistent selection, the women who chose to get screened due the recommendation would be more likely to have cancer. In this case, the recommendation would be highly effective and delaying screening would be very costly in terms of mortality. The one exercise for which this result does not hold is the incidence shift, since in this case the re-estimated mammogram decision model has one different parameter sign. As shown in Appendix Table A.11, in this case $\delta^r_{in-situ}$ is positive, implying recommendation-induced positive selection.

Figure A.1: Mammogram rate in survey and claims data, by age

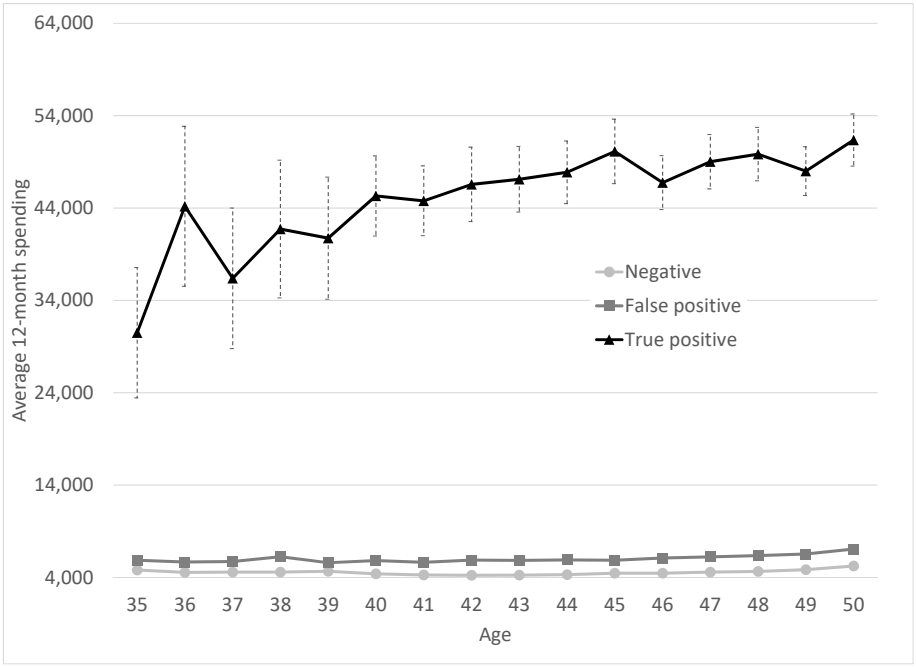


Notes: This figure shows the share of women who received a screening mammogram each year, by age. Source for survey data: Behavioral Risk Factor Surveillance System Survey (BRFSS), even years 2000-2012, restricted to women with health insurance (public or private). Source for claims data: HCCI data from 2008-2012, for mammograms between 2009-2011. Mammograms are coded in the HCCI claims data using the algorithm described in Segel et al. (2017). Mammograms are coded in the BRFSS data based on self-reports. The approximately 15-ppt discrepancy between surveyed and observed mammogram rates is consistent with the finding of Cronin et al. (2009), who document that self-reported screening rates overstated actual screening rates by 15 to 27 percentage points in a study of Vermont women.

Figure A.2: Mammogram outcomes and health care spending by age



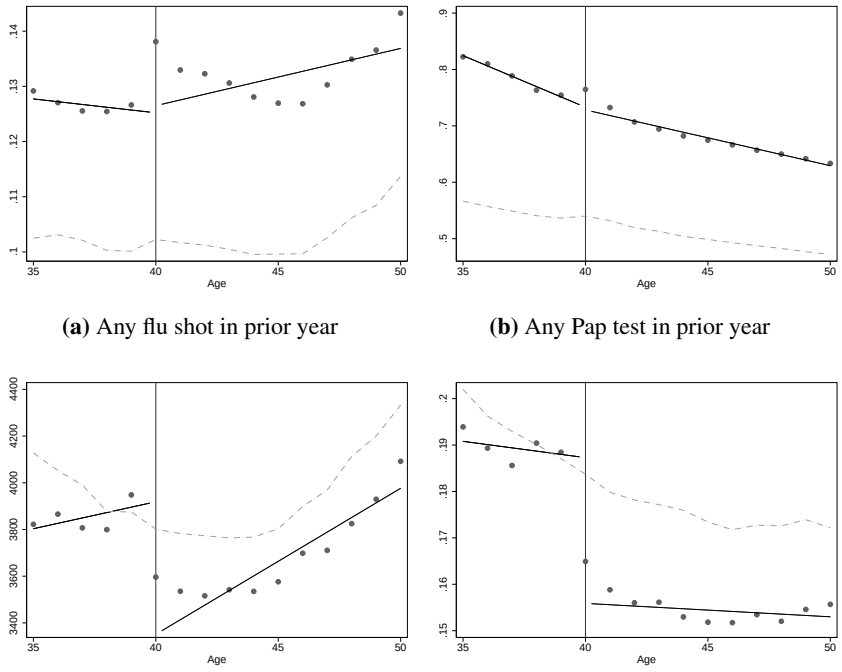
(a) Mammogram Results Conditional on Mammogram



(b) Health care spending by mammogram outcome and age

Notes: Sample is limited to the set of privately insured woman-years from the private insurance claims data who had a mammogram. N = 7,373,302 woman-years. For each age (measured by the age at the beginning of the calendar year), panel A shows the share with each mammogram outcome. Panel B focuses only on the women-years with mammograms and shows subsequent 12-month spending separately based on mammogram outcome; error bars reflect 95% confidence intervals.

Figure A.3: Health and demographic characteristics prior to mammogram by age [Health Behaviors, HCCI]



(a) Any flu shot in prior year

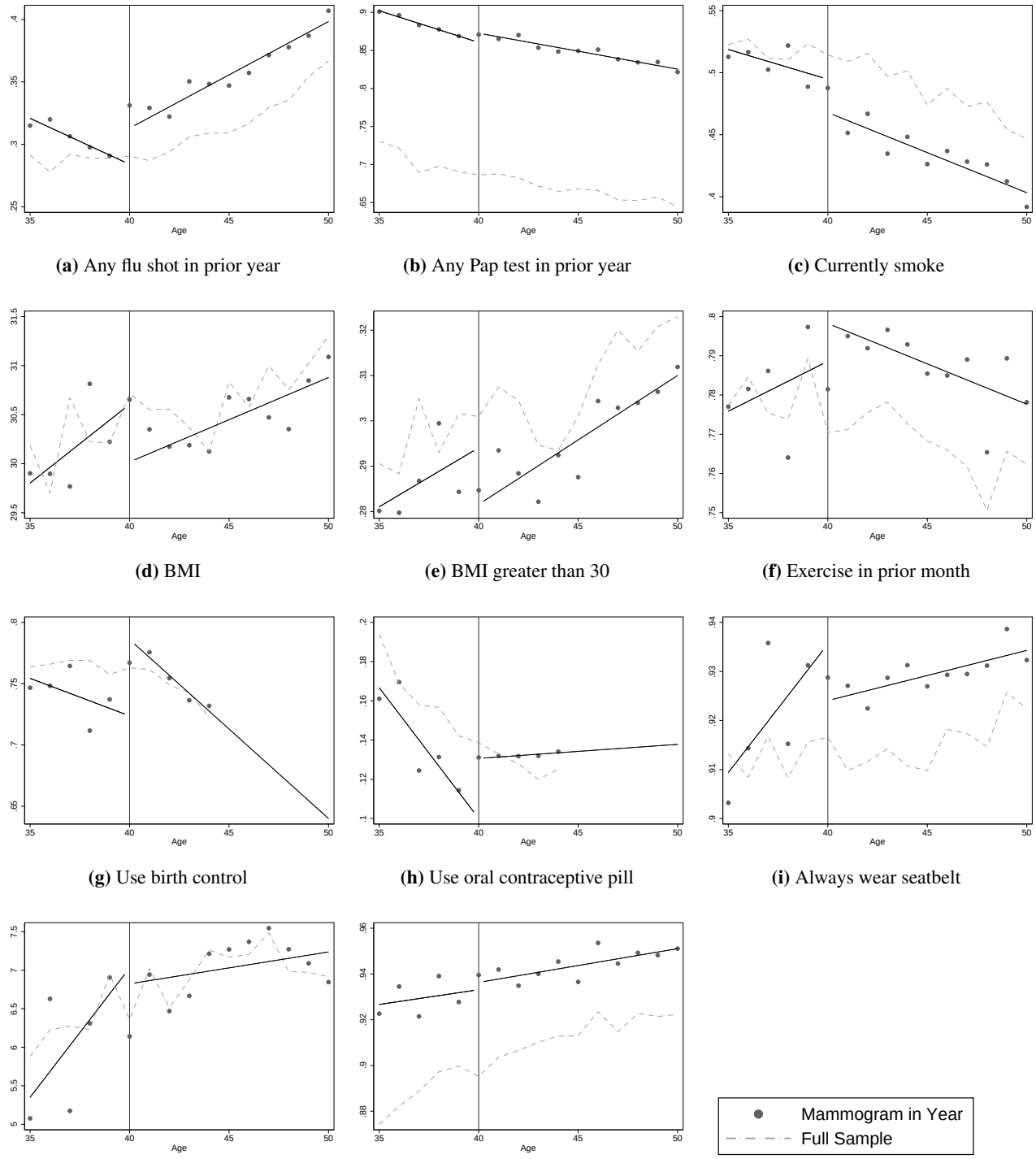
(b) Any Pap test in prior year

(c) Annual spending in prior year

(d) # of emergency room visits in prior year

Notes: Sample is insurance claims data on a set of privately insured woman-years from 2008-2012, for mammograms between 2009-2011. The x-axis plots the women's age at the time of the mammogram. In each panel, we report mean of a given outcome for women who received a mammogram at that age in solid dots. As described in Appendix B, we estimate equation (A.1) and plot the linear best-fit line for women who obtained a mammogram separately for ages below and above 40, while excluding age 40. We also plot the average share of all women with a given outcome in the dashed line. The outcome in panel (a) is an indicator for getting any flu shot in the 12 months prior to a mammogram. Panel (b) is an indicator for getting any Pap test in the 12 months prior to a mammogram. Panel (c) presents average total spending in the 12 months prior to the mammogram, not including the mammogram date. Panel (d) presents average number of emergency room visits in the 12 months prior to a mammogram. For those without a mammogram, we draw a reference date from the distribution of actual mammograms in that year. All reference dates are set to be the first of the given month. N = 7,373,302 woman-years.

Figure A.4: Health and demographic characteristics prior to mammogram by age [Health Behaviors, BRFSS]



(a) Any flu shot in prior year

(b) Any Pap test in prior year

(c) Currently smoke

(d) BMI

(e) BMI greater than 30

(f) Exercise in prior month

(g) Use birth control

(h) Use oral contraceptive pill

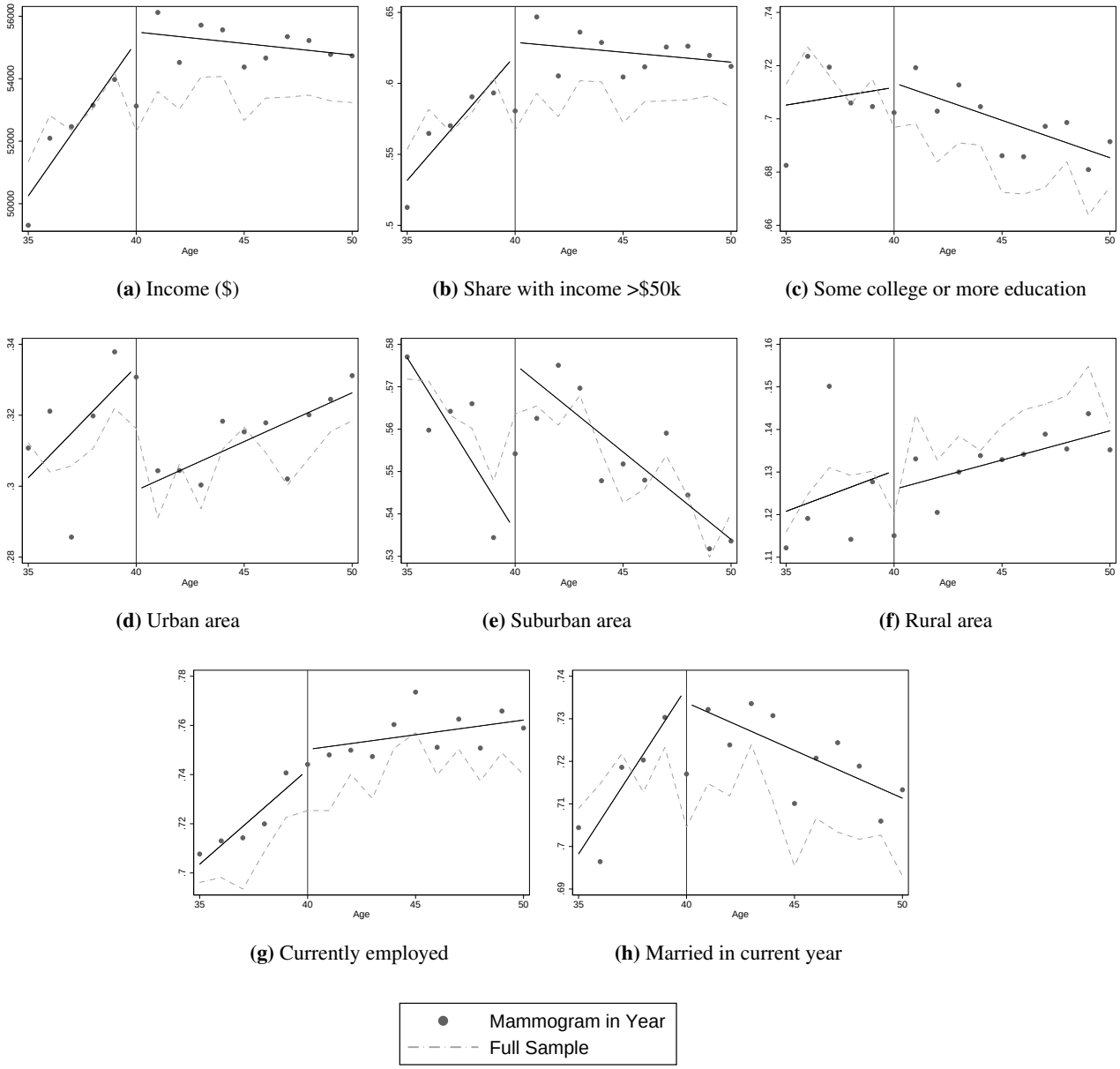
(i) Always wear seatbelt

(j) # drinks in prior month

(k) Has personal doctor

Notes: Figure replicates Appendix Figure A.3 for alternate health and demographic outcomes; see notes for Figure A.3 for implementation details. Sample is the Behavioral Risk Factor Surveillance System Survey (BRFSS), even years 2000-2012, restricted to women with any health insurance. Panel (a) is an indicator for receiving any flu shot in the prior 12 months. Panel (b) is an indicator for getting any Pap test in the prior 12 months. Panel (c) is an indicator for whether the respondent currently smokes. Panel (d) is the BMI of the respondent, computed based on the respondent's reported weight and height. Panel (e) is an indicator for a body mass index (BMI) over 30. Panel (f) is an indicator for whether the respondent participated in any physician activities for exercise in the prior month. Panel (g) is an indicator for using birth control, either for the respondent or their partner. Panel (h) is an indicator for using oral contraceptives. Panel (i) is an indicator for always wearing a seatbelt. Panel (j) is the mean number of drinks in the prior month, computed based on the number of days the respondent reported drinking and the average drinks consumed per day of drinking. Panel (k) is an indicator for whether the respondent has one person they think of as their personal doctor or health care provider. N = 382,627.

Figure A.5: Health and demographic characteristics prior to mammogram by age [Demographics, BRFSS]



(a) Income ($)

(b) Share with income >$50k

(c) Some college or more education

(d) Urban area

(e) Suburban area

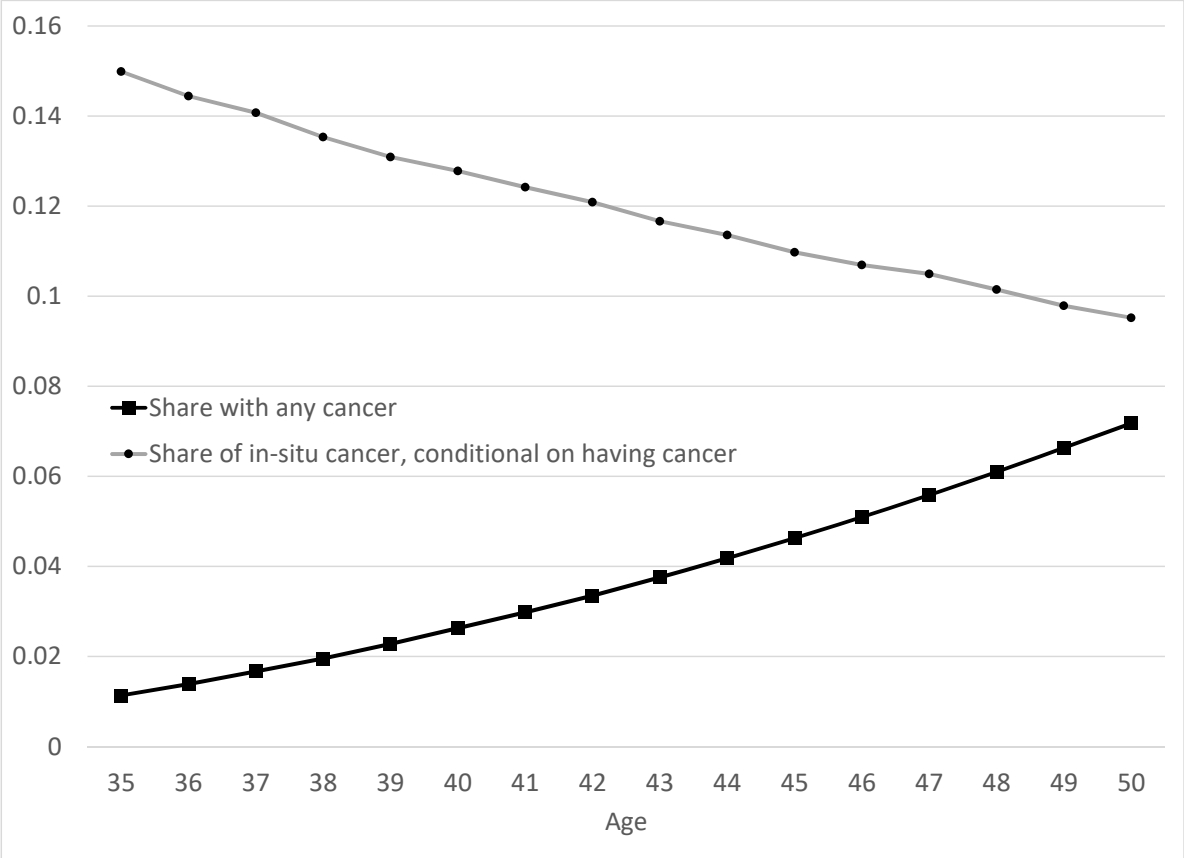
(f) Rural area

(g) Currently employed

(h) Married in current year

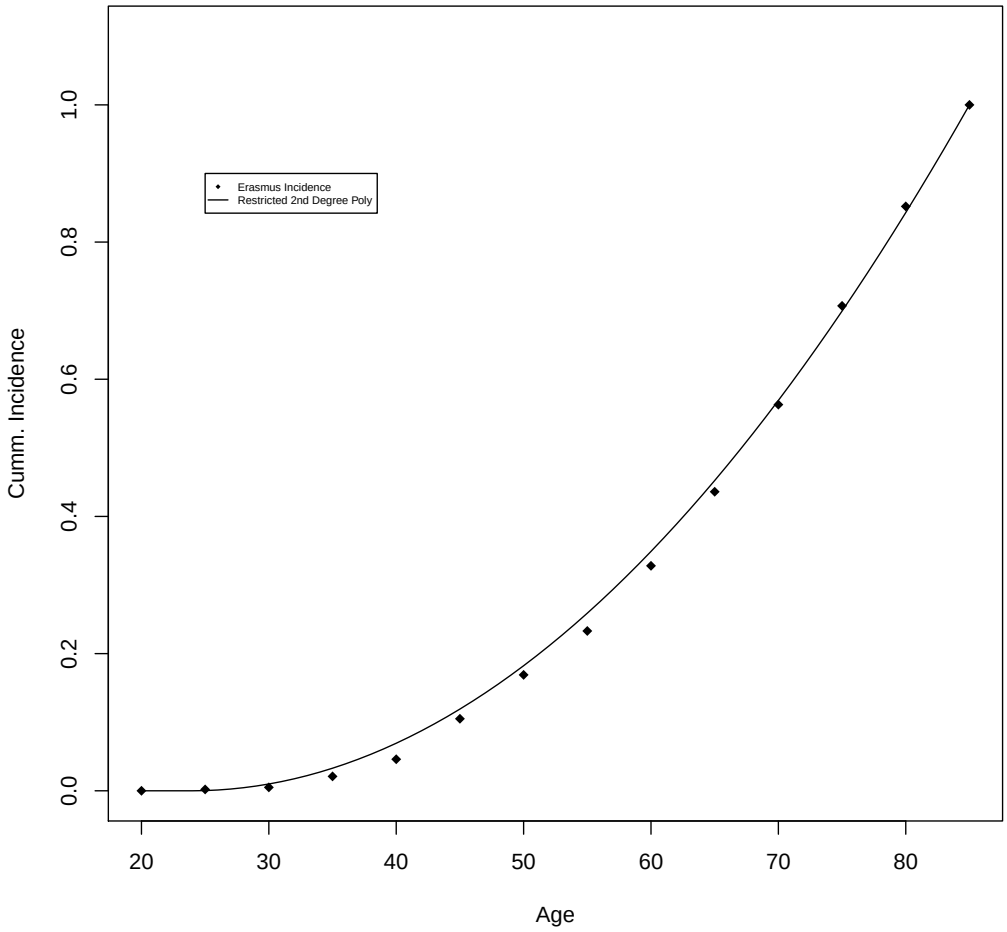
Notes: Figure replicates Appendix Figure A.3 with additional outcomes; see notes for Figure A.3 for implementation details. Sample is the Behavioral Risk Factor Surveillance System Survey (BRFSS), even years 2000-2012, restricted to women with any health insurance. Panel (a) is the average income of women who obtained a mammogram. All women in the top bin of $75,000 or above are top coded at that income. Panel (b) is an indicator for whether the income was above the median income of $50,000. Panel (c) is a binary indicator for whether the highest year of education completed included some college or technical school. Panel (d) is an indicator for whether the metropolitan status code was "in the center of an MSA." Panel (e) is an indicator for whether the metropolitan status code was "outside the city center of an MSA but inside the county containing the MSA" or "inside a suburban county of the MSA." Panel (f) is an indicator for whether the metropolitan status code was "not in an MSA." Panel (g) is an indicator for whether the respondent was currently employed for wages for self-employed. Panel (h) is a binary indicator for whether the respondent is currently married. N = 382,627.

Figure A.6: Erasmus model predictions for share with cancer and share in-situ (no screening)



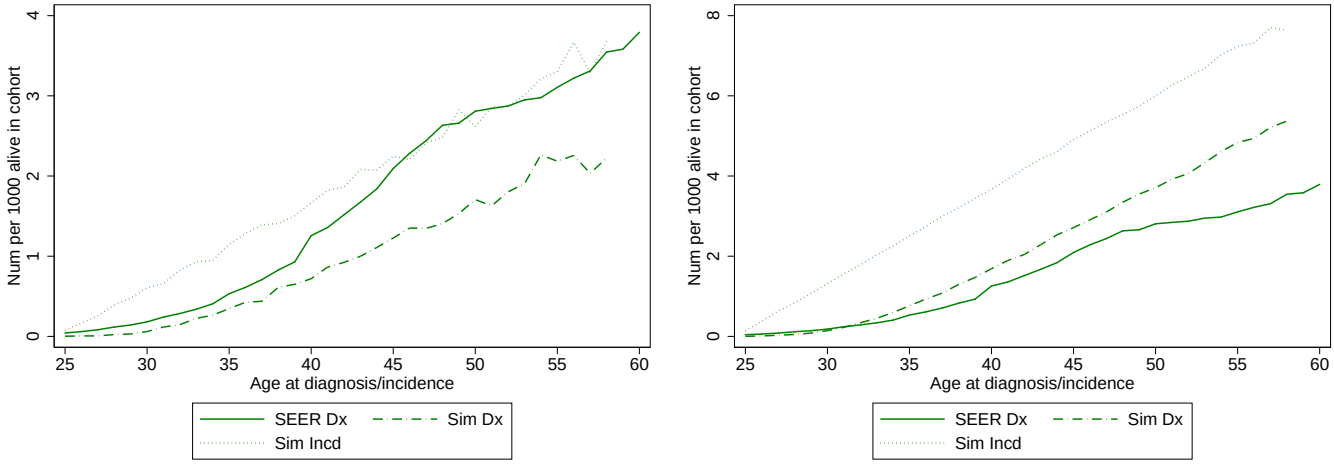
Notes: Figure presents the share with any cancer and the share of cancer in-situ in the Erasmus model, with no screening.

Figure A.7: Fitted tumor incidence by age



Notes: Figure presents the smoothed CDF of tumor incidence by age, fitted to the original Erasmus incidence in 5-year intervals.

Figure A.8: Multiplicative incidence adjustment

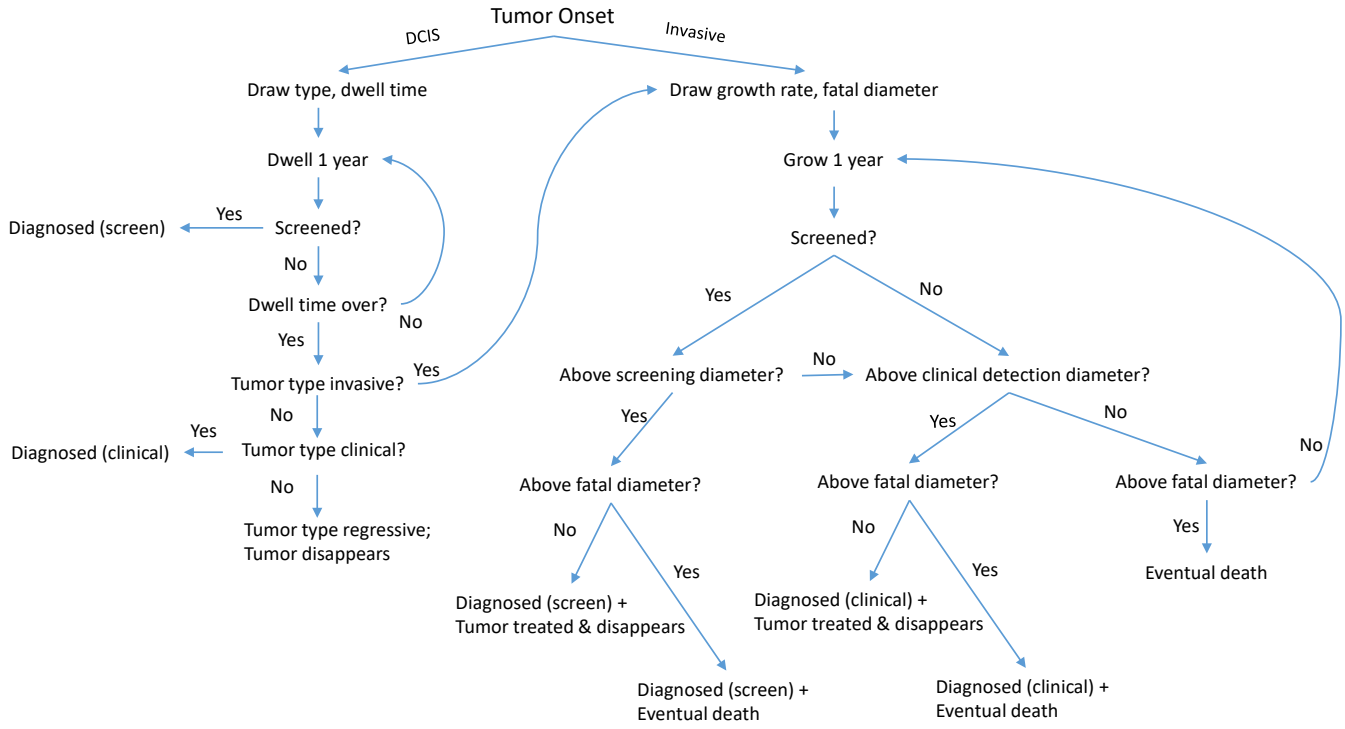


(a) Original incidence

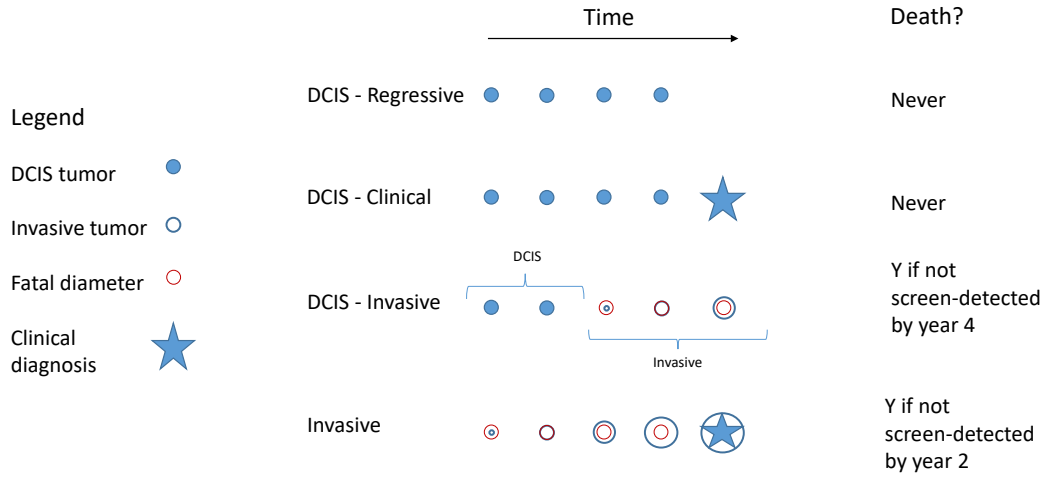
(b) Incidence multiplicatively shifted by α

Notes: Figure presents the simulated incidence and diagnosis rates compared with the SEER diagnosis rates. These are presented for both the original incidence in panel (a), and for the incidence shifted by α in panel (b). This simulation assumes no screening.

Figure A.9: Erasmus model



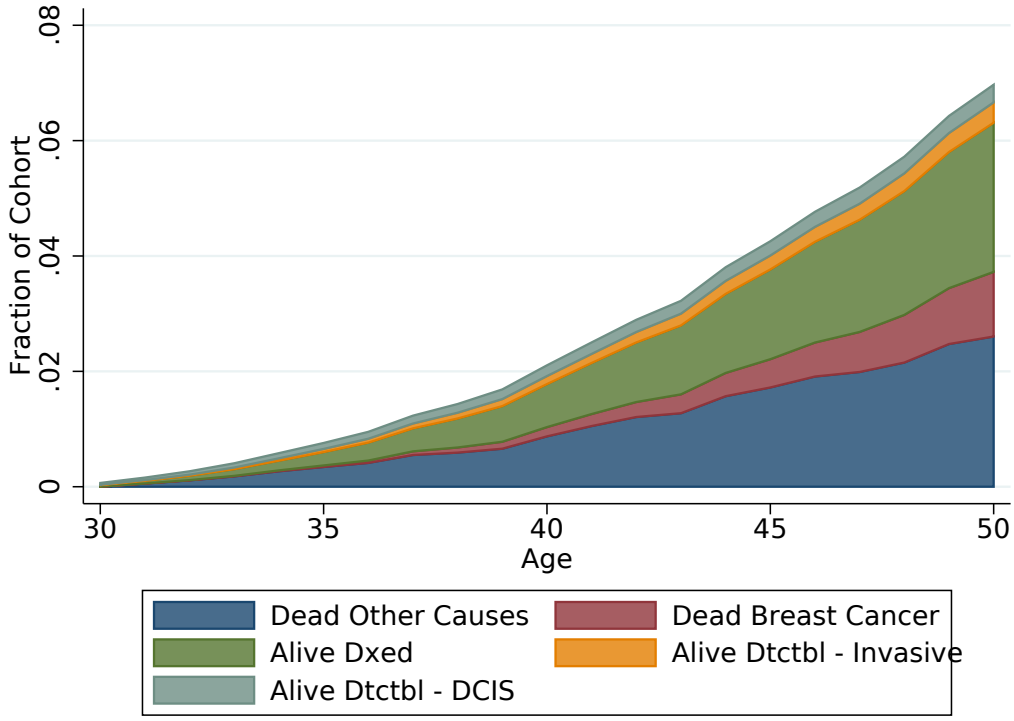
(a) Flow chart



(b) Example sequences

Notes: Panel (a) shows the flow chart of a tumor's natural history according to the Erasmus model. Panel (b) shows example sequences of progression for each different type of tumor, in the absence of screening.

Figure A.10: Cancer histories in Erasmus model



Notes: Figure shows the share of women in different categories when the Erasmus model is run without screening for birth cohorts 1950-1975, and focuses on years 2000-2005. The categories represented are "Dead Other Causes" (died due to other causes), "Dead Breast Cancer" (died due to breast cancer), "Alive Dxed" (alive and with clinically diagnosed cancer), "Alive Dtctbl - Invasive" (alive and with detectable but not yet detected invasive cancer), and "Alive Dtctbl - DCIS" (alive and with detectable but not yet detected DCIS cancer). The remainder of the population is cancer-free or has invasive or DCIS cancer that is too small to be detectable yet.

Table A.1: Codes used to identify claims

Event	Code type	CPT Codes	ICD-9 Codes
Screening mammogram	CPT procedure	77057*, G0202**	V76.12
Breast biopsy	CPT procedure	19100, 19101, 19120	85.11, 85.12, 85.20, 85.21
Breast ultrasound	CPT procedure	76645	88.73**
Radiologic breast testing	CPT procedure	76003, 77002*, 76095, 77031*, 76086, 76087, 76088, 77053*, 77054*, 76355, 76360, 76362, 77011*, 77012*, 77013*, 76098, 76100, 76101, 76102, 76120, 76125, 76140, 76150, 76350, 76365	87.35, 87.36, 87.73, 88.85
Breast cancer treatment	CPT procedure	19160, 19162, 19180, 19200, 19220, 19240, 19301**, 19303**, 19305**, 19307**, 38740, 38745	233.0, V103.0, 174.0-174.9

* indicates this code was not provided by Segel et al. (2017) but is the post-2007 analog of such a code. See http://provider.indianamedicaid.com/ihcp/Bulletins/BT200701.pdf.
** indicates this code was provided by Hubbard et al. (2015) rather than Segel et al. (2017).
Notes: This table provides the codes used to define mammograms in the HCCI and SEER-Medicare claims data. "CPT codes" are also known as "HCPCS codes."

Table A.2: Results of mammograms by diagnosis

	Diagnosed in SEER-Medicare	
	Yes	No
Negative	0.001	0.226
False Positive	0.001	0.014
True Positive	0.501	0.002
No Mammogram	0.497	0.759
N	80,408	3,327,642

Notes: This table summarizes the outcomes of mammograms for SEER-Medicare women who are diagnosed with breast cancer in that year (column 1) and not diagnosed with breast cancer in that year (column 2). Breast cancer diagnoses are recorded in the SEER linked data. Mammogram outcomes (negative, false positive, true positive, and no mammogram) are coded using the Segel et al. (2017) algorithm as described in Appendix A. We restrict to those who were diagnosed between 2007 and 2013. Sample includes both 65+ and disabled.

Table A.3: Diagnosis status by true positive result

	True positive mammogram (Conditional on screened)	
Time of Diagnosis	Yes	No
Prior to mammogram	0.001	0.000
In year of mammogram	0.722	0.000
In year following mammogram	0.145	0.022
More than 1 year after mammogram	0.016	0.142
Never diagnosed	0.116	0.836
N	55,799	952,292

Notes: This table summarizes the time of diagnosis in the linked SEER data for women who were coded as having a true positive mammogram in the SEER-Medicare data. We restrict this analysis to women who received a screening mammogram in the SEER-Medicare data, as coded in the Segel algorithm as described in Appendix A. For these women, we use the SEER-Medicare claims and the Segel algorithm to determine whether the woman had a true positive mammogram. We then compare the timing of this claims-related diagnosis with the SEER diagnosis, if any occurred. The rows refer to the year the woman was coded as having breast cancer in the SEER linked data. Source: SEER-Medicare data, diagnoses between 2007-2013.

Table A.4: Tumor characteristics

Invasive	DCIS
Size s_i^y (cm)	Dwell time w_i (years)
Growth rate g_i (1/years) *	
Screen detection diameter r_i^{ay} (cm)	
Clinical diagnosis diameter c_i (cm) *	
Fatal diameter f_i (cm)	
Survival duration since fatal u_i (years) *	

Note: This table lists the tumor characteristics for invasive and DCIS tumors. Starred variables (*) have correlated distributions; see Appendix Table A.5. Parameter values listed in Appendix Tables A.4 to A.7 are taken from Tan et al. (2006) or the extended CISNET description of the same model.

Table A.5: Model parameters

All women	Notation	Values
Probability of death from other causes	Q_y^b	Derived following Rosenberg (2006)
Probability of any breast cancer	C_b	Quadratic fit to Table A.8 plus further optimization
Age-specific probability of onset (given any onset)	S_a	Quadratic fit to values in Table A.6
Probability of invasive tumor (given tumor onset)	I_a	See Table A.9
Probability of DCIS tumor sub-type (summing to 1 - I_a)	V_a, R_a, C_a	See Table A.9
Invasive Tumors		
Mean of log of growth rate g_i	μ_G	0.062
SD of log of growth rate g_i	σ_G	0.87
Scale parameter for screen detection r_i^{ay}	β_R^{ay}	see Table A.7
Shape parameter for screen detection r_i^{ay}	η_R	2.95
Mean of log of clinical diagnosis diameter c_i	μ_C	0.97
SD of log of clinical diagnosis diameter c_i	σ_C	0.63
Scale parameter for fatal diameter f_i	β_F^y	Linear between 1915 and 1975 (0.8 in 1915; 4.0 in 1975); 4.0 after 1975
Shape parameter for fatal diameter f_i	η_F	0.95
Mean of log of survival duration u_i	μ_U	2.43
SD of log of survival duration u_i	σ_U	1.13
Correlation between g_i and c_i	ρ_{gc}	+0.41
Correlation between g_i and u_i	ρ_{gu}	-0.90
Correlation between c_i and u_i	ρ_{cu}	-0.43
DCIS Tumors		
Mean of tumor dwell time w_i[4]	W	5.22 - (time to grow from 1975 to current year screening diameter)
Screening sensitivity	E_y	Linear from 1975-2000 (0.4 in 1975, 0.8 in 2000) and 0.8 from 2001-2010

Note: This table lists the parameters of the tumor growth model, along with their values where applicable.

[4]Dwell time w_i (time from in-situ onset to invasive onset) is calculated by subtracting the time it takes the invasive tumor to grow from the 1975 screening threshold to the current screening threshold from a random draw from an exponential distribution with mean 5.22.

Table A.6: Tumor incidence by age

Age	Cumulative incidence	Age	Annual probability of incidence
25	0.002	20-24	0.0004
30	0.005	25-29	0.0006
35	0.021	30-34	0.0032
40	0.046	35-39	0.0050
45	0.105	40-44	0.0118
50	0.169	45-49	0.0128
55	0.233	50-54	0.0128
60	0.328	55-59	0.0190
65	0.436	60-64	0.0216
70	0.563	65-69	0.0254
75	0.707	70-74	0.0288
80	0.852	75-79	0.0290
85	1.00	80-85	0.0247

Note: This table shows the age distribution of the incidence of the onset of pre-clinical breast cancer (including ductal carcinoma in-situ). Source: Tan et al. (2006); author's calculations.

Table A.7: Screening diameter scale parameter

	Parameter value for age and year screened			
	30-49	50-59	60-69	70-85
1975	2.2	1.7	1.3	1.0
		(linear interpolation)		
2000	1.5	1.1	0.9	0.6

Note: This table shows the age- and screening-year-dependent values of the scale parameter for the screening diameter Weibull distribution. Linear interpolation is applied between years 1975 and 2000.

Table A.8: Tumor incidence by birth cohort: original Erasmus values

Birth cohort	Cumulative incidence
1900-04	0.122
1905-09	0.132
1910-14	0.141
1915-19	0.154
1920-24	0.169
1925-29	0.176
1930-34	0.182
1935-39	0.200
1940-44	0.220
1945-49	0.223
1950-54	0.204
1955-59	0.198
1960-64	0.193
1965-69	0.189
1970	0.187

Note: This table shows the cumulative probability (up to age 85) of the onset of pre-clinical breast cancer by birth cohort. Source: Tan et al. (2006)

Table A.9: Tumor type distribution

Age at onset	Invasive	DCIS-invasive	DCIS-regressive	DCIS-clinical
20-34	0.76	0.15	0.03	0.06
35-79		(linear interpolation)		
80-85	0.92	0.05	0.01	0.02

Note: This table shows the age-dependent proportions of incident tumor types. Linear interpolation is applied between ages 35 and 79.

Table A.10: Sensitivity checks for impact of changing mammogram recommendation age from 40 to 45

	A. Estimated Selection			B. No Selection			C. Consistent Selection		
	Dead by age 50 (per 1,000 women)			Dead by age 50 (per 1,000 women)			Dead by age 50 (per 1,000 women)		
	Recommendation at		Diff	Recommendation at		Diff	Recommendation at		Diff
	Age 40	Age 45		Age 40	Age 45		Age 40	Age 45	
Baseline Estimate	15.98	16.03	0.05	15.84	16.02	0.18	15.54	15.99	0.45
Decrease cancer incidence to:									
(1) Erasmus original level	10.66	10.68	0.02	10.67	10.68	0.01	10.65	10.68	0.03
Decrease share of in-situ tumors that become invasive:									
(2) from 62.5% to 28%	15.20	15.22	0.02	15.06	15.21	0.15	14.78	15.18	0.41
(3) from 62.5% to 14%	14.89	14.90	0.01	14.75	14.89	0.14	14.48	14.86	0.39
Increase share of non-malignant tumors:									
(4) from 6% to 19%	15.08	15.12	0.03	14.90	15.10	0.20	14.63	15.07	0.44
(5) from 6% to 42%	12.69	12.70	0.02	12.53	12.68	0.15	12.36	12.66	0.30

Notes: Table reports model predictions under alternate sensitivity assumptions. The first three columns in the first row replicate the results from Panel A of Table 4 in the main text, on the impact of changing the mammogram recommendation age from 40 to 45 based on the estimated selection patterns. We report only the impact on the death rate by age 50. The second three columns replicate the results from Panel B of Table 4 where we instead assume "no selection" (i.e. we set $\delta^r = \delta^o = 0$.) The last set of columns reflect Panel C of Table 4 where we assume "consistent selection" (i.e. we set δ^requal to our estimates of δ^o in Table A.11). Each row tests the sensitivity of these estimates under alternate natural history assumptions, as discussed in Appendix F.

Table A.11: Sensitivity checks for parameter estimates

		Sensitivity Checks				
		(1)	(2)	(3)	(4)	(5)
	Baseline		Share In-situ to Invasive		Share Non-Malignant	
Parameter	Estimate	Incidence	28%	14%	19%	42%
α^o	-5.21	-4.81	-5.20	-5.20	-4.67	-3.33
γ^o	0.10	0.09	0.10	0.10	0.09	0.05
$\delta^o_{\text{in-situ}}$	0.36	1.15	0.36	0.36	0.09	-0.16
$\delta^o_{\text{invasive}}$	1.13	10.89	1.13	1.13	1.15	1.58
α^r	0.29	-0.06	0.26	0.26	0.03	-0.73
γ^r	-0.03	-0.02	-0.02	-0.02	-0.02	0.00
$\delta^r_{\text{in-situ}}$	-0.01	0.58	-0.01	0.00	-0.55	-1.06
$\delta^r_{\text{invasive}}$	-4.67	22.28	-5.10	-12.80	-6.67	-13.67

Notes: Table shows the parameter estimates from the mammogram decision model under alternate sensitivity assumptions. Specifics for each of the columns are discussed in Appendix F.

References

Abadie, A., 2002: Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models. *Journal of the American Statistical Association*, **97 (457)**, 284–292, doi:10.1198/016214502753479419, URL https://doi.org/10.1198/016214502753479419, publisher: Taylor & Francis _eprint: https://doi.org/10.1198/016214502753479419.

Abadie, A., 2003: Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, **113 (2)**, 231–263, doi:10.1016/S0304-4076(02)00201-4, URL http://www.sciencedirect.com/science/article/pii/S0304407602002014.

American Cancer Society, 2017: Breast Cancer Facts & Figures 2017-2018. American Cancer Society, Inc., URL https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf.

Burstein, H. J., K. Polyak, J. S. Wong, S. C. Lester, and C. M. Kaelin, 2004: Ductal Carcinoma in Situ of the Breast. *New England Journal of Medicine*, **350 (14)**, 1430–1441, doi:10.1056/NEJMra031301, URL https://doi.org/10.1056/NEJMra031301.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler, 2016: The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, **315 (16)**, 1750–1766, doi:10.1001/jama.2016.4226, URL https://jamanetwork.com/journals/jama/fullarticle/2513561, publisher: American Medical Association.

Clarke, L. D., S. K. Plevritis, R. Boer, K. A. Cronin, and E. J. Feuer, 2006: Chapter 13: A Comparative Review of CISNET Breast Models Used To Analyze U.S. Breast Cancer Incidence and Mortality Trends. *JNCI Monographs*, **2006 (36)**, 96–105, doi:10.1093/jncimonographs/lgj013, URL https://academic.oup.com/jncimono/article-lookup/doi/10.1093/jncimonographs/lgj013.

Cronin, K. A., and Coauthors, 2009: CEBP Focus on Cancer Surveillance: Bias Associated With Self-Report of Prior Screening Mammography. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, **18 (6)**, 1699–1705, doi:10.1158/1055-9965.EPI-09-0020, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2771779/.

Ernster, V. L., and Coauthors, 2002: Detection of ductal carcinoma in situ in women undergoing screening mammography. *Journal of the National Cancer Institute*, **94 (20)**, 1546–1554.

Eusebi, V., E. Feudale, M. Foschini, A. Micheli, A. Conti, C. Riva, S. Di Palma, and F. Rilke, 1994: Long-term follow-up of in situ carcinoma of the breast. Vol. 11, 223–235, URL https://europepmc.org/article/med/7831534.

Freeman, J., C. Klabunde, N. Schussler, J. Warren, B. Virnig, and G. Cooper, 2002: Measuring Breast, Colorectal, and Prostate Cancer Screening with Medicare Claims Data. *Medical Care*,

40 (8), (Supplement):IV–36–IV–42, URL https://journals.lww.com/lww-medicalcare/Abstract/2002/08001/Measuring_Breast,_Colorectal,_and_Prostate_Cancer.5.aspx.

Fryback, D. G., N. K. Stout, M. A. Rosenberg, A. Trentham-Dietz, V. Kuruchittham, and P. L. Remington, 2006: The Wisconsin Breast Cancer Epidemiology Simulation Model. *Journal of the National Cancer Institute. Monographs*, **(36)**, 37–47, doi:10.1093/jncimonographs/lgj007.

Hubbard, R. A., W. Zhu, S. Balch, T. Onega, and J. J. Fenton, 2015: Identification of abnormal screening mammogram interpretation using Medicare claims data. *Health Services Research*, **50 (1)**, 290–304, doi:10.1111/1475-6773.12194.

Kowalski, A. E., 2019: Behavior within a Clinical Trial and Implications for Mammography Guidelines. Working Paper 25049, National Bureau of Economic Research. doi:10.3386/w25049, URL http://www.nber.org/papers/w25049, series: Working Paper Series.

Marmot, M. G., D. G. Altman, D. A. Cameron, J. A. Dewar, S. G. Thompson, and M. Wilcox, 2013: The benefits and harms of breast cancer screening: an independent review. *British Journal of Cancer*, **108 (11)**, 2205–2240, doi:10.1038/bjc.2013.177, URL https://www.nature.com/articles/bjc2013177.

Page, D. L., W. D. Dupont, L. W. Rogers, and M. Landenberger, 1982: Intraductal carcinoma of the breast: Follow-up after biopsy only. *Cancer*, **49 (4)**, 751–758, doi:10.1002/1097-0142(19820215)49:4<751::AID-CNCR2820490426>3.0.CO;2-Y, URL https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%2819820215%2949%3A4%3C751%3A%3AAID-CNCR2820490426%3E3.0.CO%3B2-Y, _eprint: https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0142%2819820215%2949%3A4%3C751%3A%3AAID-CNCR2820490426%3E3.0.CO%3B2-Y.

Rosenberg, M. A., 2006: Competing risks to breast cancer mortality. *Journal of the National Cancer Institute. Monographs*, **(36)**, 15–19, doi:10.1093/jncimonographs/lgj004.

Roth, M. Y., J. G. Elmore, J. P. Yi-Frazier, L. M. Reisch, N. V. Oster, and D. L. Miglioretti, 2011: Self-detection remains a key method of breast cancer detection for U.S. women. *Journal of Women's Health (2002)*, **20 (8)**, 1135–1139, doi:10.1089/jwh.2010.2493.

Segel, J., R. Balkrishnan, and R. Hirth, 2017: The Effect of False-positive Mammograms on Antidepressant and Anxiolytic Initiation. *Medical Care*, **55 (8)**, 752–758, URL https://journals.lww.com/lww-medicalcare/Abstract/2017/08000/The_Effect_of_False_positive_Mammograms_on.3.aspx.

Tan, S. Y. G. L., G. J. van Oortmarssen, H. J. de Koning, R. Boer, and J. D. F. Habbema, 2006: The MISCAN-Fadia Continuous Tumor Growth Model for Breast Cancer. *JNCI Monographs*, **2006 (36)**, 56–65, doi:10.1093/jncimonographs/lgj009, URL https://academic.oup.com/jncimono/article/2006/36/56/917299.