Quantifying Diminishing Return to Mammography Screenings Using Individual Medical Histories

Przemyslaw Jeziorski† Shelley Hwang§ Sadat Reza‡ Teck-Hua Ho*

February 22, 2023

Abstract

The promotion of mammography greatly expanded screening coverage; however, increased cancer detection ignited concerns regarding overdiagnosis, thus prompting a reevaluation of mammography screening guidelines. We use medical records of all Singaporean women and plausibly exogenous variation of screening coverage to quantify the marginal efficacy of screening and the extent of overdiagnosis. Women predisposed to aggressive cancers are 4x more likely to screen than those susceptible to non-aggressive forms. Self-selection results in overdiagnosis and negligible marginal efficacy once population coverage exceeds 50%. Conversely, transitioning from biennial to annual screening and boosting coverage in underserved communities proves beneficial regardless of population coverage.

Keywords: mammogram; breast cancer; diminishing returns; screening coverage; self-selection

JEL Classification: I12, I18, J16

†Corresponding author: University of California, Berkeley, przemekj@berkeley.edu
§Duke University, shelley.hwang@duke.edu
‡Nanyang Technological University, Singapore, sreza@ntu.edu.sg
*Nanyang Technological University, Singapore, President@ntu.edu.sg

1 Introduction

Economics and medical literature have long emphasized the importance of mammography screening as an effective means of preventing breast cancer deaths. However, because screening is costly and may result in false positives (see Rosenbaum, 2014)), the optimal screening approach is a continuing debate among academics and medical professionals. The issue focuses on two main aspects: who should be screened and at what age should screening begin. In other words, the policymakers strive to choose screening coverage[1] (e.g., screening 60 vs. 80 percent of the population of a given age) and the cut-off age for screening recommendation (e.g., 40 vs. 45 years old), balancing the cost of screening and its effectiveness.

Screening is effective if it reduces breast cancer-related mortality. If mammography is equally effective for all women, higher coverage rates and lower age mandates should reduce mortality. However, somewhat surprisingly, Bleyer and Welch (2012) find that the 1987-2008 increase in U.S. mammography coverage from 29% to 67.1% results in a significant increase in early-stage cancer diagnosis but virtually no decrease in cancer deaths. The authors conclude that screening has little effect in lowering mortality because of overdiagnosis.[2] These results provoked skepticism about recommending frequent screening for all women. For instance, in 2009 U.S. Preventive Services Task Force altered the prior screening recommendation for women aged 50 to 74 from annual to biannual mammography.[3]

In this paper, we revisit the effectiveness of mammography by evaluating the degree of overdiagnosis using the medical data of all Singaporean women aged 40 and older. We estimate the marginal benefit of screening and the extent of overdiagnosis as a function of screening coverage. Our estimates imply diminishing marginal screening efficacy and increasing overdiagnosis rates as screening coverage increases. In particular, expanding mammography screening is highly effective when the screening rates are low (between 0% and 35% of women screening once every two years) but has negligible effectiveness if the current screening rate is moderate or high (above 50% of women screening once every two years). In other words,

[1] Screening coverage is defined as a percentage of women that screened in the last two years.
[2] Overdiagnosis is defined as positive mammography, which detects cancer that would not progress to a late stage if left undiagnosed.
[3] See Nelson, Pappas, Cantor, Griffin, Daeges, and Humphrey (2016).

expanding screening coverage beyond 50% would predominantly lead to overdiagnosis, replicating the results of Bleyer and Welch (2012).[4]

We show that self-selection to undergo screening based on a latent propensity to develop aggressive cancer is the mechanism underlying the diminishing returns from screening. Specifically, women at risk for aggressive malignancies are four times more likely to undergo screening than women at risk for non-aggressive cancers. By definition, screening is only helpful for women at risk for aggressive cancer, whereas it leads to overdiagnosis for women at risk for non-aggressive cancer. Under the current degree of private information, most vulnerable women are already screened at a relatively low screening coverage. Consequently, screening is less effective for marginal women at medium and high screening coverage. We estimate that just 23% of women are susceptible to aggressive cancer; thus, if women had complete information about their potential cancer type, the system might achieve near 100% screening efficacy with only 23% coverage.

Our findings have implications for the development of breast cancer screening guidelines. Primarily, they explain the negative findings of Bleyer and Welch (2012) while allowing for the non-negligible benefits of screening for some women. In addition, our model reconciles the null effects with the research demonstrating the advantage of screening (for instance, Einav, Finkelstein, Oostrom, Ostriker, and Williams (2020) show that lowering the age requirement from 45 to 40 years old results in significantly increased early diagnosis and lowered mortality). Moreover, our findings caution that using the estimates of the marginal effectiveness of screening when designing policies applicable to inframarginal women may be misleading. For example, ethnic groups with limited coverage may benefit from expanding screening, even if the estimates of screening effectiveness in the general population indicate negligible efficacy. This duality occurs because the relevant marginal effect of screening may differ between subpopulations due to variables limiting access to mammography that are orthogonal to the risk of acquiring late-stage cancer. For instance, we demonstrate that Singapore's Malay population has low screening coverage and benefits the most from screening expansion.

[4] The quoted estimate of recommended coverage assumes the negligible monetary cost of screening and overdiagnosis. Using our effectiveness tabulations to obtain optimal coverage for a given degree of the monetary cost of screening and overdiagnosis is relatively easy.

Obtaining viable estimates of self-selection is challenging. First, screening decisions are often endogenous, leading to simultaneity in observed screening coverage and screening effectiveness. Second, the data may only identify a local average treatment effect of screening at the observed coverage rate. Because screening coverage in many previous studies was moderate or high, most marginal women might be non-susceptible to aggressive cancer. In such a case, even if the screening decision is exogenous, the data may be insufficient to identify self-selection depending on cancer susceptibility. We address these issues by analyzing a unique dataset of all Singaporean women aged 40 and older. The Singaporean data indicates age, ethnicity, screening events, and cancer progression. Conveniently, due to plausible external socioeconomic and religious factors, screening coverage differs significantly between the observed ethnic groups. The within-group screening rates also experience a plausible two-year exogenous increase due to the government's massive marketing push. This variation allows us to estimate the causal impact of screening on mortality and demonstrate how the marginal benefit varies as the screening rate increases from low to moderate and high.

We postulate a screening and cancer progression model to identify self-selection and obtain policy counterfactuals. The model estimates are driven by the outcomes of cancers detected at various coverage rates in the data. Consistently with self-selection, cancers detected in years with high screening coverage have reduced mortality rates. We also detect lower mortality rates amongst stage-one cancers detected early at higher coverage. Decreasing mortality amongst early-detected cancers rules out the possibility that higher coverage may be related to lower mortality simply because it promotes early detection.

Recent research reveals self-selection when undergoing preventative screening (Oster, 2020) and mammography (Kim and Lee, 2017; Kowalski, 2018; Einav, Finkelstein, Oostrom, Ostriker, and Williams, 2020). For instance, Kim and Lee (2017) examine the influence of public sector incentives for mammography in Korea. They find that the incentives provided by Korea's National Screening Program increased coverage and cancer detection. Nevertheless, similarly to Bleyer and Welch (2012), the program failed to decrease mortality, possibly because the individuals who decided not to participate were significantly more susceptible to cancer.

Kowalski (2018) examine data from the randomized controlled trial conducted for the Canadian National Breast Screening Study (CNBSS) among women 40-49 years of age. They find evidence that healthier women screen more often and are more susceptible to overdiagnosis. Moreover, their results do not indicate a significant improvement in future breast cancer rates (measured 20 years forward) among women who screened when 40-49 years old compared to those who did not.

Complementary work by Einav, Finkelstein, Oostrom, Ostriker, and Williams (2020) focuses on the policy decision to begin screening at age 40 or 45. We focus on the marginal effectiveness of screening as coverage increases and how this may impact under-screened subpopulations. Consistent with our narrative, the authors provide evidence of "positive selection" in mammograms. Specifically, women who choose screening without recommendation have a higher incidence of aggressive cancer. In addition, they demonstrate significant disadvantages associated with shifting recommendations from age 40 to 45. This result is also consistent with our findings because, in our paradigm, lowering the screening age and increasing the frequency may be beneficial even if the marginal effect of increasing coverage is negligible. For instance, we demonstrate that at 50% coverage, reducing the screening interval from every two years to once a year reduces the incidence of late-stage cancer by approximately 20 cases per 100,000 women.

Kowalski (2021) examines the evidence on the effect of mammography on breast-cancer-caused mortality and all-cause mortality over 20 years using CNBSS data and finds little evidence of early-stage screening resulting in a significant reduction in mortality. She argues that while it may be challenging to estimate the extent of and the harms caused due to overdiagnosis, the evidence suggests harms may outweigh the benefits of increased utilization (which increases cancer detection) of mammography screening. Notably, several studies in the medical literature reported overdiagnosis of breast cancer due to mammography, using both observational data Bleyer and Welch (2012) and randomized experiment data (Zackrisson, Andersson, Janzon, Manjer, and Garne, 2006; Baines, To, and Miller, 2016; van den Ende, Oordt-Speets, Vroling, and van Agt, 2017).

Researchers also examined the effect of insurance coverage and mammography screening utilization. For instance, Finkelstein, Taubman, Wright, Bernstein, Gruber, Newhouse, Allen, Baicker, and Group (2012) examined the effect of Medicaid expansion, and Bitler and Carpenter (2016) studied the effect of state insurance mandates requiring coverage insurance to find that such expanded coverage increased mammography utilization rate.

2 Mammography screening in Singapore

Breast cancer affects 1 in 11 women in Singapore and causes approximately 400 deaths a year. Between 2005 and 2009, breast cancer was a leading cause of cancer-related deaths; thus, detecting breast cancers at their early stage is an essential factor in reducing mortality. Mammogram screenings are a critical tool facilitating early diagnosis, so screening coverage is a widely used statistic that summarizes the level of preventive care. Coverage is typically defined as a percentage of women between 40 to 69 years old that underwent mammography in the last two years. According to Loy, Molinar, Chow, and Fock (2015), mammography coverage in Singapore was approximately 30% in 2001. In comparison, mammography coverage in the United States in 2001 was over 70%.[5]

To increase mammography coverage, in 2001, the government of Singapore introduced the National Breast Cancer Screening Programme, called Breast Screen Singapore (BSS).[6] The program aims to provide all women aged 40 to 64 with subsidized screening mammography. The procedure is performed in specialized clinics across the country. After implementing BSS, all women aged 50 to 64 received a personalized invitation to attend mammography. BSS also included community outreach to encourage self-examination. Following BSS, the screening coverage increased to 39% (Singapore Ministry of Health, 2011), with more than 46% of mammograms performed in BSS clinics and the remainder performed at restructured hospitals, private hospitals, and private X-ray centers (Loy, Molinar, Chow, and Fock, 2015).

We obtained a complete set of medical histories between 1993 and 2013 from the Singapore Cancer Registry. The data includes all women with detected breast cancer and indicates unique

[5] Source: https://www.cdc.gov/nchs/data/hus/2010/086.pdf
[6] Currently BSS is a part of the broader 'Screen for Life' program of the Health Promotion Board, Ministry of Health, Singapore.

patient identifier, age at the initial detection, date of death (if died due to breast cancer), and ethnic group.

Figure 1(A) contains the time trend of breast cancer incidence. Early- and late-stage cancer incidence soared between 1993 and 2001, potentially due to an aging population and decreasing mortality for non-breast-cancer-related causes. Between 2001 and 2013, we observe a further increase in early cancer incidences, partly due to more extensive screening coverage. We also observe a somewhat flatter but still increasing trend in late-stage cancer incidences.

--- Insert Figure 1 here ---

Our data on cancer incidences are supplied with a complete list of screenings done via BSS. Figure 1(B) contains the time trend in breast cancer coverage (screened in the last two years) computed using the BSS sample. We observed an initial increase in screening coverage in 2003 and 2005. However, this increase was followed by a slight decrease in coverage from 2007-2013. According to our conversations with the program administrators, later years of BSS had less intense media campaign efforts.

According to the Singapore Ministry of Health (2002), in 2001, 52% of women aged 40 to 69 years were aware of mammography. However, this awareness varies amongst three major ethnic groups in Singapore. Specifically, Chinese women had a higher level of awareness (54%) compared with Indian women (42%) and Malay women (41%). Figure 1(B) disaggregates the BSS screening coverage for Malay and non-Malay populations and shows a persistent but somewhat shrinking gap in Malay and non-Malay BSS coverage. In addition, Figure 1(C) shows the trend in late-stage cancer incidence between Malay and non-Malay populations. Taking this evidence together, we conclude that the Malay population screens at a lower rate and experiences a significantly larger incidence of late-stage cancer.

The reasons for ethnic disparities in cancer survival rates have been long studied in the American healthcare system. Li, Malone, and Daling (2003) identify that “blacks, American Indians, Hawaiians, Indians and Pakistanis, Mexicans, South and Central Americans, and Puerto Ricans had 1.4- to 3.6-fold greater risks of presenting with stage IV breast cancer”

relative non-Hispanic whites. Lannin, Mathews, Mitchell, Swanson, Swanson, and Edwards (1998) show that the majority of the gap can be attributed to economic factors, such as: having low income, never having been married, having no private health insurance, delaying seeing a physician because of money, or lacking transportation; and cultural beliefs, such as, "air causes cancer to spread," "the devil can cause a person to get cancer," or "women who have breast surgery are no longer attractive to men." Smith-Bindman, Miglioretti, Lurie, Abraham, Barbash, Strzelczyk, Dignan, Barlow, Beasley, and Kerlikowske (2006) link the ethnic gap in the US mortality rate due to breast cancer with differences in access to mammography screening. They find that "African-American, Hispanic, Asian, and Native American women were more likely than white women to have received inadequate mammographic screening" and that "The observed differences in advanced cancer rates between African American and white women were attenuated or eliminated after the cohort was stratified by screening history." Similarly, Mitchell, Lannin, Mathews, and Swanson (2002) find that strong religious beliefs supporting "religious intervention in place of treatment" may contribute to later cancer detection and a higher rate of advanced-stage cancer diagnosis.

The above regularities identified in the US population parallel Singapore. As mentioned before, the indigenous Malayan population faces stark disadvantages in access to mammography and exhibits a more significant incidence of late-stage breast cancer than non-Malayans. At the same time, Malayans are subject to several socio-economic factors that were shown to lower mammography adoption. They have a relatively low university degree percentage, have lower household income, and are more driven by religious beliefs (see Table 1).

The premise of our empirical analysis is to utilize the pre-BSS cross-sectional variation in exposure to screening across ethnic groups. BSS increased screening exposure for all ethnic groups, enabling us to estimate the marginal benefit of extra screening for various levels of initial coverage. The following section presents descriptive evidence that the benefits of extra screening are decreasing as coverage increases.

--- Insert Table 1 here ----

3 Descriptive analysis

We focused on women who were at least 40 years of age. It is helpful to use the following stylized model of breast cancer incidence. Suppose that breast cancers vary by their progression rates. Some cancers are non-aggressive; they do not progress to the late stage or progress slowly. Some cancers are aggressive and progress to the late stage faster. The population is assumed to consist of three types of individuals, NC (no cancer), NA (non-aggressive cancer), and AC (aggressive cancer), with varying susceptibilities to develop non-aggressive and aggressive cancer. NC individuals do not contract cancer. NA individuals are susceptible to contracting non-aggressive cancer but are not susceptible to contracting aggressive cancer. Lastly, AC individuals are susceptible to contracting aggressive cancer. All individuals detected with late-stage cancer are of the AC type. However, individuals detected with early-stage cancer can be either AC or NA. According to some hazard rates, NA individuals may become AC individuals as they age. In other words, initially, non-aggressive cancer may become aggressive. Importantly, if cancer is detected early, mammography provides little information about the aggressiveness of cancer. Therefore, early detection of NA-type cancer may lead to overdiagnosis.

--- Insert Figure 2 here ----

Our primary empirical evidence relies on the characteristics of women that undergo screening as screening coverage varies (across subpopulations and over time). We define cancer susceptibility as the proportion of AC-type women. The critical factor in deciding optimal screening coverage is the susceptibility of the marginal woman (i.e., the marginal benefit of screening) at various coverage levels. If the marginal susceptibility is high, increasing screening may be beneficial. If the marginal susceptibility is low, increasing screening may not be beneficial and may lead to overdiagnosis. Figure 2 depicts two relevant cases: constant and decreasing marginal susceptibility. If all types of individuals are equally likely, or if screening is randomly allocated, the relative composition of screened types will remain constant as the screening rate increases. Panels (A) and (B) describe the case of constant susceptibility with low and high screening coverage, respectively. A flat solid line describes how marginal susceptibility varies with current screening coverage. The dotted line depicts the average susceptibility of women that screen for a given coverage given by a gray rectangle. Since marginal susceptibility is

constant, the average susceptibility does not change as the gray screening coverage area expands. In other words, the composition of screened individuals is identical to the proportions of types in the population regardless of the screening rate. Similarly, the proportion of unscreened aggressive and non-aggressive cancers will be constant.

However, if AC individuals are more likely to undergo screenings than NA and NC individuals, the composition of screened individuals will vary as the screening rate increases. When the screening rate is low, the proportion of AC individuals among screened patients may be higher than the population proportion. Finally, as the screening rate approaches 100%, the proportion of AC individuals among screened patients may decline and reach the population proportion. As a result, mammography screenings will have diminishing returns in detecting AC individuals.

Panels (C) and (D) of Figure 2 describe the case of decreasing marginal susceptibility at low and high screening coverage, respectively. In contrast to Panels (A) and (B), the average susceptibility of screened women, depicted by the dotted line, decreases as screening coverage increases. A natural empirical test for decreasing returns could be measuring the proportion of AC type amongst the screened women as the screening coverage changes. Unfortunately, cancer type is not directly observed in the data. Instead, we utilize various measures of cancer progression to proxy for the cancer type.

--- Insert Figure 3 here ---

Figure 3 (Panel A) shows an inverse relationship between the proportion of late-stage cancer among cancers detected and the screening rate from 2003-2013. Specifically, the proportion of late-stage cancer detected declines as the screening rate increases. Figure 3 (Panel B) plots the five-year mortality rates among individuals detected with cancer in a particular year and the corresponding annual screening rates from 2003-2009. Again, we see an inverse relationship: average breast cancer mortality amongst screened women decreases as the screening coverage increases. There are two possible explanations for this fact. First, if treatment effectiveness does not depend on coverage and all types of individuals are equally likely to undergo screenings, we should observe a horizontal line. The decreasing line supports self-selection consistent with Panels

(C) and (D) of Figure 2. Second, more extensive coverage may lead to a higher percentage of early cancer detections and, thus, more effective treatment and lower mortality.

To account for the latter mechanism and isolate diminishing returns, we restrict our attention to cancers detected early, resulting in similar treatment effectiveness. Figure 3 (Panel C) shows that the mortality rate following early-stage cancer detection decreases as screening coverage increases. If AC and NA individuals were equally likely to undergo screening, the proportion of aggressive cancers amongst those early detected should not depend on coverage. On the contrary, we find higher mortality when the coverage is low, indicating that a high proportion of early-detected cancers prove aggressive. As coverage increases, a decrease in mortality indicates that the proportion of aggressive cancers decreases. Our results imply that AC individuals have a higher screening propensity than NA individuals. Since screening is the most beneficial for aggressive cancers detected early, a decreasing proportion of these cases suggests that the coverage has diminishing returns.

Importantly, analysis using Panel C allows for the possibility that early detection changes the course of the disease if the inherent efficacy of treatment of early-detected aggressive cancers does not depend on contemporary screening coverage. In our setting, this possibility is unlikely since the screening coverage does not exhibit a systematic time trend (see Figure 1).

In the next section, we present a stylized model of cancer progression and screening that we use to identify the percentage of women susceptible to aggressive cancer and screening selection function.

4 Model

The country is composed of N individuals indexed by n. Time is indexed by t and is divided into monthly intervals. Time $t = 0$ denotes the first month after turning 40. An individual n can have one of the three cancer states at time t, denoted by $c_n^t \in \{S_0, S_1, S_2\}$, where S_0, S_1 and S_2 denote "no breast cancer," "early-stage breast cancer" and "late-stage breast cancer," respectively. Every individual is assumed to have no breast cancer at $t = 0;\ that$ is, $c_n^0 = S_0$.

Everyone is born with one of the three latent types $\bar{\alpha}_n \in \{\text{NC}, \text{NA}, \text{AC}\}$, with probabilities p_{NC}, p_{NA} and p_{AC} (which sum to 1). NC (no cancer) individuals never develop breast cancer; NA (non-aggressive cancer) individuals can contract early-stage but not late-stage cancer; and AC (aggressive cancer) individuals can contract early-stage cancer that can progress to the late stage. In each period, NA individuals may become AC individuals with probability κ. Thus, individuals not susceptible to late-stage cancer may become susceptible at an older age. This also implies that non-aggressive tumors may become aggressive. We denote the type of the individual at t as α_n^t. Note that $\alpha_n^0 = \bar{\alpha}_n$.

Cancer stages and individual types follow a Markov chain. Individual n may obtain early-stage breast cancer with probability $P(c_n^{t+1} = S_1 | c_n^t = S_0, \alpha_n^t)$. We assume that an individual must contract early-stage cancer before progressing to late-stage cancer, so

$$P(c_n^{t+1} = S_2 | c_n^t = S_0, \alpha_n^t) = 0\ , \forall \alpha_n^t. \tag{1}$$

As mentioned before, an NC individual will not contract breast cancer so

$$P(c_n^{t+1} = S_1 | c_n^t = S_0, \alpha_n^t = \text{NC}) = 0. \tag{2}$$

Also, an NA individual does not contract late-stage cancer, so

$$P(c_n^{t+1} = S_2 | c_n^t = S_0, \alpha_n^t = \text{NA}) = 0. \tag{3}$$

Additionally, we assume that

$$P(c_n^{t+1} = S_1 | c_n^t = S_0, \alpha_n^t = \text{NA}) = P(c_n^{t+1} = S_1 | c_n^t = S_0, \alpha_n^t = \text{AC}) = \lambda_{01} \tag{4}$$

and denote

$$P(c_n^{t+1} = S_2 | c_n^t = S_1, \alpha_n^t = \text{AC}) = \lambda_{12}. \tag{5}$$

As a result, the cancer transition process is fully characterized by two numbers λ_{01} and λ_{12}. The λ_{12} specifies the distribution of, so-called adjourn time – the time cancer spends between pre-clinical and clinical stages. Adjourn time is distributed as a geometric random variable with the mean $\frac{1}{\lambda_{12}}$. We allow λ_{12} (and thus sojourn time) to differ by age. We estimate two parameters λ_{12}^Y for younger individuals and λ_{12}^O for older individuals. Average sojourn time is given as $E[\tau -$

$t|c_n^t = S_1]$, where the stage transition stopping time $\tau = \min\{t: c_n^t = S_2\}$. Note that the average sojourn time can be decomposed into sojourn time for AC and NC individuals at time t,

$$\begin{aligned} E[\tau - t|c_n^t = S_1] \\ &= E[\tau - t|c_n^t = S_1, \alpha_n^t = AC].P(\alpha_n^t = AC) \\ &+ E[\tau - t|c_n^t = S_1, \alpha_n^t = NA].P(\alpha_n^t = NA). \end{aligned} \tag{6}$$

Since NA individuals must become AC to transition to the late stage, we know that AC sojourn times are shorter:

$$E[\tau - t|c_n^t = S_1, \alpha_n^t = AC] \leq E[\tau - t|c_n^t = S_1, \alpha_n^t = NA]. \tag{7}$$

Also, $E[\tau - t|c_n^t = S_1, \alpha_n^t = NA] > 0$ if $\kappa > 0$.

Breast cancer is detected through screenings which are assumed to reveal c_n^t to the doctor and the individual. The decision to get screened is endogenous and depends on the type α_n^t, latent cancer stage c_n^t and the ethnic origin of the individual e_n. We assume that each period type transition occurs first, followed by the cancer stage transition and the decision to get screened. This way, individuals incorporate the most current type and cancer stage when deciding to get screened. In addition, we incorporate two ethnicities, *M* and *NM*, which denote Malayans and non-Malayans, respectively. As mentioned earlier, Malayans screen less than non-Malayans.

Let $d_n^t = 1$ if an individual gets screened at time t, and $d_n^t = 0$ otherwise. Denote the screening probability of an individual of type α, cancer stage c, and ethnic origin e, as

$$\delta_c^\alpha(e) = P(d = 1|c, \alpha, e). \tag{8}$$

We impose a set of functional form restrictions on the screening probability $\delta_c^\alpha(e)$. We allow type, cancer stage and ethnic origin to affect the probability of screening multiplicatively. We also allow screening to depend on a calendar year to capture the effects of marketing programs on the screening rate. Formally,

$$P_{Year}(d_n^t = 1|c_n^t, \alpha_n^t, e_n) = \gamma_1(c_n^t) \times \gamma_2(\alpha_n^t) \times \gamma_3(e_n) \times \gamma_4(Year). \tag{9}$$

We assume that everything else being equal, the screening probability with early-stage cancer is at least as high as the screening probability with no cancer, *i.e.,* $\gamma_1(c_n^t = S_1) \geq \gamma_1(c_n^t = S_0)$. We also assume that late-stage cancer is screened and detected instantaneously so that

$$P_{Year}(d_n^t = 1 | c_n^t = S_2, \alpha_n^t, e_n) = 1. \tag{10}$$

Next, we allow for self-selection to get screened and hence different screening propensities based on type α_n^t. Without a loss of generality, we normalize $\gamma_2(\alpha_n^t = \text{NC})$ to equal 1. We also allow $\gamma_2(\alpha_n^t = \text{NA})$ and $\gamma_2(\alpha_n^t = \text{AC})$ to be different from 1 and each other. We conjecture that an individual at risk may get screened more, that is:

$$\gamma_2(\alpha_n^t = \text{AC}) \geq \gamma_2(\alpha_n^t = \text{NA}) \geq 1. \tag{11}$$

We do not make the above assumption in the estimation. Instead, we let the data reveal the above ranking.

Without loss of generality, we normalize $\gamma_3(e_n = NM)$ to equal 1. We allow the screening rate among Malays to be different from that among non-Malays, that is, $\gamma_3(e_n = M)$, which is different (possibly smaller) than 1. Finally, we allow for a constant multiplicative trend in screenings, the surge in screenings in 2004 and 2006, and the slump in 2009. Formally,

$$\gamma_4(Year) = \tau_1^{Year-1993} \times \begin{cases} \tau_2 & \text{if } Year = 2004 \\ \tau_3 & \text{if } Year = 2006 \\ \tau_4 & \text{if } Year = 2009 \\ 1 & \text{otherwise.} \end{cases} \tag{12}$$

There can be three outcomes from a screening: no cancer detected, early-stage cancer detected, late-stage cancer detected. If no cancer is detected, nothing changes, and the individual is returned to the screening pool. Such an individual can decide to get screened again at a later period and her cancer stage c_n^t and type α_n^t may change in the future. If cancer is detected, an individual is assigned to a treatment regime for the rest of her life. We do not explicitly model cancer progression in the treatment regime. Instead, we impose a hazard rate of death due to cancer if the

individual is of type *AC*. This rate is ρ_1 if the stage of cancer is S_1, or ρ_2 if the stage of cancer is S_2 at the point of detection. We do not need to specify the hazard rate for individuals of type *NC* because they never test positive in a mammogram. The hazard rate for individuals of type *NA* is zero. However, even *NA* individuals (but not *NC* individuals) can eventually die of breast cancer because type *NA* individuals may progress to type *AC*. In this case, regardless of being in the treatment regime or not, former *NA* individuals would continue as *AC* individuals. For individuals not in the treatment regime, this would mean an increase in the hazard rate of contracting late-stage cancer from zero to λ_{12}. Similarly, this would mean an increase in the death hazard rate changes from zero to zero to ρ_c, where c was the cancer stage at the initial detection for the individuals in the treatment regime.

The model is estimated using a maximum likelihood estimator. The parameters to be estimated are:

$$\theta = (\kappa, \lambda_{01}, \lambda_{02}^{Y}, \lambda_{02}^{O}, \gamma_1, \gamma_2, \gamma_3, \tau_1, \tau_2, \tau_3, \tau_4, \rho_1, \rho_2, p_{NA}, p_{NC}, p_{AC}). \quad (13)$$

The estimation corrects that some people drop out from the sample because they pass away for reasons unrelated to breast cancer using death hazard rates from the census. We consider that one of the reasons why someone does not screen or die of breast cancer is that she could already be dead for reasons unrelated to breast cancer.

For everyone, we observe screening except for two cases: (i) if the screening happened before 2002 and was negative, (ii) if the screening happened outside BSS and was negative. Both issues are addressed in the estimation, and the details are contained at the end of the next section.

Figure 4 illustrates the transition probabilities for types of *NC*, *NA*, and *AC*. The probability of getting screened is denoted as δ and the probability of contracting cancer is denoted by λ. The left panel shows that *NC* individuals have a positive probability of getting screened, but zero probability of progressing to early-stage cancer. The middle panel shows that *NA* individuals can progress to early-stage cancer, but they never progress to late-stage cancer or die of cancer. Incidences of *NA* testing positive are the crucial component in our definition of overdiagnosis. Lastly, the right panel shows that *AC* individuals have a positive probability of transitioning from early-stage to late-stage cancer. Suppose an AC individual develops cancer, depending on the stage, the screening may either result in early-stage detection (followed by a death hazard ρ_1), or late-stage detection (followed by a death hazard ρ_2).

--- Insert Figure 4 here ---

5 Estimation details

In this section, we explain the estimation procedure. Beyond time-invariant ethnicity, each individual is characterized by a state $\omega_n^t = (\alpha_n^t, d_n^t, c_n^t, m_n^t, e_n)$, where m_n^t is equal to 1 if the individual died due to breast cancer, and 0 otherwise. The state is not fully observed in each period, and we denote the observable part of the state (an observation) for individual n at time t as x_n^t. For each n and t, we observe the screening decision d_n^t, cancer mortality m_n^t and ethnicity e_n. However, we do not observe cancer stage $c_n^t = \{S_0, S_1, S_2\}$ unless the individual gets screened, that is if $d_n^t = 1$. We also do not directly observe type α_n^t. We only know that an individual is *AC* if late-stage cancer is detected, and that individual is either *AC* or *NA* if early-stage cancer is detected. The individual can be of any type if no cancer is detected.

For every parameter vector θ, and under the initial conditions

$$\omega_n^0 = (\alpha_n^0 = \bar{\alpha}_n, d_n^0 = 0, c_n^0 = S_0, m_n^0 = 0, e_n = e),$$

the model implies a Markov process over ω_n^t for every $t = 1, \dots, T_n$. The kernel of that process is given by $P(\omega_n^t | \omega_n^{t-1}, \theta)$. The log-likelihood function of the data $x_n^1, \dots, x_n^{T_n}$ is given by:

$$\sum\nolimits_n \log \sum\nolimits_{\bar{\alpha}_n} P(\bar{\alpha}_n) P\left(x_n^{T_n}, \dots, x_n^1 \middle| \omega_n^0, \theta\right).$$

This joint probability can be computed recursively using the fact that the latent process ω_n^t is Markov. The procedure is standard when estimating hidden Markov processes and is briefly described here for completeness.

First, denote all states ω_n^t that are consistent with observing x_n^t as $\Omega(x_n^t)$. The likelihood computation is explained in two steps. First, we obtain distributions of latent states ω_n^t for each t, conditional on past and current observed information $x_n^t, x_n^{t-1}, \dots, x_n^1$ using the recursive equations

$$P(\omega_n^t|x_n^t,\dots,x_n^1,\omega_n^0,\theta)$$
$$= \begin{cases} \dfrac{P(\omega_n^t|x_n^{t-1},\dots,x_n^1,\omega_n^0,\theta)}{P(\omega_n^t \in \Omega(x_n^t)|x_n^{t-1},\dots,x_n^1,\omega_n^0,\theta)} & \text{if } \omega_n^t \in \Omega(x_n^t) \\ 0 & \text{otherwise} \end{cases}$$

and

$$P(\omega_n^{t+1}|x_n^t,\dots,x_n^1,\omega_n^0,\theta) = \sum\nolimits_{\omega_n^{t+1}} P(\omega_n^{t+1}|\omega_n^t)P(\omega_n^t|x_n^t,\dots,x_n^1,\omega_n^0,\theta).$$

Next, we obtain a relevant conditional likelihood

$$P(x_n^t|x_n^{t-1},\dots,x_n^1,\omega_n^0,\theta) = \sum\nolimits_{\omega_n^t \in \Omega(x_n^t)} P(\omega_n^t|x_n^{t-1},\dots,x_n^1,\omega_n^0,\theta)$$

which yields a joint likelihood

$$P\left(x_n^{T_n},\dots,x_n^1\middle|\omega_n^0,\theta\right) = P\left(x_n^{T_n}\middle|x_n^{T_n-1},\dots,x_n^1,\omega_n^0,\theta\right) \times$$
$$\vdots$$
$$P(x_n^t|x_n^{t-1},\dots,x_n^1,\omega_n^0,\theta) \times$$
$$\vdots$$
$$P(x_n^1|\omega_n^0,\theta)$$

There are a few other details of the estimation procedure that are omitted from the above presentation. First, for some observations the type of cancer is unobserved, that is, the cancer stage is marked as unknown. We assume that the event of an undetermined screening is random and exogenous, the likelihood of detecting an unknown stage is a joint likelihood of early and late-stage cancer. We use ρ_U as the hazard death rate due to breast cancer after an undetermined screening is observed.

Second, the above description omits the Markov chain's three-dimensional sub-state with exogenous transitions: calendar year, current age, and a 0-1 variable indicating death from reasons unrelated to breast cancer. The calendar year is used to obtain screening probabilities which depend on the current year. The current age is used in conjunction with age-dependent death hazard rates (from the Singapore census) to obtain the probability that the patient may be dead for reasons

unrelated to breast cancer for each t. When computing this probability, we also account for the fact that observed screenings must mean that the person is still alive. If the person is dead for reasons other than breast cancer, the patient would not get screened or die of breast cancer; thus, the likelihood of such events is appropriately adjusted.

Third, we do not observe negative screening events with probability ϕ, because some screenings are administered outside of BSS. Formally, the event of a negative screening in the data arrives with probability $\phi\delta_0^\alpha$ instead of δ_0^α, where $\phi \leq 1$. We account for this in the estimation, by calibrating ϕ to 40%, which was obtained from a survey administered by BSS. Consequently, we replace δ_0^α with $0.4\ \delta_0^\alpha$ when computing the likelihood of negative screening. Importantly, we observe all positive screenings; thus, δ_1^α and δ_2^α do not need to be deflated. We also do not observe negative screenings before 2002, and consequently, we set $\phi = 0$ for t before the year 2002.

6 Results

In this section, we present the result of the model, discuss its goodness of fit, and conduct counterfactuals.

6.1 Model estimates

We fit the model described in the previous section to match individual medical records. Such procedure fits stylized facts from Figure 3 to deliver the estimates of the proportion of types in the population and the degree of self-selection to screen conditional on the underlying type.

Our analysis yields two major findings. The first major finding is confirmation that the empirical proportion of three types of individuals in the population is non-trivial.[7] Our estimates indicate that the overall proportions of types are 54% (*NC*), 23% (*NA*), and 22% (*AC*). Each year, the probability of an *NA* type becoming an *AC* type is about 1%. Thus, there are more individuals susceptible to aggressive cancer among older women. Among the 40-year-old women, there are 27% and 19% of *NA* and *AC* types. The proportion of AC types grows to 22% for 50-year-olds and 24% for 60-year-olds. Even though fewer young women are susceptible to developing late-stage cancer, our analysis confirms earlier findings that the cancers of susceptible younger women

[7] We estimate that 45% of women have a positive likelihood of developing breast cancer at some point in their lives; however, we also estimate that 10% to 12% of women actually develop breast cancer over the course of their lives.

have shorter sojourn times (similar findings reported by Tabar, Fagerberg, Chen, Duffy, Smart, Gad, and Smith, 1995). The likelihood of stage-1 cancer growing to stage-3 within a year is 92% for *AC* women aged 40-45 and 76% for *AC* women aged 46+. These hazard rates imply average population sojourn times of 1.6, 2.1. 1.9 years for stage 1 cancers (1.1, 1.3, and 1.3 for *AC* individuals) for 40, 50, and 60-year-old women. The sojourn times initially increase because the likelihood of cancer growth among AC women decreases with age (λ_{12}^{Y} vs. λ_{12}^{O}). However, later in life, sojourn times decrease because of the higher likelihood of being *AC* type.[8]

The second major finding is that the screening rate differs significantly among the three types. In other words, individuals self-select into a screening regime based on their own type. Results indicate that every two years, *AC* individuals are 34 times more likely, and *NA* individuals are 8 times more likely than *NC* to undergo screening (*NC*:*NA*:*AC* is 1:8:34). This implies that *AC* individuals are approximately 4 times more likely to undergo screening than *NA* individuals.

--- Insert Table 2 here ----

6.2 Model fit

We validate our model by comparing the predicted number of cancer incidences vis-a-vis the actual number for several cohorts of women. We have cancer incidence data from 1993 but screening data only from 2002. Thus, we restrict our goodness-of-fit analysis to individuals who were at most 40-49 years old in 2002. We fit the predicted cancer incidence rates by stage to the ten cohorts of patients within this age group. We find that the predicted numbers of late-stage cancer fit the actual number very well for all but one outlier cohort. As for early-stage incidences, the quality of fit varies. The model underpredicts for the first five cohorts by about 20%. However, predictions for the subsequent four cohorts are well within ±/-10% of actual incidences. The fit of our model by cohort is presented in Figure 5.

--- Insert Figure 5 here ----

[8] In the Appendix B, we consider alternative assumptions regarding the relationship between age and hazard rates λ_{12}.

6.3 Counterfactuals

Heterogeneity in the rate of cancer progression and the propensity to screen results in the varying composition of screened types as the screening rate increases. Our estimates imply that at a 100% screening rate, the composition of the types is 54% (*NC*), 23% (*NA*), and 22% (*AC*). Compositions of screened types for intermediate coverages are obtained using counterfactual experiments produced by our model. These experiments aim to isolate the impact of increasing coverage on the efficacy of screening and the degree of overdiagnosis.

The screening policies are obtained in the following way. The relevant population is all Singapore women of more than 40 years of age in 2002. This population is tracked starting at 40 and is initially subdivided into latent subpopulations *NC*, *NA*, and *AC* according to the estimated initial proportion of types. To obtain the set of women screening for a given x% coverage rate, we draw x% of women without replacement maintaining a 1:8:32 ratio of the propensity to screen across types. Drawn women are flagged as "covered." The remaining women are flagged as "not covered." Covered women screen once every two years, while not covered women never screen. Given such screening policies for different coverage rates, we simulate the progression of types, cancer progression, and detection.[9]

--- Insert Figure 6 here ---

--- Insert Table 3 here ---

Figure 6 presents the composition of types at various screening coverages. Due to self-selection, at low screening rates, a large proportion of screened individuals are *AC*. However, as the screening rate increases, more *NA* and *NC* individuals decide to undergo screenings. For example, at a 25% screening rate, the ratio of types (*NC*:*NA*:*AC*) is 9:27:64; at a 50% screening rate, the ratio is 19:37:44; and at a 75% screening rate, the ratio is 39:31:30.

To measure the degree of overdiagnosis, we estimate the number of cancer incidences by the types of individuals. In Table 3, columns 2-4, we report the estimated detection rates of early-stage cancer, decomposed into non-aggressive (*NA*) and aggressive cancers (*AC*) and late-stage cancers

[9] Since the coverage status is determined for each woman at the age of 40, it remains to specify what happens to coverage when the type of women changes. In the baseline specification, we redraw the coverage status for all women that change types using updated propensity to the screen. We also consider two boundary conditions: keeping coverage status constant regardless of changing the type, or marking the individual as "covered" after every type change. As presented in the Appendix A, our results are robust to this assumption.

(*AC*) for various screening rates. Out of 100,000 individuals, a 25% screening rate will detect 56 incidences of early-stage cancer, 26 of which are in *AC* individuals; a 50% screening rate will detect 110 incidences, 32 of which are in *AC* individuals; and a 75% screening rate will detect 128 incidences, 31 of which are in *AC* individuals. Therefore, when the screening rate exceeds 50%, there is an insignificant increase in the detection rate of early-stage cancer in *AC* individuals.[10]

We define overdiagnosis at a given screening rate as the ratio of non-aggressive cancers to the total incidence rate of cancer (both early and late-stage). Therefore, the degree of overdiagnosis can be estimated using the predicted cancer incidences by stage and type. As a result of self-selection, the proportion of screened NA individuals increases as the screening rate increases. Consequently, the degree of overdiagnosis increases as the screening rate increases. In the last column of Table 3, we report the estimated degrees of overdiagnosis by screening rate. We see that even at low screening rates, there is some overdiagnosis. For instance, at a 5% screening rate, 110 cases are detected per 100,000 women, among which 4.5 cases are NA. In other words, 4.5 cases are overdiagnosed. As the screening rate increases to 25%, the degree of overdiagnosis increases to 21%. At a 50% screening rate, overdiagnosis increases to 43%. And, at a 75% screening rate, the degree of overdiagnosis increases even more to 50%.

Mammography screening will only help AC individuals, who account for 22% of the population. With a perfect targeting technology, a 22% screening rate could screen all *AC* individuals. On average, with current targeting technology, increasing the screening rate from 50% to 55% (extra 2,000 individuals screened) decreases late-stage occurrences by merely 0.7 cases per 100,000 women. The decrease is computationally negligible beyond 70% coverage.

Increasing screening coverage is more beneficial for older individuals. In columns 6-8, we present the number of late-stage cancer occurrences as a function of coverage for different age groups. Increasing the screening rate from 50% to 55% has a negligible impact (0.3 fewer cases per 100,000) on late-stage cancer occurrences for individuals between 40 and 49 years old. This change in screening coverage results in 1 fewer and 1.1 fewer late-stage cases for individuals 50-59 and 60-69 years old. Thus, increasing the screening rate beyond 50% is nearly three times more

[10] The individuals' type may change over time, and we report the type at the initial cancer detection. Consequently, the overall number of AC types recorded at the initial detection would decrease as coverage increases because more NA individuals are detected with early cancer before they progress to AC type. This decrease is also the reason why the number of AC early-stage detections decreases beyond 45% coverage.

effective for older individuals. At the same time, increasing screening rates beyond 75% have negligible benefits regardless of age. The last row of the table presents a simulation that evaluates an increase in screening frequency. At 50% coverage, moving from screening every two years to once a year has measurable benefits of decreasing late-stage cancer incidence by approximately 20 cases. The change is present for every age group.

In the estimation, we set the same distribution of initial types across ethnic groups. However, our identification of the self-selection parameters does not depend on this assumption. Instead, identification may be obtained using a diff-in-diffs variation in screening propensities within each type (detailed in Section 3). In Appendix C, we investigate the robustness of our results to heterogenous type distribution across ethnic groups. We re-estimate the model and re-simulate our counterfactuals allowing the Malayan population to have different baseline susceptibility to aggressive cancer. Our estimates imply that screening is slightly more effective if we assume that the Malayan population is significantly more susceptible to aggressive cancer. Nevertheless, even if heterogeneity across types is assumed to be extensive (+10pp higher susceptibility of the Malay population), our analysis confirms diminishing returns and muted impact of screening when increasing the coverage beyond 50% of the population.

7 Conclusion

Individuals have varying susceptibilities to developing aggressive breast cancer. Early screenings are critical for those more susceptible, but screenings may not be as beneficial for those with low or no susceptibility. In this paper, we propose a simple model to estimate the probabilities of individuals with various levels of susceptibility to aggressive cancer. We use a detailed data set from the Singapore Cancer Registry and the Health Promotion Board of Singapore. In addition, we exploit the variation in screening tendencies across various ethnic groups and temporal spikes in screening rates across years due to a government screening program.

Our data show that individuals with greater susceptibility are more likely to screen early. Therefore, as the population screening rate increases, the composition of individuals being screened and the effectiveness of screening change. We find that increasing the screening rate up to 50% is effective, while increasing the screening rate beyond 50% is, on average, ineffective in preventing late-stage cancer incidences and results in overdiagnosis. While we can detect small

but measurable benefits of increasing screening rates beyond 50% only for individuals beyond 60 years of age, increasing screening coverage beyond 75% has no measurable benefits regardless of age. Conversely, increasing screening frequency from every two years to yearly has measurable benefits regardless of age. Finally, with perfect targeting technology, only 22% of women must undergo screenings. Until such technology is available, a screening rate of up to 50% will likely save lives.

These results have policy implications. American screening rate reached above 70%. At this coverage in 2008 alone, Bleyer and Welch (2012) estimated the degree of overdiagnosis to be around 70,000. Such findings questioned the widespread promotion of mammography screenings for breast cancer prevention. Nonetheless, empirical evidence of overdiagnosis and related harm has not been strong enough to reject mammography as a routine screening strategy outright. Therefore, a recent IARC (International Agency on Research on Cancer) committee concluded that regular screening mammography should still be recommended for women 50-69 (Lauby-Secretan, Scoccianti, Loomis, Benbrahim-Talla, Bouvard, Bianchini, and Straif, 2015).

Our results caution against applying marginal screening effectiveness obtained at high coverage to cases with moderate or low coverage. In countries with high screening rates, overdiagnosis is undoubtedly a significant concern. However, not all countries have such high rates. For instance, the screening rate in Singapore was estimated to be just below 25% in 2010. At this rate, overdiagnosis is only modest; thus, efforts by the Health Promotion Board of Singapore to promote regular mammography screening through their Breast Screen Singapore program were effective in reducing mortality. The same reasoning holds when designing mammography promotion across segments with varying coverage. If the coverage varies for reasons plausibly orthogonal to genetic cancer susceptibility (such as income or possibly ethnicity), increasing coverage for underserved populations will prevent cancer deaths, even if the overall population coverage is high.

References

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.

BITLER, M. P., AND C. S. CARPENTER (2016): "Health insurance mandates, mammography, and breast cancer diagnoses," *American Economic Journal: Economic Policy*, 8(3), 39–68.

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.

EINAV, L., A. FINKELSTEIN, T. OOSTROM, A. OSTRIKER, AND H. WILLIAMS (2020): "Screening and selection: The case of mammograms," *American Economic Review*, 110(12), 3836–70.

FINKELSTEIN, A., S. TAUBMAN, B. WRIGHT, M. BERNSTEIN, J. GRUBER, J. P. NEWHOUSE, H. ALLEN, K. BAICKER, AND O. H. S. GROUP (2012): "The Oregon health insurance experiment: evidence from the first year," *The Quarterly journal of economics*, 127(3), 1057–1106.

KIM, H. B., AND S.-M. LEE (2017): "When public health intervention is not successful: Cost sharing, crowd-out, and selection in Korea's National Cancer Screening Program," *Journal of health economics*, 53, 100–116.

KOWALSKI, A. E. (2018): "Behavior within a clinical trial and implications for mammography guidelines," Discussion paper, National Bureau of Economic Research.

_____ (2021): "Mammograms and Mortality: How Has the Evidence Evolved?," *Journal of Economic Perspectives*, 35(2), 119–40.

LANNIN, D. R., H. F. MATHEWS, J. MITCHELL, M. S. SWANSON, F. H. SWANSON, AND M. S. EDWARDS (1998): "Influence of Socioeconomic and Cultural Factors on Racial Differences in Late-Stage Presentation of Breast Cancer," *JAMA*, 279(22), 1801–1807.

LAUBY-SECRETAN, B., C. SCOCCIANTI, D. LOOMIS, L. BENBRAHIM-TALLAA, V. BOU-vARD, F. BIANCHINI, AND K. STRAIF (2015): "Breast-cancer screening—viewpoint of the IARC Working Group," *New England journal of medicine*, 372(24), 2353–2358.

LI, C. I., K. E. MALONE, AND J. R. DALING (2003): "Differences in Breast Cancer Stage, Treatment, and Survival by Race and Ethnicity," *Archives of Internal Medicine*, 163(1), 49–56.

LOY, E. Y., D. MOLINAR, K. Y. CHOW, AND C. FOCK (2015): "National Breast Cancer Screening Programme, Singapore: evaluation of participation and performance indicators," *Journal of medical screening*, 22(4), 194–200.

MITCHELL, J., D. R. LANNIN, H. F. MATHEWS, AND M. S. SWANSON (2002): "Religious beliefs and breast cancer screening," *Journal of Women's Health*, 11(10), 907–915.

OSTER, E. (2020): "Health recommendations and selection in health behaviors," *American Economic Review: Insights*, 2(2), 143–60.

ROSENBAUM, L. (2014): "Invisible Risks, Emotional Choices–Mammography and Medical Decision Making," *The New England journal of medicine*, 371(16), 1549–1552.

SINGAPORE DEPARTMENT OF STATISTICS (2011): "Census of population 2010 statistical release 2: Households and housing," .

SINGAPORE MINISTRY OF HEALTH (2002): "National Health Surveillance Survey 2001," .

_____ (2011): "National Health Survey 2010," .

SMITH-BINDMAN, R., D. L. MIGLIORETTI, N. LURIE, L. ABRAHAM, R. B. BARBASH, J. STRZELCZYK, M. DIGNAN, W. E. BARLOW, C. M. BEASLEY, AND K. KER-LIKOWSKE (2006): "Does utilization of screening mammography explain racial and ethnic differences in breast cancer?," *Annals of internal medicine*, 144(8), 541–553.

TABAR, L., G. FAGERBERG, H.-H. CHEN, S. W. DUFFY, C. R. SMART, A. GAD, AND R. A. SMITH (1995): "Efficacy of breast cancer screening by age. New results swedish two-county trial," *Cancer*, 75(10), 2507–2517.

VAN DEN ENDE, C., A. M. OORDT-SPEETS, H. VROLING, AND H. M. VAN AGT (2017): "Benefits and harms of breast cancer screening with mammography in women aged 40–49 years: A systematic review," *International journal of cancer*, 141(7), 1295–1306.

ZACKRISSON, S., I. ANDERSSON, L. JANZON, J. MANJER, AND J. P. GARNE (2006): "Rate of over-diagnosis of breast cancer 15 years after end of Malmo¨ mammographic screening trial: follow-up study," *Bmj*, 332(7543), 689–692.

Appendix

A Changing coverage status

Depending on her initial type, the screening coverage status is determined for each woman at 40 years of age. However, it remains to specify what happens to the coverage status when the type of women changes. In the baseline specification, we redraw the coverage status for all women that change types using updated screening propensity. We also consider two alternative assumptions. In alternative scenario 1, we keep the coverage status constant regardless of the changing type. Because the type may only change from NA to AC, which could increase screening propensity, this scenario constitutes a lower bound on screening effectiveness. We also consider scenario 2, in which we always mark the individual as covered after changing the type. Scenario 2 constitutes an upper bound on screening effectiveness. The results of these alternative simulations are presented in Table 4. The alternative specifications deliver quantitatively similar results to the baseline specification. Thus, we conclude that our results are robust.

--- Insert Table 4 here ---

B Age-dependent cancer progression hazard rates

In the model, we allow the early-state to late-stage cancer hazard rates to depend on the woman's age. This dependence allows our model to generate inverse U-shaped cancer sojourn types as a function of age previously identified in the literature. We estimate two numbers λ_{12}^{Y} for younger individuals and λ_{12}^{O} for older individuals. The baseline specification allows for the change in the hazard rate as the woman reaches 45 years old. In this Appendix, we investigate alternative

assumptions. Table 5 contains simulations using a re-estimated model assuming homogeneous hazard rates and age breaks at 43 and 47 instead of 45. The results are nearly identical.

--- Insert Table 5 here ----

C Heterogeneous susceptibility propensities

In the baseline model we calibrate the baseline type propensities at birth to be equal across ethnic groups. We investigate the robustness of our findings to this assumption by calibrating the model using heterogeneous type propensities; that is, we set

$$\mathrm{P}(\mathrm{AC}|\mathrm{Malay}) = \mathrm{P}(\mathrm{AC}|\mathrm{Non-Malay}) + X \qquad \mathrm{P}(\mathrm{NA}|\mathrm{Malay}) = \mathrm{P}(\mathrm{NA}|\mathrm{Non-Malay}) - X.$$

Subsequently, we recompute the simulated composition of screened types and the probability of overdiagnosis. The results are shown in Table 6. In all cases, including extreme heterogeneity in baseline rates, we find that *AC* individuals are significantly more likely to screen than *NC* and *NA* individuals. This self-selection results in an increasing rate of overdiagnosis as screening coverage increases. We estimate that at 25%, coverage overdiagnosis ranges from 18% to 28%. At 50% coverage, the overdiagnosis ranges from 30% to 56%. Regardless of the assumed heterogeneity in the composition of types across ethnicities, we find that increasing screening coverage up to 50% is effective in preventing late-stage cancer. We also confirm that increasing screening coverage beyond 50% is less effective on average but may be somewhat effective for older individuals.

--- Insert Table 6 here ----

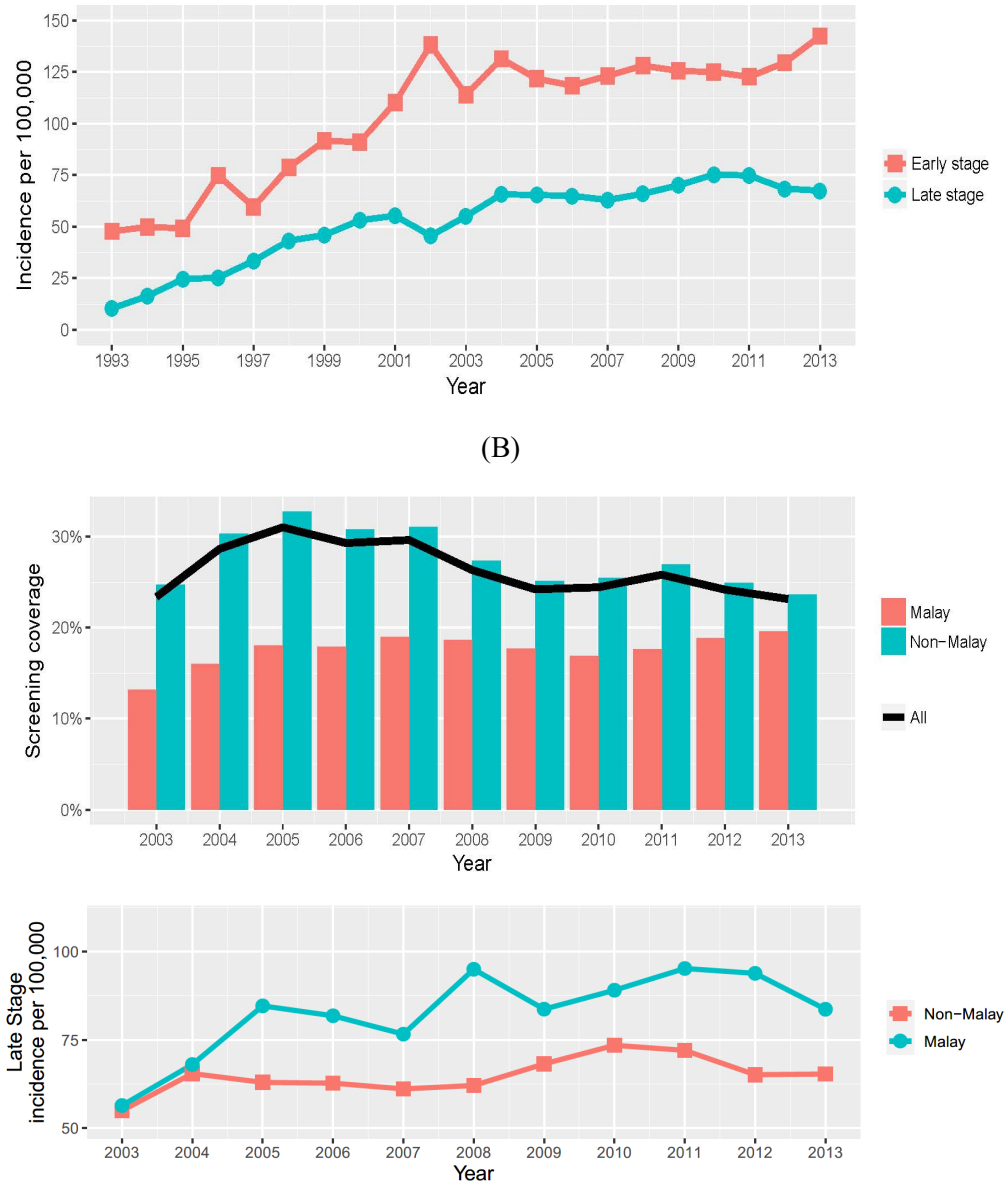

Figure 1: (A) Trends in aggregated cancer detection by stage. (B) Trends in aggregated screening rates for Malays and non-Malays. (C) Trends in late-stage cancer incidence for Malays and non-Malays.

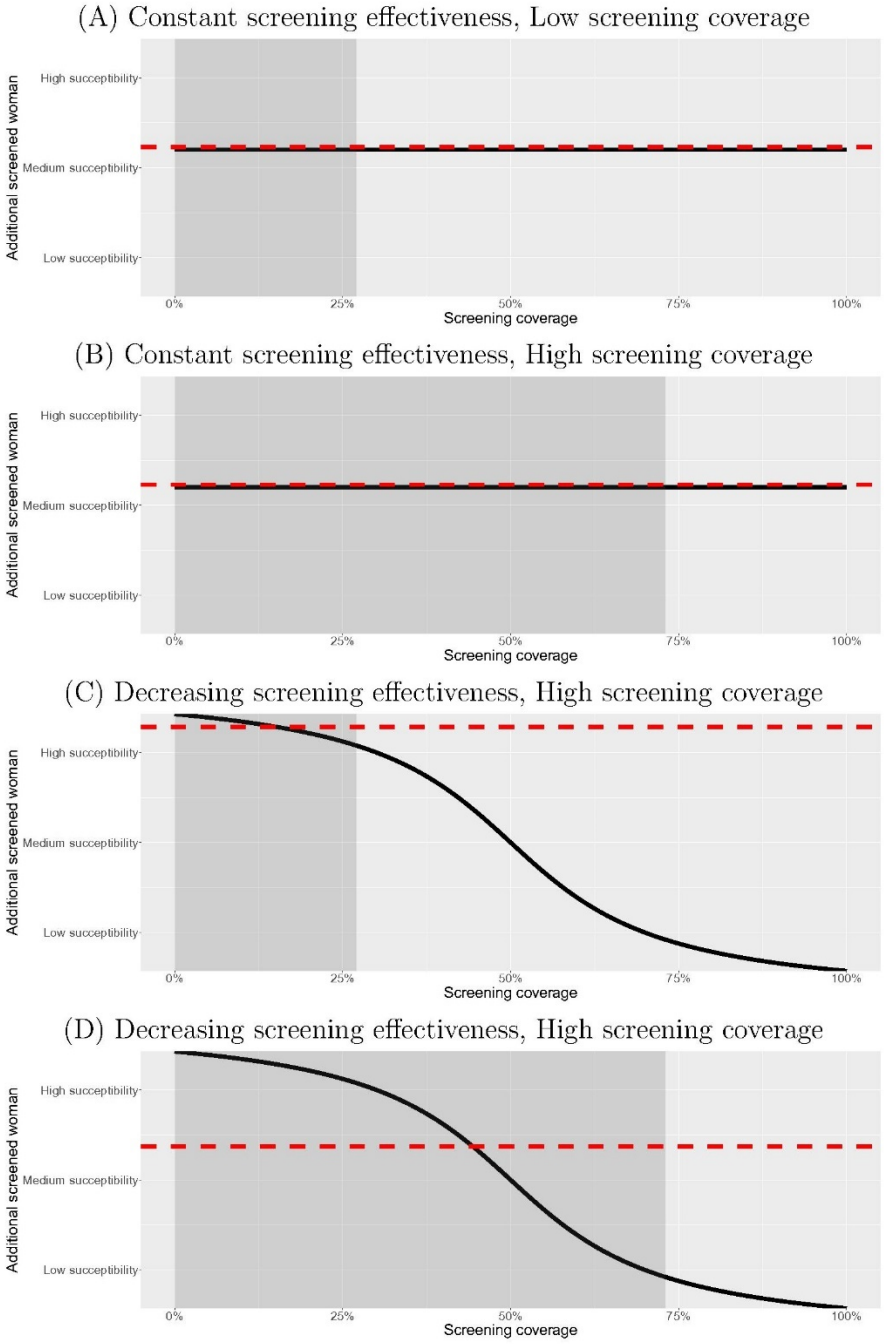

Figure 2: Composition of screened types

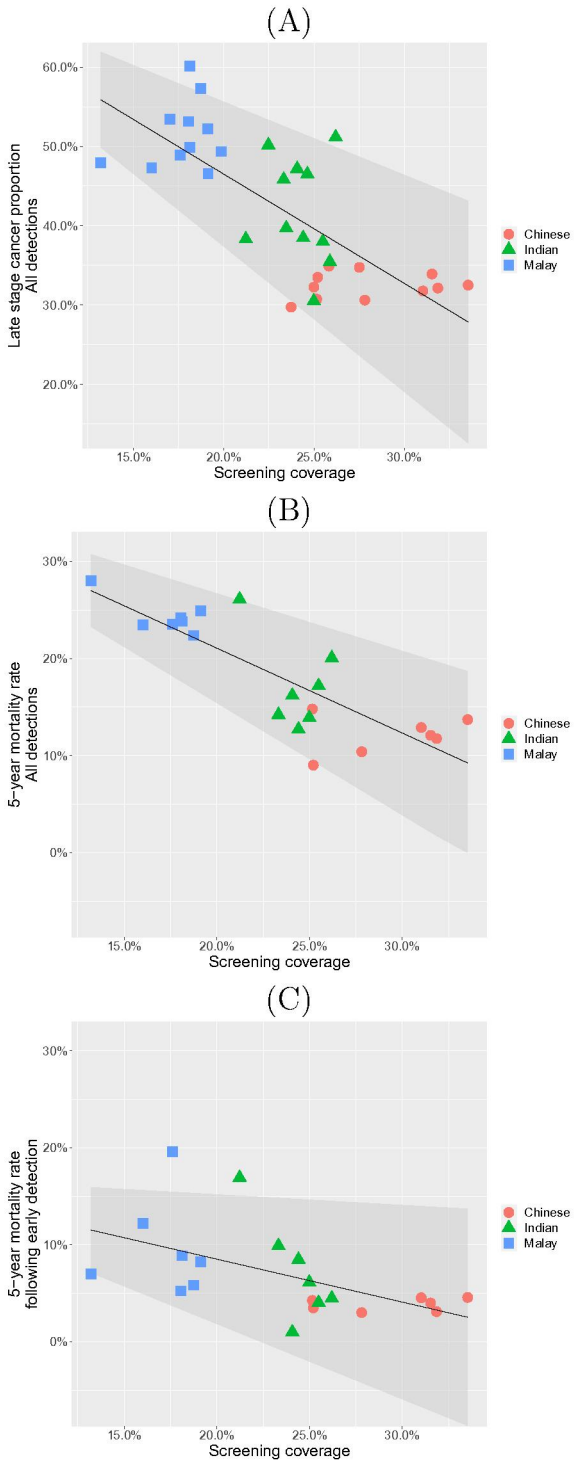

Figure 3: (A) Screening rate and proportions of late-stage cancer. (B) Screening rate and five-year mortality following cancer detection. (C) Screening rate and five-year mortality following early (stage 1 and unknown stage) cancer detection. The gray area represents a 95% confidence interval. The downward sloping upper bound signifies a statically significant negative relationship with *p*-value of less than 0.05.

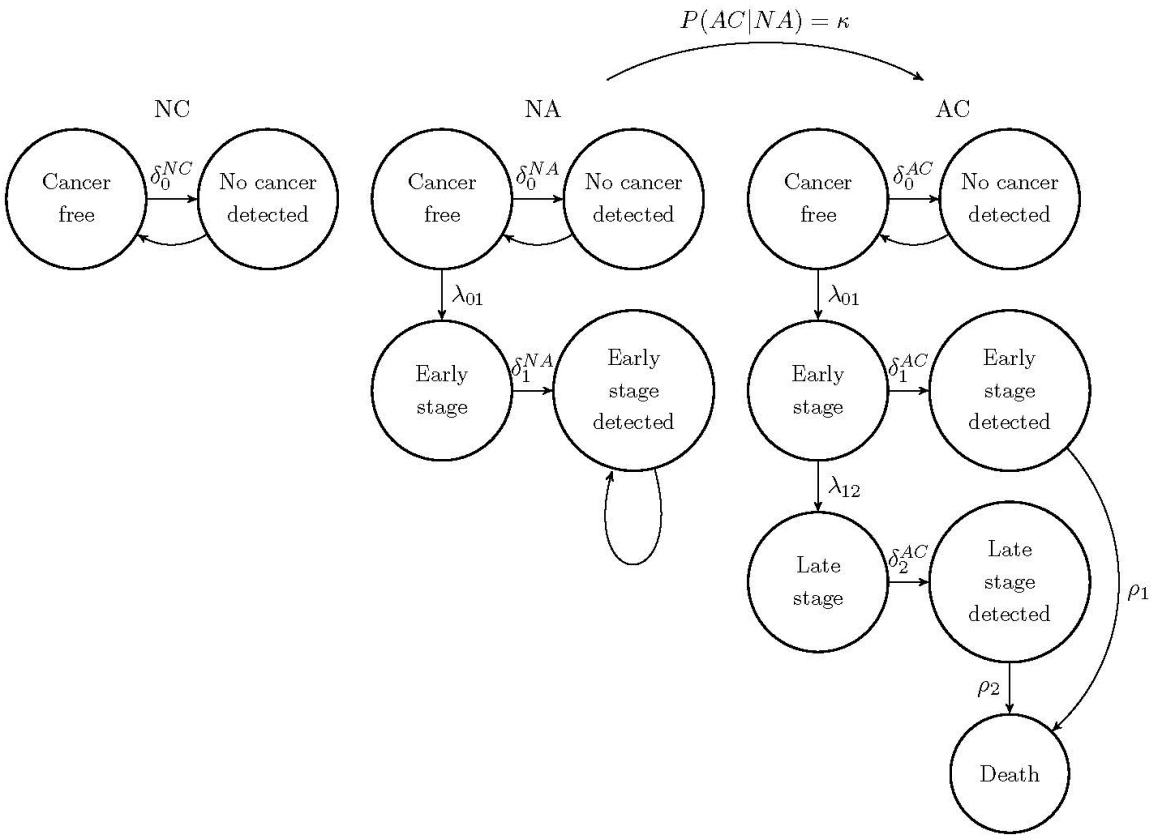

Figure 4: Transition probabilities of cancer stages for NC, NA, and AC individuals.

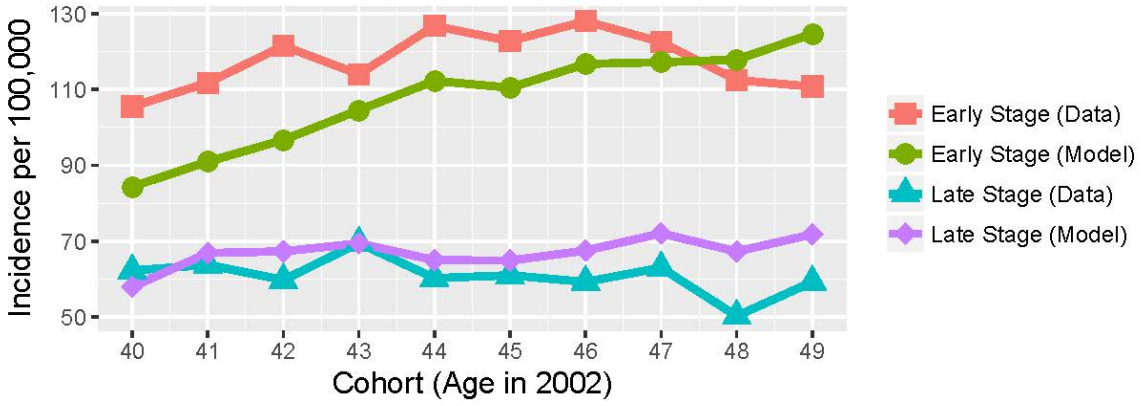

Fig. 5. Fit of the model by cohort.

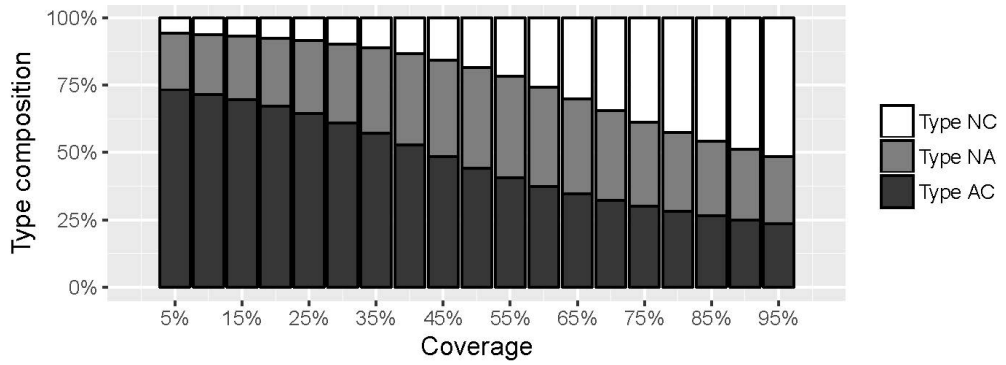

Fig 6: Composition of screened types

	Total	Chinese	Malay	Indian	Others
Ethnic Composition (%)	100.0	74.1	13.4	9.2	3.3
Age: Below 15 Years (%)	17.4	15.7	22.4	21.5	22.8
15 - 64 Years (%)	73.7	74.2	71.5	72.8	73.7
65 Years & Over (%)	9.0	10.1	6.1	5.7	3.5
Average Monthly Household Income ($)	7,214	7,326	4,575	7,664	11,518
Education: Below Secondary (%)	32.4	33.8	37.0	22.5	10.8
Secondary (%)	18.9	18.2	27.1	17.2	9.9
Post-Secondary (%)	11.1	9.9	19.2	11.2	7.6
Diploma (%)	14.8	15.5	11.6	14.1	13.3
University (%)	22.8	22.6	5.1	35.0	58.4
Religion: Christianity (%)	18.3	20.1	0.7	12.8	57.6
Buddhism/Taoism (%)	44.2	57.4	0.2	0.8	20.8
Islam (%)	14.7	0.4	98.7	21.7	9.2
Hinduism (%)	5.1	-	0.1	58.9	0.8
Other Religions (%)	0.7	0.37	0.1	4.6	1.1
No Religion (%)	17.0	21.8	0.2	1.1	10.6

Table 1: Socio-economic differences between ethnic groups in Singapore in 2010. Source Singapore Department of Statistics (2011).

		Estimate	Std. Error
Proportion of NC individuals at birth	p_{NC}	0.539	0.0068
Proportion of NA individuals at birth	p_{NA}	0.272	0.0062
Proportion of AC individuals at birth	p_{AC}	0.189	0.0018
Transition prob. 'No Cancer' to 'Early Stage'	λ_{01}	0.00037	0.00004
Transition prob. 'Early Stage' to 'Late Stage' (>45yo)	λ_{12}^{Y}	0.190	0.0140
Transition prob. 'Early Stage' to 'Late Stage' (≤45yo)	λ_{12}^{O}	0.113	0.0070
Probability of Screening, given 'No Cancer'	$\gamma_1(c_n^t = S_0)$	0.00076	0.00007
Probability of Screening, given 'Early Stage'	$\gamma_1(c_n^t = S_1)$	0.00076	0.00489
Screening probability adjustment factor for NA vs NC	$\gamma_2(\alpha_n^t = NA)$	9.144	0.3862
Screening probability adjustment factor for AC vs NC	$\gamma_2(\alpha_n^t = AC)$	51.400	2.2646
Screening probability adjustment factor for Malays	$\gamma_3(\alpha_n^t = M)$	0.644	0.0033
Time trend	τ_1	1.029	0.0102
Trend 2004	τ_2	1.357	0.0067
Trend 2006	τ_3	1.236	0.0061
Trend 2012	τ_4	0.807	0.0044
Hazard rate of death, given 'Early Stage'	ρ_1	0.002	0.0001
Hazard rate of death, given 'Late Stage'	ρ_2	0.003	0.0001
Hazard rate of death, given 'Unknown Stage'	ρ_U	0.002	0.0001

Table 2. Parameter estimates – all transition and screening probabilities are month-to-month. Standard errors are calculated using standard MLE sandwich formula.

	Cancer detections per 100,000 (first diagnosis)						
Screening coverage	Early-stage NA (A)	Early-stage AC (B)	Late-stage AC (C)	Late-stage AC 40y-49y	Late-stage AC 50y-59y	Late-stage AC 60y-69y	Mortality vs 5% coverage per 100,000
5%	3.8	7.4	98.8	83.4	102.9	112.5	-
10%	7.7	14.5	92.8	80.0	95.8	104.2	-0.4
15%	11.9	21.4	87.1	76.5	89.2	96.6	-0.8
20%	16.6	27.6	82.0	73.7	83.0	89.3	-1.3
25%	22.2	33.7	77.2	70.9	77.1	82.8	-1.7
30%	28.3	38.9	73.4	68.6	72.6	77.7	-2.0
35%	35.5	43.4	70.3	66.7	69.0	73.7	-2.3
40%	43.1	46.9	68.2	65.6	66.5	70.6	-2.6
45%	50.7	49.7	66.7	64.9	64.8	68.2	-2.8
50%	57.9	51.9	65.7	64.5	63.8	66.6	-3.0
55%	63.8	53.6	64.9	64.2	63.1	65.4	-3.1
60%	68.3	54.9	64.4	64.0	62.5	64.7	-3.2
65%	70.8	55.6	64.1	63.8	62.1	64.3	-3.3
70%	72.0	56.0	63.9	63.8	61.9	64.1	-3.3
75%	72.4	56.0	63.9	63.8	61.9	64.0	-3.3
80%	72.5	56.1	63.9	63.8	61.9	64.0	-3.3
85%	72.5	56.1	63.9	63.8	61.9	64.0	-3.3
90%	72.5	56.1	63.9	63.8	61.9	64.0	-3.3
95%	72.5	56.1	63.9	63.8	61.9	64.0	-3.3
Screening every year (double frequency)							
50%	59.2	73.4	45.9	47.5	43.1	44.1	-4.7

Table 3: Composition of screened types, detection and overdiagnosis.

		Cancer detection per 100,000 (first diagnosis)				
	Screening coverage	Early stage NA (A)	Early stage AC (B)	Late stage AC (C)	Late stage AC 60y-69y	Mortality vs 5% coverage per 100,000
Constant screening regime	25%	22.2	28.9	81.6	90.6	-1.9
	50%	57.9	50.0	67.4	69.4	-3.5
	75%	72.3	56.0	63.9	64.0	-4.0
Stochastic screening regime	25%	22.2	33.7	77.2	82.8	-1.7
	50%	57.9	51.9	65.7	66.6	-3.0
	75%	72.4	56.0	63.9	64.0	-3.3
"Always" screening regime	25%	22.2	35.7	75.4	79.9	-1.7
	50%	57.9	51.9	65.7	66.6	-2.8
	75%	72.4	56.0	63.9	64.0	-3.2

Table 4. Composition of screened types, detection, and mortality -- changing types do not start screening.

		Cancer detection per 100,000 (first diagnosis)				
	Screening coverage	Early stage NA (A)	Early stage AC (B)	Late stage AC (C)	Late stage AC 60y-69y	Mortality vs 5% coverage per 100,000
Homogeneous	25%	21.7	34.2	76.9	84.9	-1.9
	50%	57.6	52.8	65.0	68.8	-3.4
	75%	73.2	57.2	63.1	65.8	-3.8
40-43yo	25%	22.1	34.0	77.5	82.7	-1.8
	50%	57.8	52.3	66.0	66.3	-3.1
	75%	72.3	56.4	64.2	63.8	-3.4
40-45yo regime	25%	22.2	33.7	77.2	82.8	-1.7
	50%	57.9	51.9	65.7	66.6	-3.0
	75%	72.4	56.0	63.9	64.0	-3.3
40-47yo	25%	22.3	33.6	76.4	83.1	-1.6
	50%	58.1	51.8	64.9	67.1	-3.0
	75%	72.6	55.9	63.1	64.4	-3.2

Table 5. Composition of screened types, detection, and mortality -- changing types always

	Screening coverage	Cancer detection per 100,000 (first diagnosis) Early stage NA (A)	Early stage AC (B)	Late stage AC (C)	Late stage AC 60y-69y	Mortality vs 5% coverage per 100,000
	25%	23.3	30.9	90.2	89.2	-1.3
X=-0.1	50%	38.9	43.4	80.9	77.6	-2.2
	75%	39.3	43.5	80.8	77.5	-2.3
X=-0.05	25%	22.6	28.7	87.7	87.5	-0.4
	50%	37.2	40.4	78.9	76.9	-0.8
	75%	37.7	40.6	78.8	76.7	-0.8
X=-0.01	25%	23.3	32.8	77.4	81.6	-1.5
	50%	56.9	50.1	66.2	66.0	-2.8
	75%	64.9	52.4	65.2	64.6	-3.0
X=-0.005	25%	22.5	33.1	77.5	82.4	-1.6
	50%	57.2	50.7	66.1	66.6	-2.9
	75%	67.7	53.8	64.8	64.7	-3.2
	25%	22.1	34.0	76.7	82.5	-1.7
X=-0.001	50%	57.7	52.3	65.0	66.3	-3.0
	75%	71.4	56.2	63.4	63.8	-3.3
	25%	22	34.1	77.2	83.2	-1.7
X=0.001	50%	58.1	52.6	65.4	67.0	-3.0
	75%	73.7	57.0	63.6	64.1	-3.3
	25%	20.6	38.7	75.2	83.7	-2.4
X=0.005	50%	57.7	58.3	62.6	66.6	-4.2
	75%	80.8	64.8	59.9	62.5	-4.7
	25%	17.9	45.7	75.3	85.8	-3.5
X=0.01	50%	54.2	66.3	61.8	67.5	-6.0
	75%	86.3	74.9	58.4	62.0	-6.5
	25%	15.8	62.3	65.7	80.1	-6.1
X=0.05	50%	50.0	80.8	54.1	63.8	-9.6
	75%	88.3	90.4	50.9	58.5	-10.4
	25%	21.6	62.8	58.3	75.1	-5.6
X=0.1	50%	63.3	78.9	49.9	62.3	-8.7
	75%	107.2	90.0	46.2	56.3	-9.4

Table 6. Composition of screened types, detection, and mortality for different baseline rates.