Appendix:

Analysis of mammography screening schedules under varying resource constraints for planning breast cancer control programs in low- and middle- income countries: a mathematical study

Shifali Bansal[1], BS, Vijeta Deshpande[1], MS, Xinmeng Zhao[1], BS, Jeremy A. Lauer[2], PhD, Filip Meheus[3], PhD, André Ilbawi[2], MD, Chaitra Gopalappa[1], PhD

Table of Contents

A. Two-step Markov process (TSMP) methodology for parametrization of the natural onset and progression of cancer

Parameterization of a cancer natural history model consists of estimation of three sets of parameters that vary by age: a) onset rates- the rates of transitioning from healthy to carcinoma in-situ (CIS); b) progression rates- the rates of transitioning between preclinical disease stages in the absence of diagnosis, and c) diagnostic rates- the current rates of diagnosis in the absence of intervention. Though there are multiple mathematical models presented in the literature for parameterization of natural history models, most are applied to HICs and are based on the use of longitudinal data from cancer registries (1), (2), (3), (4) or population-based screening studies. (5), (6) The pre- and post- screening data provide references for the estimation process. Data that are usually available for most LMICs are only the nationally representative annual rates of cancer incidence and mortality, i.e., the numbers of newly diagnosed cases of cancers and deaths per 1000 women, estimated through the Global Cancer Observatory. (7), (8) There are usually no data on how people are diagnosed, which could vary according to population-specific parameters, such as population's awareness and knowledge in recognizing symptoms and access to health care, in addition to disease-specific parameters such as occurrence of symptoms. Therefore, in this study, we used a new two-step Markov process methodology developed specifically for parameterization of cancer progression models in LMICs where longitudinal cancer registry databases are not available. (9)

The TSMP method uses as inputs, country-specific incidence estimates by age, which are publicly available through Global Cancer Observatory, and country-specific stage at diagnosis distributions, which were obtained from studies in the literature. Among the three sets of parameters needed for the model (discussed above), we assumed that the second sets of rates, progression rates, are disease-specific and do not vary by country, and used estimates from models applied to high-income countries that were presented in the literature (see Table S3). We then estimated onset rates and diagnostic rates specific to Peru using the TSMP method. Technical details of the theory and proofs of the TSMP are presented in (9) and its application for the cost-effectiveness analysis of the 'Best Buy' interventions, for breast cancer, cervical cancer, and colorectal cancer for updating the Appendix 3 of the NCD Global Action Plan (10) (11) are presented in (12). In these previous work, the model in (9) was applied to sub-Saharan Africa and Southeast Asia regions using data from 2008 to 2012, when there was not much screening in these regions. However, in the case of Peru, certain populations underwent screening prior to 2012. Therefore, we modified the model in (9) to consider this difference, which we discuss here. For completeness, we first present the earlier version of the model formulation before discussing the modifications specific to Peru.

A.1. Overview of the two-step Markov process method for parametrization of natural history model specific to LMICs

The TSMP divides the estimation of population-specific onset rates of disease and diagnostic rates into two steps, each defined by a Markov process model but with different state spaces. In the first step, we define the disease onset and progression as a discrete-time Markov process $\boldsymbol{X} = \{X_t; t \geq 0, \Omega, \mathbb{P}\}$ with a collapsed state space $\Omega = \{[H_a], [U_a], [D_a]\}$ representing age a and health states H_a =healthy, U_a= undiagnosed, and D_a= diagnosed, without differentiating between disease stages; and $\mathbb{P}$ is the transition probability matrix. We assume that $\mathbb{P}$ is heterogeneous by age, i.e., the probabilities for disease onset, diagnosis, and stage progression were modeled as a function of age. Therefore, the size of the state space is 300, 3 health states time 100 ages. We estimate age-specific onset rates using an iterative analytical model derived using the Markov chain.

In the second step, we estimated diagnostic rates in each stage of cancer, i.e., transition rates from preclinical to clinical states ($d_{i,a}$), by using a simulation-based optimization of the Markov process $\boldsymbol{Y} = \{Y_t; t \geq 0, Z, \mathbb{Q}\}$, with state space $Z = \{[H_a], [U_{i,a}], [D_{i,a}]\}$, which is an expansion of the state space in equation (1) to include stage $i \in \{0 = CIS, 1 = Local, 2 = Regional, and\ 3 = Distant\}$ and age a; and rate matrix $\mathbb{Q}$, which corresponds to the flow diagram in Figure S1.

We discuss each of these steps below.

A.1.1. Estimation of disease onset rates

We use a two-step Markov process for estimation of disease onset rates and diagnostic rates. In this first step, for estimation of the onset rates, we define disease onset and progression as a discrete-time Markov process,

$$\boldsymbol{X} = \{X_t; t \geq 0, \Omega, \mathbb{P}\} \tag{1}$$

with a collapsed state space $\Omega = \{[H_a], [U_a], [D_a]\}$ representing age a and health states H_a =healthy, U_a= Undiagnosed, and D_a= diagnosed, (see Figure S1 for a flow diagram, and Table S1 for a list of notations), without differentiating between disease stages; and $\mathbb{P}$ is the transition probability matrix. Then, using steady state Markov properties we can write

$$\pi_k = \sum_{j \epsilon \Omega,} \pi_j \, P_{jk} \;\; ; 0 \leq \pi_k \leq 1; \sum_{k \epsilon \Omega} \pi_k = 1 \tag{2}$$

where, P_{jk} are the probabilities of transitioning from state j to state k, i.e., elements of the matrix $\mathbb{P}$, and π_k are the elements of the steady-state distribution vector $\boldsymbol{\pi}$. Our prime element of interest in this Markov process is $P_{H_a U_a}$, the risk or probability of developing the disease in age a, i.e., an element of $\mathbb{P}$ representing the probability of transitioning from H_a to U_a. Using the standard definition of risk to rate conversion that assumes that the underlying distributions governing transition probabilities are exponential, the rate of disease onset in age group a can be written as $\theta_a = -\ln(1 - P_{H_a U_a})$. Based on the above structure of the Markov process we derived an analytical expression for estimation of $P_{H_a U_a}$ as

$$P_{H_a U_a} = \frac{I_{D_a} c_a - \sum_{k=0}^{a-1} \left(\pi_{H_k} P_{H_k U_k} \left[\sum_i s_i \left(1 - e^{-(a-k)\lambda_i} \right) - \sum_i s_i \left(1 - e^{-(a-1-k)\lambda_i} \right) \right] \left(\prod_{j=k:a+1} e^{-\mu_j} \right) \right)}{A_a \left[\sum_i s_i \left(1 - e^{-\lambda_i} \right) \right] (e^{-\mu a}) - I_{D_a} c_a} \tag{3}$$

and developed an iterative process for estimation of $P_{H_a U_a}$ starting with the lowest age. We present the notations and the iterative process for estimating of $P_{H_a U_a}$and eventually θ_a in Tables S1 and S2, respectively. The details of the derivation are presented elsewhere. (9)

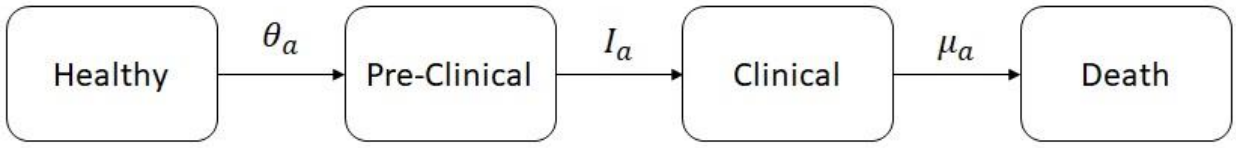


Figure S1: Flow diagram for the collapsed state space in the Markov process used for parameterization of disease progression

Table S1: Summary of notations for estimation of onset rate using algorithm in Table S2

Notation	Description
Model 1: $\boldsymbol{X} = \{X_t; t \geq 0, \Omega, \mathbb{P}, \boldsymbol{\pi}\}$	$\boldsymbol{X}$ is a Markov process with state space Ω, its underlying discrete time Markov chain given by the one-step transition probability matrix $\mathbb{P}$ and steady-state distribution vector $\boldsymbol{\pi}$.
$[H_a], [U_a], [D_a]$	Age-vectors representing states of healthy, pre-clinical disease (i.e., undiagnosed cancer state), and clinical disease (i.e., diagnosed cancer state), respectively, for age a
π_j	Element of vector $\boldsymbol{\pi}$ representing steady-state probability for state j
P_{jk}	Element of the matrix $\mathbb{P}$ representing one-step probability of transitioning from state j to k
$P_{H_aU_a} = 1 - e^{-\theta_a}$	$P_{H_aU_a}$ is the risk of developing disease at age a, and is defined as the one-step probability of transitioning from healthy to preclinical disease (H_a to U_a); θ_a is the rate of disease onset per person-year among persons in age a (used in Model 2)
T	T is a random variable denoting the time taken to transition to clinical disease state from the time of disease onset (sojourn time); $T \sim hyperexponential(\lambda_1, s_1, \ldots, \lambda_4, s_4)$, s_i is the probability that T will take the form of the exponential distribution with rate λ_i
S	S is a random variable denoting time of natural survival past the age at disease onset (i.e., the person does not die from any other disease during this time); $S \sim exponential(\mu_k)$
I_{D_a}	Cancer incidence, defined as the number of new cases of cancer diagnoses in age a each year divided by the number of people in age a
c_a	The proportion of the total population in age a
A_a	Among persons in age group a, the proportion in healthy state or pre-clinical disease states
$p_{i,a}$	Rate of progression from disease stage i to $i+1$ (also used in Model 2)
μ_a	Disease-free mortality rate at age a
Model 2: $\boldsymbol{Y} = \{Y_t; t \geq 0, \mathrm{Z}, \mathbb{Q}, \boldsymbol{\rho}\}$	$\boldsymbol{Y}$ is a continuous time Markov process with state space Z, generator matrix $\mathbb{Q}$, and steady state distribution vector $\boldsymbol{\rho}$.
$[H_a], [U_{i,a}], [D_{i,a}]$	Age-vector representing states of healthy, pre-clinical disease (i.e., undiagnosed cancer state), and clinical disease (i.e., diagnosed cancer state), respectively, at age a and cancer stage i
$\theta_a = -\ln(1 - P_{H_aU_a})$	Rate of disease onset in age a; (see Model 1 for $P_{H_aU_a}$)
$d_{i,a}$	Diagnostic rates, defined as the rates of transitioning from pre-clinical stage i to clinical-stage i per person-year for persons in age a
$p_{i,a}$	Rate of progression from disease stage i to $i+1$ (also used in Model 1)
μ_a	Disease-free mortality rate at age a
$\bar{\mu}_{i,a}$	Mortality rates when not on treatment and at disease stage i and age a
$\bar{\bar{\mu}}_{i,a}$	Mortality rates on treatment and at disease stage i and age a
I_a	Cancer incidence by age a
s_i	Proportion diagnosed in stage i in screening-naïve population

Table S2: Overview of the algorithm for computing age-specific onset rate of cancer

Initialize $\pi_{H_0} = A_0;\ \pi_{U_0} = 0$; and $P_{H_0U_0} = 0$; Set $a = 1$, the youngest age-group of cancer onset (we assumed age 15 for breast cancer).

Step 1: Calculate in-situ onset rate

$$P_{H_aU_a} = \frac{I_{D_a}c_a - \sum_{k=0}^{a-1}\left(\pi_{H_k}P_{H_kU_k}\left[\sum_i s_i\left(1-e^{-(a-k)\lambda_i}\right) - \sum_i s_i\left(1-e^{-(a-1-k)\lambda_i}\right)\right]\left(\prod_{j=k:a+1} e^{-\mu_j}\right)\right)}{A_a\left[\sum_i s_i(1-e^{-\lambda_i})\right](e^{-\mu_a}) - I_{D_a}c_a}$$

Where.
$\frac{1}{\lambda_i} = \sum_{j=0}^{i}\frac{1}{p_j}$; if p_j are a function of age at disease onset then $\frac{1}{\lambda_{i,a}} = \sum_{j=0}^{i}\frac{1}{p_{j,a}}$
Then, disease onset rate at age a is estimated as

$$\theta_a = -\ln\left(1 - P_{H_aU_a}\right)$$

Step 2: Calculate prevalence of healthy state:

$$\pi_{H_a} = \frac{A_a - \sum_{k=0}^{a-1}\left(\pi_{H_k}P_{H_kU_k}P(T\geq a-k)P(S\geq a-k)\right)}{1+P_{H_aU_a}};$$

$$P(T \geq a-k)P(S \geq a-k) = \sum_i s_i\left(1-e^{-(a-1-k)\lambda_i}\right)\prod_{j=k:a} e^{-\mu_j}$$

Where,
$\frac{1}{\lambda_i} = \sum_{j=0}^{i}\frac{1}{p_j}$; if p_j are a function of age at disease onset then $\frac{1}{\lambda_{i,a}} = \sum_{j=0}^{i}\frac{1}{p_{j,a}}$

Step 3: Increment a by 1; if a is less than the maximum age goes to step 1, else stop.

A.1.2. Estimation of diagnostic rates

In the second step of the two-step Markov process, for estimation of diagnostic rates, we reformulate the discrete-time Markov process $\boldsymbol{X}$**,** in previous section that defined disease onset and progression, into a continuous-time discrete-state Markov process $\boldsymbol{Y} = \{Y_t; t \geq 0, Z, \mathbb{Q}\}$, with more granular discretization of the state space as $Z = \{[H_a], [U_{i,a}], [D_{i,a}]\}$, for stage $i \in \{0 = CIS, 1 = Local, 2 = Regional, and\ 3 = Distant\}$ and age a, and rate matrix $\mathbb{Q}$. We estimated diagnostic rates in each stage of cancer, i.e., transition rates from preclinical to clinical states ($d_{i,a}$), by using a simulation-based optimization of the Markov process $\boldsymbol{Y}$**.**

The objective of the simulation-based optimization model is to minimize the sum of square errors between the simulated cancer incidence ($\bar{I}_a$) and the GLOBOCAN predicted incidence (I_a). (13) The details of the model are presented in (9), which were applied to sub-Saharan Africa and Southeast Asia regions using data from 2008 to 2012, when there was not much screening in these regions. However, in the case of Peru, certain populations underwent screening prior to 2012. Therefore, we modified the model in (9) to consider this difference, which we disucss here. For completeness, we first present the earlier version of the model formulation before discussing the modifications specific to Peru. The objective function was formulated as

$$Minimize_{d_{i,a}} \sum_a (\bar{I}_a - I_a)^2,\ d_{i,a} \geq 0,\ \forall i, a \qquad (4)$$

As the analytical form of $\bar{I_a}$ are unknown, we used a numerical optimization solution method where the objective function value can be evaluated numerically through simulation at any specific values of the decision parameters $d_{i,a} \geq 0$, $\forall i, a$. Here, for any specific $d_{i,a}$ values, we simulated the Markov Process $\boldsymbol{Y}$ over time t using $\boldsymbol{\rho}_{t+1} = \boldsymbol{\rho}_t + \boldsymbol{\rho}_t \mathbb{Q} \Delta t$ until it reached state steady, i.e.,

$$\boldsymbol{\rho} = \boldsymbol{\rho} + \boldsymbol{\rho} \mathbb{Q} \Delta t \tag{5}$$

where $\boldsymbol{\rho}$ is a vector of state distribution at steady state and $\mathbb{Q}$ is the rate matrix. We estimated $\bar{I_a}$ using $\bar{I_a} = \sum_i \rho_{U_{i,a}} d_{i,a}$, where $\rho_{U_{i,a}}$ is the steady state value for state $U_{i,a}$ (denoting the prevalence in pre-clinical cancer stage i at age a), which can be estimated by expansion of equation (5) as

$$\rho_{U_{i,a}} = \rho_{U_{i,a}} + \rho_{U_{i-1,a-1}} \lambda_{i-1,a} - \rho_{U_{i,a-1}} \left(\lambda_{i,a} + d_{i,a} + \mu_{i,a}\right) \tag{6}$$

In the previously presented model in (9), because of the assumption that diagnosis occurs only based on symptoms and that the probability of showing symptoms are higher in advanced disease stages, i.e., $d_{i,a} > d_{i-1,a}$, the distribution of the stage at diagnosis was a good approximation for the ratio of stage-specific diagnostic rates. That is, $\frac{d_{i,a}}{d_{3,a}} = \sum_{j=0}^{i} s_j$, where s_j is the proportion diagnosed in stage j, and $d_{i,a}$ is the diagnostic rate at state i and age a. Therefore, for the terms in the objective function in equation (4) we could write

$$(\bar{I_a} - I_a)^2 = \left(\sum_i \rho_{U_{i,a}} d_{i,a} - I_a\right)^2 = \left(\sum_i \rho_{U_{i,a}} \left(d_{3,a} \sum_{j=0}^{i} s_j\right) - I_a\right)^2 \approx f(d_{3,a}) \tag{7}$$

That is, the only unknown values in the objective function in equation (4) were the diagnostic rates in the last stage of cancer ($d_{3,a}$), as the steady state values in the pre-clinical states, $\rho_{U_{i,a}}$, are estimated numerically from the simulation of the Markov model in equation (5) as discussed above. The resulting objective function was

$$Minimize_{\ d_{3,a}} \sum_a \left(\sum_i \rho_{U_{i,a}} \left(d_{3,a} \sum_{j=0}^{i} s_j\right) - I_a\right)^2 \tag{8}$$

and the decision variables $d_{3,a} \forall a$ were solved iteratively for each a. However, in the case of Peru, certain populations have undergone screening based on recommendations prior to 2012 (the latest incidence data available at the time of this work was for year 2012), and thus, the assumption $d_{i,a} > d_{i-1,a}$ does not hold. Therefore, we modified the objective function in equation (7) to

$$Minimize_{\ d_{i,a}, \forall i,a} \sum_{i,a} \left(\rho_{U_{i,a}} \left(d_{i,a}\right) - I_a\right)^2, d_{i,a} \geq 0 \ \forall i, a \tag{9}$$

that is, the number of decision variables (the unknown values) now increase to include diagnostic rates $d_{i,a}$ for each stage i and age a, as the actual rates of screening currently occurring in the population are unknown. This creates many decision variables. As the number of decision variables increases, ascertaining the convergence of a solution algorithm to the global optima becomes more challenging. We address this by showing below that the optimization problem in equation (9) is separable both on i and a and thus equation (9) can be converted to ia number of sub-problems. Each sub-problem can then be solved separately but iteratively for $d_{i,a}$, iterating over each i and a (see below). We further test for the convexity of each sub-problem (see Appendix C).

Remark 1: We can rewrite equation (9) as,

$$Minimize_{\ d_{i,a}} \left(\rho_{U_{i,a}} \left(d_{i,a}\right) - I_a\right)^2, d_{i,a} \geq 0 \tag{10}$$

for each combination of i, a pair thus generating ia number of sub-problems. Each function can then be solved separately for $d_{i,a}$ but iteratively over age a starting from the youngest age and, within each age, iteratively over cancer state i starting with the earliest disease state.

Proof:

Using the expression for $\rho_{U_{i,a}}$, from the expansion of the Markov process in equation (6) discussed above, and multiplying by $d_{i,a}$ we can write

$$\rho_{U_{i,a}} d_{i,a} = \left[\rho_{U_{i,a}} + \rho_{U_{i-1,a-1}}\lambda_{i-1,a} - \rho_{U_{i,a-1}}\left(\lambda_{i,a} + d_{i,a} + \mu_{i,a}\right)\right] d_{i,a} \quad (11)$$

In equation (8), for $i = 0$ (the in-situ stage) $\lambda_{i-1,a-1} = \theta_{a-1}$ the cancer onset rate, and for all other values of i (i.e., local, regional, and distant stages) $\lambda_{i-1,a-1}$ are the progression rates (see Figure S1); and $\mu_{i,a-1}$ are the mortality rates. Values for $\lambda_{i-1,a-1}$ and $\mu_{i,a-1}$ are known. When $i = 0$ (the in-situ stage) $\rho_{U_{i-1,a-1}} = \rho_{H_{a-1}}$ denoting the steady state value in healthy (i.e., prevalence of healthy stage), and under all other values of i, $\rho_{U_{i-1,a-1}}$ are the steady state values in the pre-clinical states (i.e., prevalence of pre-clinical cancer stages). For any given i, a pair, from Remark 2 and its proof below, the steady state values for $\rho_{U_{i,a-1}}$ and $\rho_{U_{i-1,a-1}}$, and solution to $d_{i,a-1}$ are known. Therefore, for any value of $d_{i,a}$, the steady state value for $\rho_{U_{i,a}}$ can be calculated through simulation of the Markov process in equation (5). As such, the only unknown value in equation (11) will then be $d_{i,a}$.

This completes the proof.

Remark 2: If we iteratively solve for $d_{i,a}$ using equation (11) by iterating over a and, within each a, iterate over i, then, for any given i, a pair, the steady state values for $\rho_{U_{i,a-1}}$ and $\rho_{U_{i-1,a-1}}$, and the solution to $d_{i,a-1}$ are known. Thus, the only unknown term in equation (11) is $d_{i,a}$

Proof:

We prove this by applying mathematical induction on equation (11)

For i = 0, a = 1,

$$\rho_{U_{0,1}} d_{0,1} = \left[\rho_{U_{0,1}} + \rho_{H_0}\theta_1 - \rho_{U_{0,0}}\left(\lambda_{0,a} + d_{0,a} + \mu_{H,a}\right)\right] d_{0,1} \quad (12)$$

Then, the only unknown value is $d_{0,1}$ because $\rho_{U_{0,0}} = 0$ and ρ_{H_0} is the actual prevalence of healthy persons in age 0 (obtained from population demographics) as the first age for disease risk is 1, and all other parameters are known as discussed in proof of Remark 1.

Assuming the proof holds for $i = m$, $a = 1$,

for $= m + 1$, $a = 1$

$$\rho_{U_{m+1,1}} d_{m+1,1} = \left[\rho_{U_{m+1,1}} + \rho_{U_{m,0}}\lambda_{m,1} - \rho_{U_{m+1,0}}\left(\lambda_{m+1,1} + d_{m+1,1} + \mu_{m+1,1}\right)\right] d_{m+1,1} \quad (13)$$

Then, the only unknown parameter is $d_{m+1,1}$ as $\rho_{U_{m,0}} = 0$ and $\rho_{U_{m+1,0}} = 0$ as the first age of disease risk is 1.

For $= 0$, $a = 2$

$$\rho_{U_{0,2}} d_{0,2} = \left[\rho_{U_{0,2}} + \rho_{H_1}\theta_2 - \rho_{U_{0,1}}\left(\lambda_{i,2} + d_{i,2} + \mu_{i,2}\right)\right] d_{0,2} \quad (14)$$

Then, the only unknown parameter is $d_{0,2}$ because $\rho_{H_1} = \rho_{H_1} - \rho_{H_0}(\theta_1 + \mu_{H,1})$ can be estimated through steady state simulation of equation (5) and $\rho_{U_{0,1}}$ was estimated previously under $= 0$, $a = 1$.

Assuming the proof holds for $i = m + 1$, $a = 2$,

$$\rho_{U_{m+1,2}}\, d_{m+1,2} = \left[\rho_{U_{m+1,2}} + \rho_{U_{m,1}}\lambda_{m,2} - \rho_{U_{m+1,1}}\left(\lambda_{m+1,2} + d_{m+1,2} + \mu_{m+1,2}\right)\right] d_{m+1,2} \quad (15)$$

Then, the only unknown parameter is $d_{m+1,2}$ as $\rho_{U_{m,1}}$ and $\rho_{U_{m+1,1}}$ were estimated above under $= m$, $a = 1$ and $i = m + 1$, $a = 1$, respectively

Finally, assuming the proof holds for any i and $a = k$,

for any i , and $a = k + 1$

$$\rho_{U_{i,k+1}}\, d_{i,k+1} = \left[\rho_{U_{i,k+1}} + \rho_{U_{i-1,k}}\lambda_{i-1,k+1} - \rho_{U_{i,k}}\left(\lambda_{i,k+1} + d_{i,k+1} + \mu_{i,k+1}\right)\right] d_{i,k+1} \quad (16)$$

Then, the only unknown parameter is $d_{i,k+1}$ as $\rho_{U_{i-1,k}}$ and $\rho_{U_{i,k}}$ were estimated above under any i and $a = k$. This completes the proof.

A.2. Test for convexity of the optimization model for estimation of diagnostic rates

To check for the convergence of the solution to global optima we test for the convexity of the objective functions.

Specifically, we test for the commonly used convexity test, a function $f(x)$ that is twice differentiable on x is convex if it is positive semi-definite, i.e., the second derivative $f''(x) \geq 0$ at all points of x. However, we do not know the analytical form of $\bar{I}_{i,a}$ to calculate the second derivative of the objective function $\left(\bar{I}_{i,a} - I_{i,a}\right)^2$. Therefore, for each combination of cancer stage (i) and age (a) pair, we empirically generated the function for $\bar{I}_{i,a}$ by estimation at multiple points of $d_{i,a}$. See Figure S1 and Figure S2 for results on In-situ and Local stages of cancer and at multiple age groups.

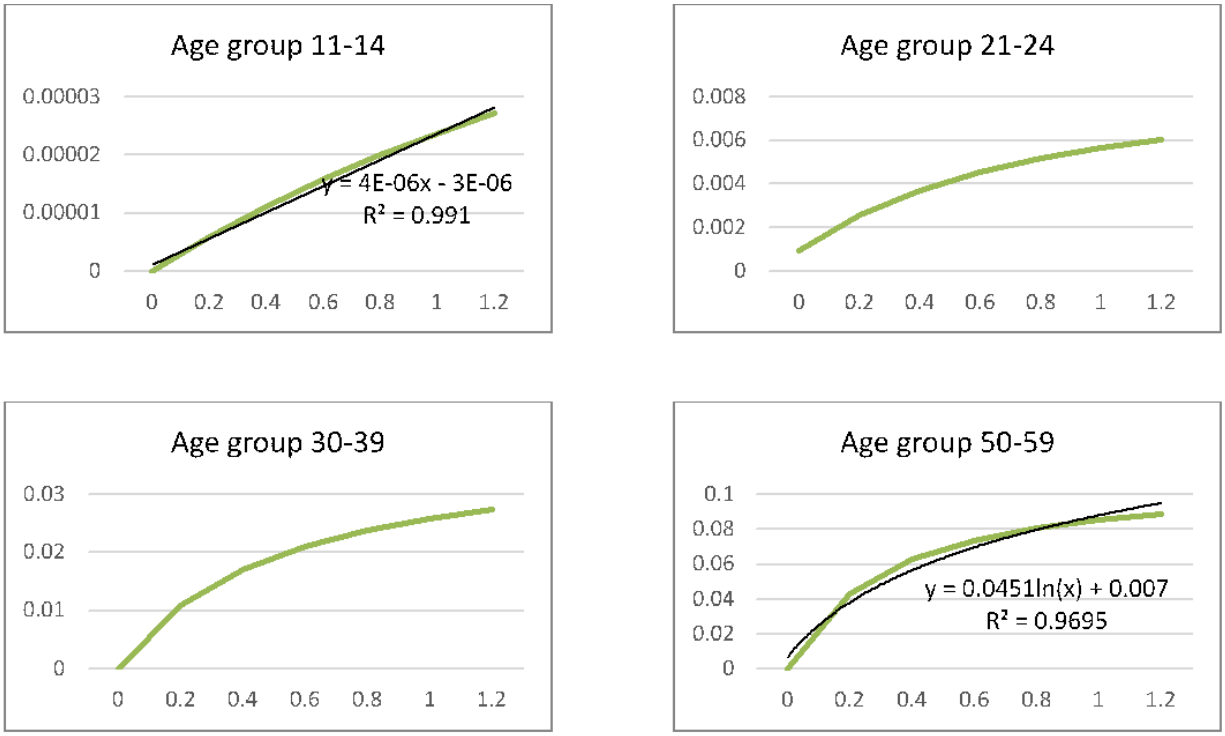


Figure S2: Incidence vs diagnostic rate for specific age-group and In-situ stage of cancer

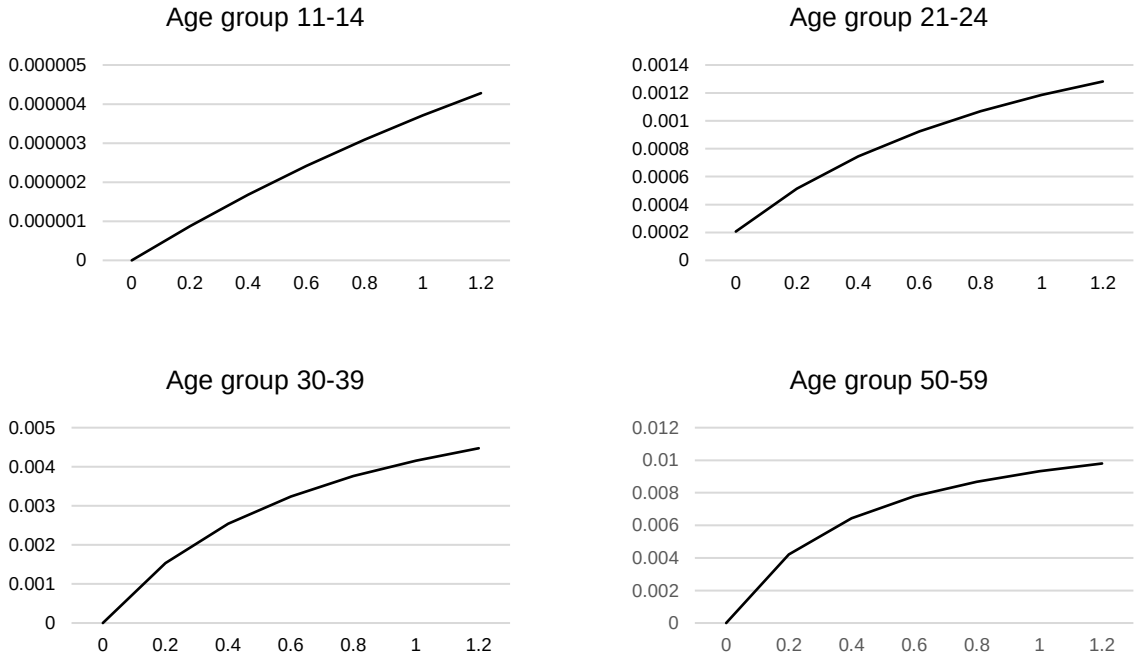


Figure S3: Incidence vs diagnostic rate for specific age-group and local stage of cancer

From the above empirical results, for any given cancer stage and age, the simulated incidence $\bar{I}_{i,a}$ is approximately a linear or a logarithmic function of diagnostic rates $d_{i,a}$, i.e.,

$\bar{I}_{i,a} \sim c\ ln(d_{i,a}) + b$ or $\bar{I}_{i,a} \sim c d_{i,a} + b$ for some constants c and b.

Writing $x = d_{i,a}$,

If $\bar{I}_{i,a} \sim c\ ln(x) + b$, the second derivative of the objective function $\left(\bar{I}_{i,a} - I_{i,a}\right)^2$ on x is

$$f''(x) = \frac{d^2}{dx^2}\left(c\ ln(x) + b - I_{i,a}\right)^2 = \frac{2\left(I_{i,a} - b - c\ ln(x) + c\right)}{x^2} > 0 \text{ as } I_{i,a} > b \quad (17)$$

And if $\bar{I}_{i,a} \sim cx + b$, the second derivative of the objective function $\left(\bar{I}_{i,a} - I_{i,a}\right)^2$ on x is

$$f''(x) = \frac{d^2}{dx^2}\left(cx + b - I_{i,a}\right)^2 = 2c(c) > 0 \quad (18)$$

thus, indicating that the objective function $\left(\bar{I}_{i,a} - I_{i,a}\right)^2$ is convex.

A.3. Data assumptions for parameterization of cancer onset and progression for Peru

Table S3 presents data specific to the US and Peru that were used for constructing cancer onset and progression models specific to these countries using the two-step Markov process methodology discussed above in A.1

Table S3: Region specific input data for parameterization

Parameters	Value	Reference
GENERAL PROGRESSION PARAMETERS		(14) (15) (16)
Progression rates (average over age)		
In-situ to Local ($p_{0,.}$)	0.19	
Local to Regional ($p_{1,.}$)	0.33	
Regional to Distant ($p_{2,.}$)	0.43	
Distant to Death ($p_{3,.}$)	0.50	
Annual mortality rate (per woman year) without treatment by stage at diagnosis (average over age)		
In-situ ($\bar{\mu}_{0,.}$)	0.08	
Local ($\bar{\mu}_{1,.}$)	0.14	
Regional ($\bar{\mu}_{2,.}$)	0.23	
Distant($\bar{\mu}_{3,.}$)	0.50	
Annual mortality rate (per woman year) with treatment by stage at diagnosis (average over age)		
In-situ ($\bar{\bar{\mu}}_{0,.}$)	0.01	
Local ($\bar{\bar{\mu}}_{1,.}$)	0.02	
Regional ($\bar{\bar{\mu}}_{2,.}$)	0.08	
Distant ($\bar{\bar{\mu}}_{3,.}$)	0.27	
Note: Here '.' denotes the age		
REGION-SPECIFIC DATA		
Pre-screening incidence per 1000 women per year (I_a)		(17) (18)
Age group	Peru	US
0-14	0.00	
15-19	0.00	0.00
20-24	0.01	
25-29	0.05	0.09
30-34	0.14	0.26
35-39	0.36	0.58
40-44	0.70	1.09
45-49	0.91	1.72
50-54	1.05	1.97
55-59	1.38	2.21
60-64	1.38	2.60
65-69	1.52	2.84
70-74	1.52	3.06
75-79	1.56	3.33

80-84	2.16	3.43	
85+	2.14		
Distribution of stage at diagnosis in base-case			(18) (19) (20) (21)
Stage	Peru	US	
In-Situ (s_0)	3%	4.70%	
Local (s_1)	43%	48.30%	
Regional (s_2)	45%	39.50%	
Distant (s_3)	9%	7.50%	

B. Markov decision process (MDP) to identify optimal screening schedules for mammography

B.1. Formulation of the problem of identifying optimal screening schedule as a MDP model

The parameterized cancer onset and progression model from section A was used in a Markov decision process model to identify an optimal screening schedule. Specifically, we formulated the problem as a finite-state, finite-horizon and discrete-time MDP defined by a 6-tuple $\{Y_t, D_t;\ Z, S, \mathbb{P}_s, R_s\}$, where $t = \{1,2,3\ldots,100\}$ are the decision-making stages; here, stages represent individual ages, i.e., $t = a$, therefore, for convenience, we will use 'age' to refer to the normally used terminology of 'stage' in MDP models, replacing t with a,

$Y_a \epsilon$ Z is the disease state at age a, defined over the state space $\mathrm{Z} = \{[H_a], [U_{i,a}], [D_{i,a}], M\}$, where $[H_a], [U_{i,a}], [D_{i,a}]$ are healthy, preclinical, and clinical states in disease stage $i \in \{0 = CIS, 1 = Local, 2 = Regional, and\ 3 = Distant\}$ and age a, as in the Markov process model in the previous section, and M denotes a mortality state,

S is the action space which is a set of possible decision choices, here we have 2 possible choices, i.e., $S = \{Screen(1), Do\ not\ Screen(0)\}$

$D_a \epsilon\ S$ is the decision taken at age a (choosing from set S),

$\mathbb{P}_s$ is the transition probability matrix corresponding to action s, specifically, each element $p(i', a, s, j)$ of the matrix $\mathbb{P}_s$ represents the probability of transitioning from state i' to state j if the person is at age a and action s is taken, and

R_s is the immediate reward matrix corresponding to action s, specifically, each element $r(i', a, s, j)$ of matrix R_s represents the immediate reward of taking action s when the person is in state i' at age a and as a result the person transitions to state j.

The problem is then to solve for the optimal values of D_a. Use of MDP in this context has been studied before, (22) so we do not discuss further details of the methodology here. We only show the formulation of the problem in the context of identifying optimal screening schedules for mammography considering costs of screening and monetary value per quality-adjusted life-year lived.

The above MDP was solved using dynamic programming, which is formulated as follows.

Let $V(i', a, s)$ be the value of choosing action s when the system is in state i' at age a,

$$V(i', a, s) = \sum_{i' \in k} \left(\frac{\rho_{i'}}{\sum_{m \in k} \rho_m}\right) \left[\sum_{j \in Z} p(i', a, s, j) r(i', a, s, j) + \sum_{j \in Z} p(i', a, s, j) J^*(i', a+1)\right] \quad (19)$$

$$\forall s \in S, \forall\ i' \in k = \{[H_a], [U_{i,a}]\}$$

where,

$$J^*(i', a) = r(i', a, s^*(i', a)) + \sum_{j \in Z} p(i', a, s^*(i', a), j) J^*(i', a+1) \quad (20)$$

Then, the optimal decision $s^*(i', a)$ at age a and disease state i' can be written as

$$s^*(i', a) = \begin{cases} arg\ max_{s \in S}\ V(i', a, s)\ , if\ i' = k = \{[H_a], [U_{i,a}]\}, \\ Do\ nothing, if\ i' \epsilon \{[D_{i,a}], M\} \end{cases} \quad (21)$$

Note that, with the above equations, all states in $= \{[H_a], [U_{i,a}]\}$, will have the same optimal action because, in the absence of diagnosis, we cannot distinguish between persons in preclinical cancer states $[U_{i,a}]$ from healthy state $[H_a]$. For persons in states $[D_{i,a}]$ and M, i.e., for persons in clinical cancer states and deaths, respectively, the action is to do nothing.

Transition probabilities, $p(i', a, s, j)$, are estimated using the parameterized model from section A. The specific equations are presented in Tables S4, S5, and S6. Immediate rewards incorporate the costs and benefits of screening as follows.

$$r(i', a, s = no\ screening, j) = \begin{cases} 0, if\ j\ is\ mortality \\ r_{LY}.q_j + c_d + c_i, if i' \in [U_{i,a}]\ and\ j \in [D_{i,a}] \\ r_{LY}.q_j + c_t, if i' \in [D_{i,a}]\ and\ j \in M \\ r_{LY}.q_j\ otherwise \end{cases}, \text{ and} \quad (22)$$

$$r(i', a, s = screening, j) = \begin{cases} 0\ if\ j\ is\ mortality \\ r_{LY}.q_j + c_d + c_i, if\ i' \in [U_{i,a}]\ and\ j \in \{[D_{i,a}]\} \\ r_{LY}.q_j + c_t, if i' \in [D_{i,a}]\ and\ j \in M \\ r_{LY}.q_j + c_s\ otherwise \end{cases} \quad (23)$$

where,

$c_s = -\left(\zeta_{mammography_a} c_{mammogram} + \left(1 - \zeta_{mammography_a}\right)(c_{mammogram} + c_{-diagnosis})\right)$,

$c_d = -\left(c_{mammogram} + c_{+diagnosis}\right)$,

$r_{LY} =$ value-per-QALY lived,

q_j= QALY associated with state j, $q_j = \begin{cases} 1\ if\ j = H_a \\ 0\ if\ j = M \\ 0 < q_j < 1\ otherwise \end{cases}$, (24)

$\zeta_{mammography_a}$ is the specificity of mammography at age a,

$c_{mammogram}$ is the unit cost of mammography per person,

$c_{-diagnosis}$ is the cost of follow-up diagnostic tests for a false positive (per person)

$c_{+diagnosis}$ is the cost of follow-up diagnostic tests for a true positive (per person)

c_i is the initial treatment cost per person,

c_t is terminal treatment cost per person, which was applied at the final year of life for women who die from breast cancer.

Table S4: Notation used in transition probability matrix

$\theta_{i,a}$	Onset rate of breast cancer
$\lambda i, a$	Dwell rate for cancer stage i and age a
$d_{i,a}$	Diagnostic rate of cancer in stage i and age a
μ_a	Natural mortality rate at age a
$\overline{\overline{\mu_{i,a}}}$	Diseased mortality in cancer stage i and age a

Table S5: Transition probability matrix for action = no screening

$$
\begin{bmatrix}
e^{-\mu_a}\left(\frac{1}{\theta_a+1}\right) & e^{-\mu_a}\left(\frac{\theta_a}{\theta_a+1}\right) & 0 & 0 & 0 & 0 & 0 & 1-e^{-\mu_a} \\
0 & e^{-\mu_a}\left(\frac{1}{d_{0,a}+\lambda_{0,a}+1}\right) & e^{-\mu_a}\left(\frac{\lambda_{0,a}}{d_{0,a}+\lambda_{0,a}+1}\right) & 0 & e^{-\mu_a}\left(\frac{d_{0,a}}{d_{0,a}+\lambda_{0,a}+1}\right) & 0 & 0 & 1-e^{-\mu_a} \\
0 & 0 & e^{-\mu_a}\left(\frac{1}{d_{1,a}+\lambda_{1,a}+1}\right) & e^{-\mu_a}\left(\frac{\lambda_{1,a}}{d_{1,a}+\lambda_{1,a}+1}\right) & 0 & e^{-\mu_a}\left(\frac{d_{1,a}}{d_{1,a}+\lambda_{1,a}+1}\right) & 0 & 1-e^{-\mu_a} \\
0 & 0 & 0 & e^{-\mu_a}\left(\frac{1}{d_{2,a}+\lambda_{2,a}+1}\right) & e^{-\mu_a}\left(\frac{\lambda_{2,a}}{d_{2,a}+\lambda_{2,a}+1}\right) & 0 & e^{-\mu_a}\left(\frac{d_{2,a}}{d_{2,a}+\lambda_{2,a}+1}\right) & 1-e^{-\mu_a} \\
0 & 0 & 0 & 0 & e^{-\mu_a}\left(\frac{1}{d_{3,a}+1}\right) & 0 & e^{-\mu_a}\left(\frac{d_{3,a}}{d_{3,a}+1}\right) & 1-e^{-(\mu_a+\lambda_{3,a})} \\
0 & 0 & 0 & 0 & e^{-(\mu_a+\bar{\bar{\mu}}_{0,a})} & 0 & 0 & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{0,a})} \\
0 & 0 & 0 & 0 & 0 & e^{-(\mu_a+\bar{\bar{\mu}}_{1,a})} & 0 & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{1,a})} \\
0 & 0 & 0 & 0 & 0 & 0 & e^{-(\mu_a+\bar{\bar{\mu}}_{2,a})} & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{2,a})} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{3,a})} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

Table S6: Transition probability matrix for action = screening

$$
\begin{bmatrix}
e^{-\mu_a}\left(\frac{1}{\theta_a+1}\right) & e^{-\mu_a}\left(\frac{\theta_a}{\theta_a+1}\right) & 0 & 0 & 0 & 0 & 0 & 1-e^{-\mu_a} \\
0 & e^{-\mu_a}\left(\frac{s}{\bar{d}_{0,a}+\bar{\lambda}_{0,a}+s}\right) & e^{-\mu_a}\left(\frac{\bar{\lambda}_{0,a}}{\bar{d}_{0,a}+\bar{\lambda}_{0,a}+s}\right) & 0 & 0 & e^{-\mu_a}\left(\frac{\bar{d}_{0,a}}{\bar{d}_{0,a}+\bar{\lambda}_{0,a}+s}\right) & 0 & 1-e^{-\mu_a} \\
0 & 0 & e^{-\mu_a}\left(\frac{s}{\bar{d}_{1,a}+\bar{\lambda}_{1,a}+s}\right) & e^{-\mu_a}\left(\frac{\bar{\lambda}_{1,a}}{\bar{d}_{1,a}+\bar{\lambda}_{1,a}+s}\right) & 0 & 0 & e^{-\mu_a}\left(\frac{\bar{d}_{1,a}}{\bar{d}_{1,a}+\bar{\lambda}_{1,a}+s}\right) & 0 & 1-e^{-\mu_a} \\
0 & 0 & 0 & e^{-\mu_a}\left(\frac{s}{\bar{d}_{2,a}+\bar{\lambda}_{2,a}+s}\right) & e^{-\mu_a}\left(\frac{\bar{\lambda}_{2,a}}{\bar{d}_{2,a}+\bar{\lambda}_{2,a}+s}\right) & 0 & e^{-\mu_a}\left(\frac{\bar{d}_{2,a}}{\bar{d}_{2,a}+\bar{\lambda}_{2,a}+s}\right) & 0 & 1-e^{-\mu_a} \\
0 & 0 & 0 & 0 & e^{-\mu_a}\left(\frac{s}{\bar{d}_{3,a}+s}\right) & 0 & 0 & e^{-\mu_a}\left(\frac{\bar{d}_{3,a}}{\bar{d}_{3,a}+s}\right) & 1-e^{-(\mu_a+\lambda_{3,a})} \\
0 & 0 & 0 & 0 & 0 & e^{-(\mu_a+\bar{\bar{\mu}}_{0,a})} & 0 & 0 & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{0,a})} \\
0 & 0 & 0 & 0 & 0 & 0 & e^{-(\mu_a+\bar{\bar{\mu}}_{1,a})} & 0 & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{1,a})} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & e^{-(\mu_a+\bar{\bar{\mu}}_{2,a})} & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{2,a})} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-e^{-(\mu_a+\bar{\bar{\mu}}_{3,a})} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

Where,

$$\bar{d}_{i,a} = \left(1 - \eta_{mammography}\right) d_{i,a} + \eta_{mammography} - \eta_{mammography}\left(1 - \eta_{mammography}\right) d_{i,a}$$

$$\bar{\lambda}_{i,a} = \left(1 - \eta_{mammography}\right) \lambda_{i,a}$$

$$s = \left(1 - \eta_{mammography}\right)$$

B.2. Data assumptions used for the MDP model

Country-specific data related to the natural cancer progression, specifically the transition probability matrices in Tables S5 and S6, are the same data used in the two-step Markov process methodology and are listed in Table S3. Data related to mammography (film) screening are presented in Table S7. We assumed the use of film mammography in Peru as the availability of digital mammography is limited in developing countries (23). We used data f from the Breast Cancer Surveillance Consortium (BCSC) presented in (24). Further, as mammography sensitivity and specificity varied by breast density, we used weighted average values, weighted by the proportion of persons presenting with the different breast density as reported in the BCSC.

Table S7: Parameters specific to screen-film mammography

Parameter name	**Assumption** (25) (26) (24) (27)				
$\zeta_{mammography}$ (Specificity of film mammogram) for Peru and US (24), (27)	Age	Initial	Annual	Biennial	Triennial
	<29	0.83000	0.83000	0.83000	0.83000
	30-34	0.85800	0.85800	0.85800	0.85800
	35-39	0.87500	0.87500	0.87500	0.87500
	40–49	0.85356	0.91812	0.90472	0.89606
	50-59	0.85576	0.91974	0.90498	0.90013
	60-69	0.86576	0.92974	0.91459	0.91013
	70-79	0.88384	0.93602	0.92127	0.91974
$\eta_{mammography}$ (Sensitivity of film mammogram) for Peru and US (24), (27)	Age	Initial	Annual	Biennial	Triennial
	<29	0.66700	0.66700	0.66700	0.66700
	30-34	0.81500	0.81500	0.81500	0.81500
	35-39	0.76100	0.76100	0.76100	0.76100
	40–49	0.87158	0.75644	0.8173	0.83026
	50-59	0.88126	0.77184	0.82155	0.83783
	60-69	0.90754	0.80298	0.85269	0.86897
	70-79	0.92611	0.84373	0.88126	0.8964
$\zeta_{mammography}$ (Specificity of film mammogram) for US validation (28)	Age	Specificity			
	<40	0.906			
	40-44	0.906			
	45-49	0.904			
	50-54	0.916			
	55-59	0.922			
	60-64	0.925			
	65-69	0.93			
	70-74	0.937			
	75-89	0.942			

$\eta_{mammography}$ (Sensitivity of film mammogram) for US validation and Peru sensitivity analysis (28), (29)	Age: Sensitivity <40: 0.55 40-44: 0.645 45-49: 0.701 50-54: 0.744 55-59: 0.744 60-64: 0.744 65-69: 0.744 70-74: 0.798 75-89: 0.807
$\zeta_{mammography}$ (Specificity of film mammogram) for Peru sensitivity analysis (28)	Age: Specificity <40: 0.841 40-44: 0.841 45-49: 0.823 50-54: 0.816 55-59: 0.84 60-64: 0.856 65-69: 0.86 70-74: 0.869 75-89: 0.88
c_{screen} (Screening cost) for Peru (12)	2.45 USD
c_{screen} (Screening cost) for US (24)	81.35 USD
$c_{diagnosis}$ (Cost of follow-up tests if diagnosed) for Peru (12)	True positive, $: 551.36 False positive, $: 72.18
$c_{diagnosis}$ (Cost of follow-up tests if diagnosed) for US (24)	Age group: True positive, $: False positive, $ 40–49: 2187.89: 229.1612 50–64: 2053.74: 271.6121 65–74: 2065.13: 272.353 ≥75: 1741.3: 280.5171
$c_{treatment}$ (Cost of treatment by stage at diagnosis) for US (24)	Stage: Initial, $: Terminal, $ In situ: 13055: 35335 localized: 13055: 35335 Regional: 24682: 41825 Distant: 38119: 58665
$c_{treatment}$ (Cost of treatment by stage at diagnosis) for Peru (12) for initial cost; proportion of terminal to initial cost for US used in calculation of terminal cost for Peru	Stage: Initial, $: Terminal, $ In situ: 218.01: 590 localized: 218.01: 590 Regional: 464.58: 787 Distant: 684.84: 1053
q_j= quality-adjusted life-years associated with state j	$q = [1, 1, 1, 1, 1, 0.992, 0.992, 0.971, 0.46, 0]$ corresponding to stage $[H_a, U_{in-situ}, U_{local}, U_{regional}, U_{distant}, D_{in-situ}, D_{local}, D_{regional}, D_{distant}, M]$

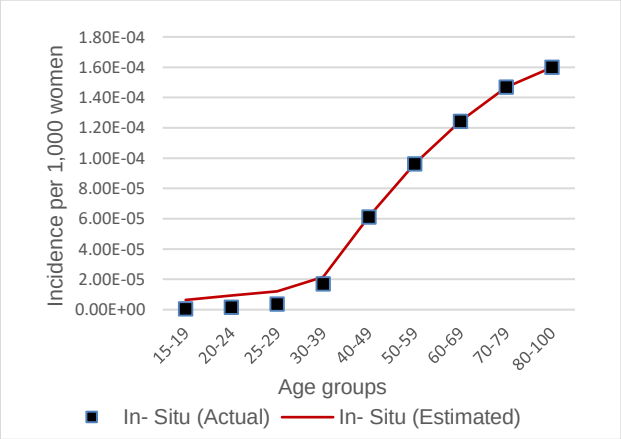
C. Model verification on the US population

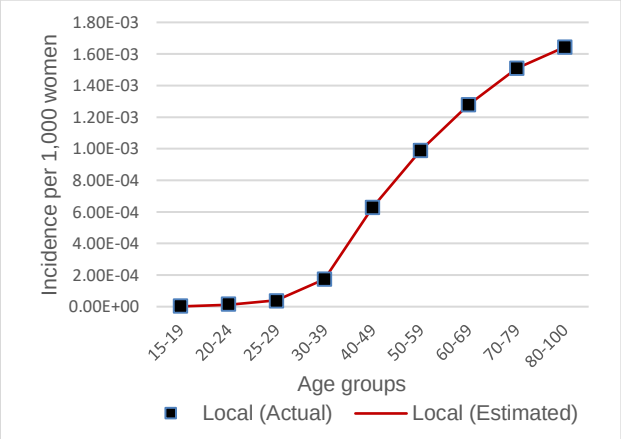
C.1. Verifying parameterization of natural history model for the US

We verified that our model outputs match data or results observed in the CISNET study models, specifically the cumulative risk by age of cancer onset, and incidence of cancers by age and stage at diagnosis.

Table S8: Cumulative probability of onset of cancer by age

Age	**US Study** (2)	**Estimations from our model**
20	0.000	0.001
25	0.002	0.003
30	0.005	0.007
35	0.021	0.019
40	0.046	0.046
45	0.105	0.099
50	0.169	0.172
55	0.233	0.258
60	0.328	0.354
65	0.436	0.457
70	0.563	0.563
75	0.707	0.670
80	0.852	0.799
85	1.000	1.000





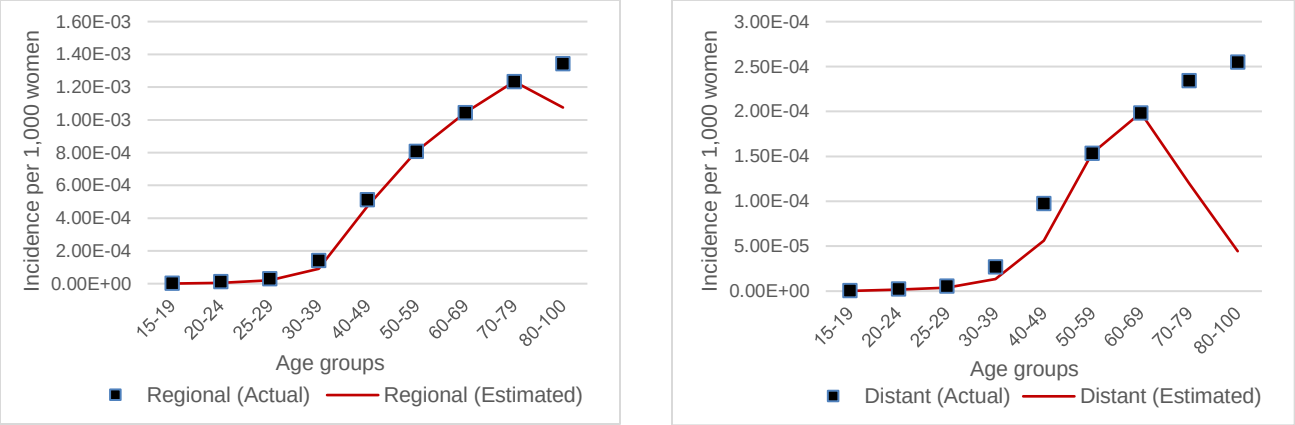


Figure S4: Comparison of estimated versus actual incidence by age and stage at diagnosis for US

C.2. Model validation on the US population

Results and analysis from our model related to mammography screening were compared with results from the CISNET study. Details of the analysis related to Figures S5, S6, and S7, and Table S8 are discussed in the main paper, under Validation section. All metrics are undiscounted.

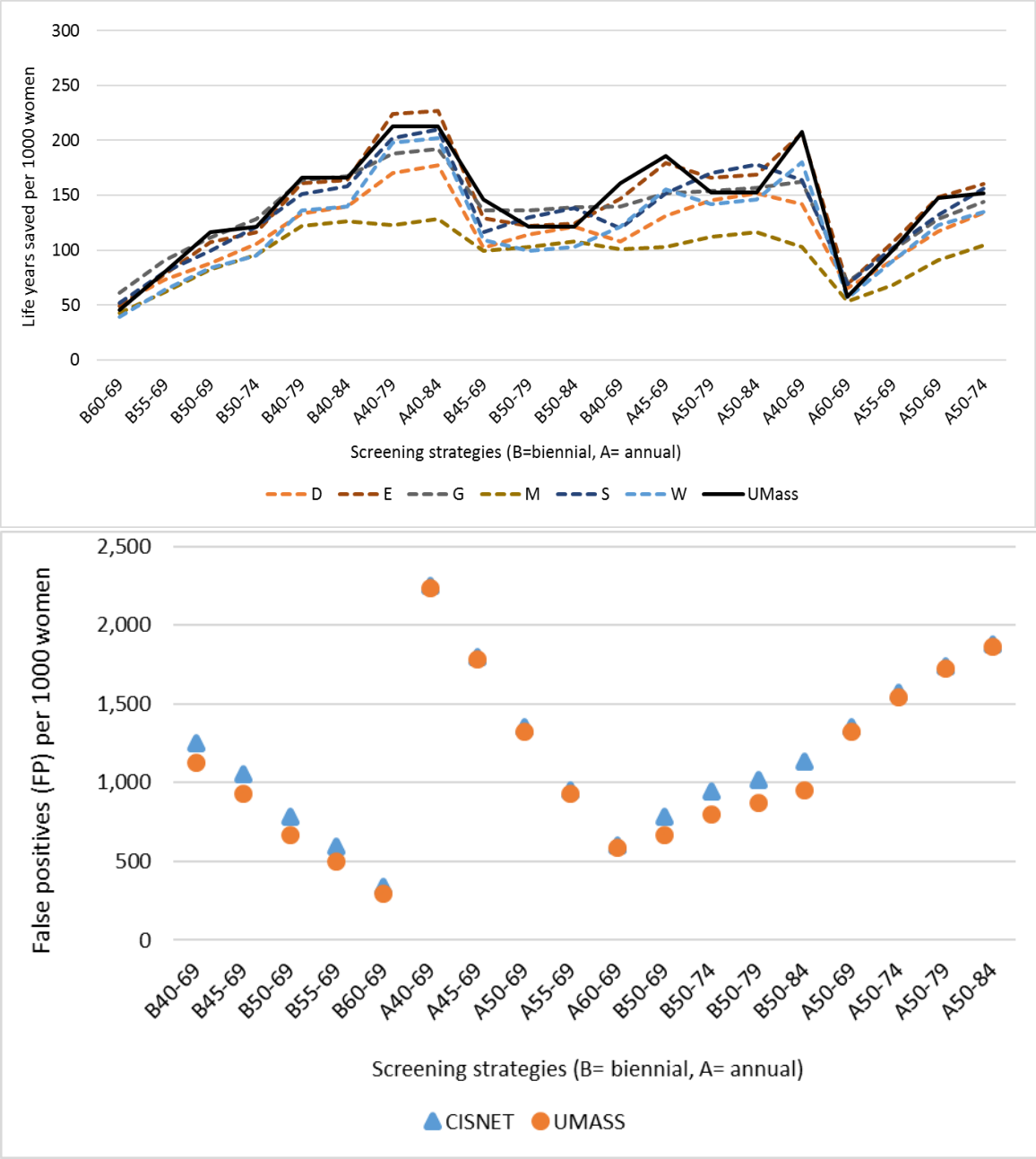


Figure S5: Model validation on the US population: Comparing benefits (life years saved per 1000 women) and harms (false positives) between our model (UMass) and CISNET model estimations. The x-axis presents the different screening strategies, biennial (B) or annual (A), and ages to screen. CISNET model group abbreviations: D = Dana-Farber Cancer Institute; E = Erasmus Medical Center; G = Georgetown University; M = M.D. Anderson Cancer Center; S = Stanford University; W = University of Wisconsin/Harvard.

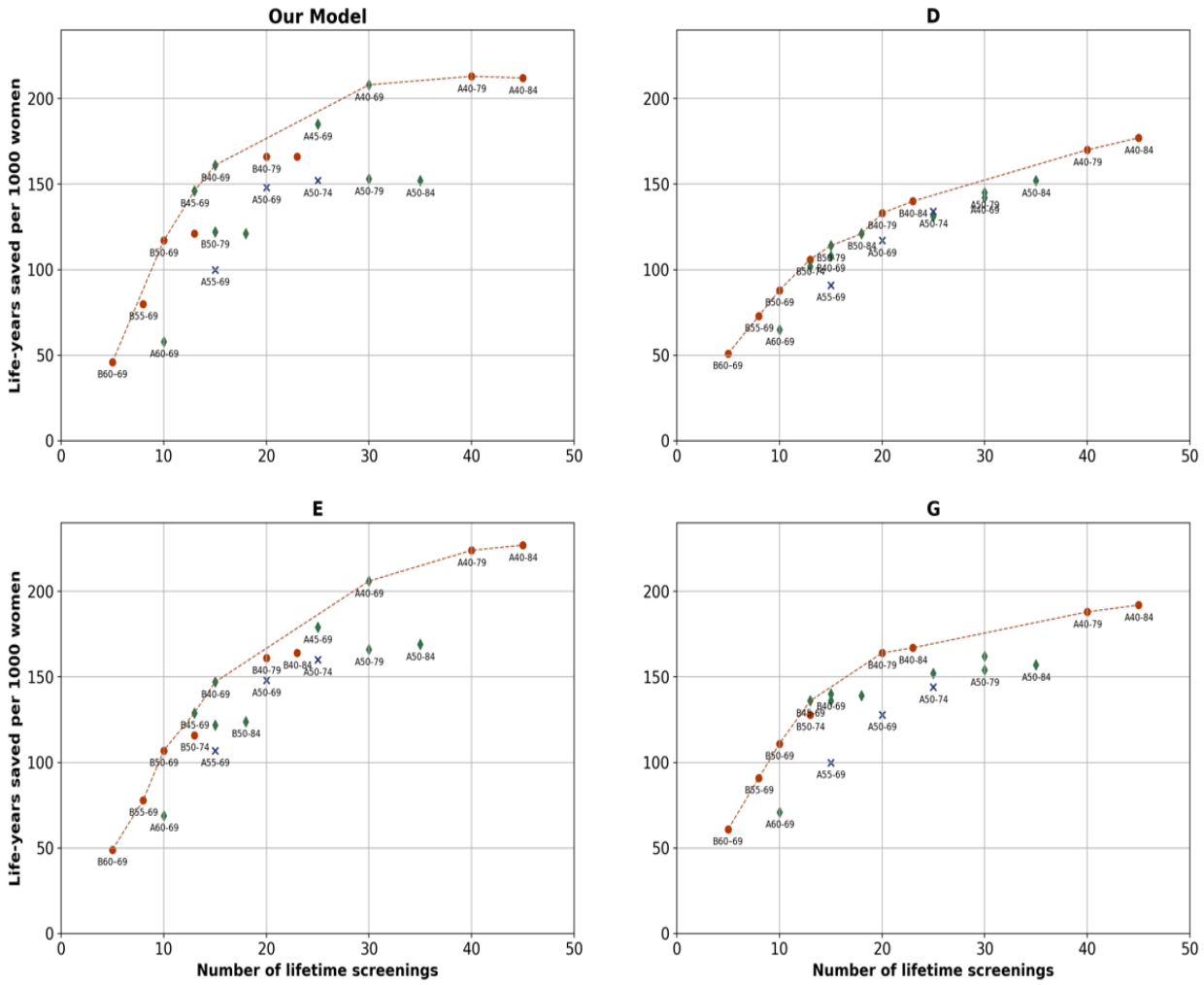


Figure S6: Comparing efficiency frontiers from our model with 3 CISNET models for pre-selected scenarios evaluated in the CISNET study for the US population. All metrics are undiscounted.

Figure legend: In the CISNET study, the screening schedules with red full circles were categorized as efficient (non-dominated in at least 5 CISNET models), scenarios in green diamond were categorized as borderline (dominated in 2-4 models), and the scenarios in blue "cross" were categorized as inefficient (dominated in all the models); Model Group Abbreviations: D (Dana Farber Cancer Center), E (Erasmus Medical Center), G (Georgetown U); B= biennial screening, A=annual screening

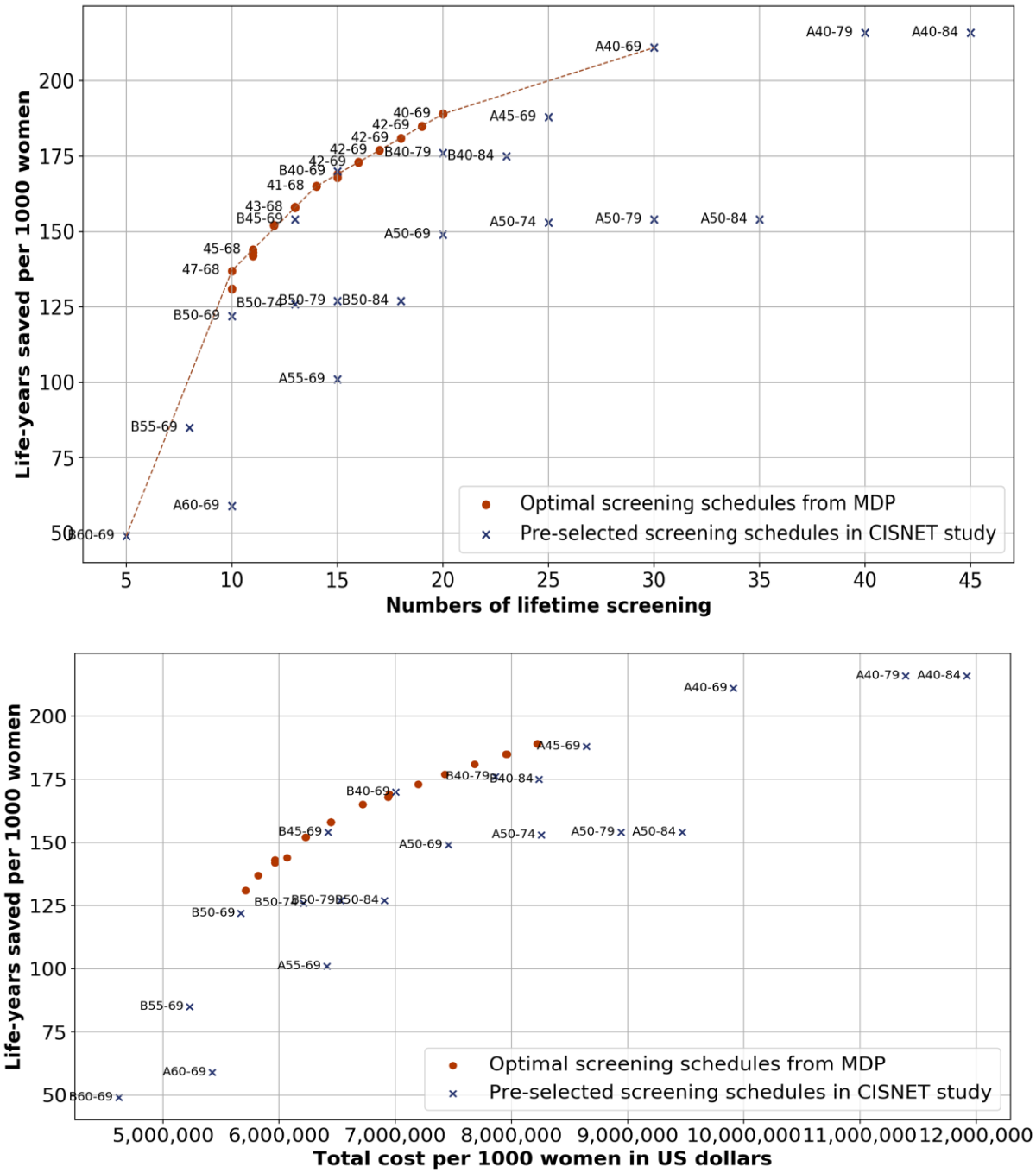


Figure S7: Efficiency frontier plotted by combining pre-selected scenarios in CISNET study with optimal scheduled generated by the model for the US population. All metrics are undiscounted.

Table S9: Benefits, harms, and costs under alternative screening schedules for the US.

Number of lifetime screenings	Age interval for screening	False positives per 1000 women	Life-years saved per 1000 women	QALYs saved per 1000 women	Total cost per 1000 women (undiscounted) (USD)	Cost per life-year saved (USD)	Cost per QALY saved (USD)
0	NA	NA	reference	reference	3,730,421	reference	reference
5 (B60-69)	60-69	388	49	69.23491	4,618,171	18,254	12,830
8 (B55-69)	55-69	636	85	114.9399	5,227,876	17,630	13,028
10	47-68	841	132	167.7138	5,710,988	15,099	11,806
11	45-68	931	142	180.5029	5,963,972	15,775	12,374
12	44-68	1022	152	190.8303	6,225,760	16,424	13,072
13	43-68	1097	158	197.9314	6,443,835	17,220	13,705
14	41-68	1190	165	205.5111	6,720,099	18,111	14,543
15	41-69	1264	168	209.3046	6,938,201	19,143	15,321
16	41-69	1349	173	215.6531	7,195,497	20,000	16,062
17	42-69	1437	177	220.4828	7,427,262	20,842	16,761
18	40-69	1513	181	224.1628	7,681,212	21,887	17,618
19	40-69	1600	185	229.3496	7,952,445	22,857	18,440
20	40-69	1693	189	233.9974	8,222,971	23,710	19,192

Note: Costs and benefits are undiscounted

D. Sensitivity analysis

D.1. Impact of mammography sensitivity and specificity

In the base case results presented in the main manuscript, we assumed the use of film mammography in Peru, as the availability of the more advanced digital mammography in developing countries is limited (23). Estimates from the breast cancer surveillance consortium (BCSC) suggests that mammography sensitivity and specificity have been increasing over time, representing the advancements in diagnostic tools. Therefore, we used mammography specificity and sensitivity from the 1995-1999 era (presented in Table S7) to test the impact of the unavailability of the most recent technology and human expertise. The results are presented in Table S10 below and discussed in the main manuscript.

Table S10: Sensitivity analysis using lower values of mammography test sensitivity and specificity- Comparison of model outputs between basecase and sensitivity analysis case for Peru

Number of lifetime screenings	Age interval for screening (base case)	Age interval for screening	False positives per 1000 women (base case)	False positives per 1000 women	Life-years saved per 1000 women (base case)	Life-years saved per 1000 women	QALYs saved per 1000 women	Total cost per 1000 women (undiscounted) (USD) (base case)	Total cost per 1000 women (undiscounted) (USD)	Cost per life-year saved (USD) (base case)	Cost per life-year saved (USD)	Cost per QALY saved (USD)
0	NA	NA	NA	NA	reference	reference	reference	52,644	46,967	reference	reference	reference
1	51-51	51-51	129	168	19	16	19	135,788	153,307	4,376	6,535	5,670
2	50-56	50-57	220	310	31	28	32	192,122	243,310	4,499	7,116	6,081
3	46-57	47-58	315	469	44	38	44	252,467	344,421	4,541	7,925	6,825
4	45-60	46-60	395	619	54	47	54	303,885	439,667	4,653	8,416	7,275
5	44-61	44-61	489	743	62	55	63	363,927	518,721	5,021	8,628	7,530
7	41-62	42-62	676	1,069	78	69	78	484,359	725,455	5,535	9,887	8,698
8	42-63	42-63	766	1,212	84	75	85	542,360	816,670	5,830	10,312	9,094
9	42-64	42-63	842	1,362	88	80	90	591,269	912,208	6,121	10,865	9,593
10	41-64	41-64	939	1,507	95	84	95	654,089	1,004,946	6,331	11,350	10,060
11	42-65	42-65	978	1,644	99	89	100	680,196	1,091,992	6,339	11,791	10,447
12	41-66	42-65	1,055	1,765	101	91	103	730,554	1,169,479	6,712	12,311	10,885
13	40-66	40-66	1,132	1,914	105	96	108	780,415	1,264,450	6,931	12,640	11,251
14	41-67	41-66	1,213	2,091	107	99	111	832,693	1,376,901	7,290	13,446	11,948
15	40-67	40-66	1,302	2,240	112	103	115	890,442	1,472,028	7,480	13,856	12,361
16	40-67	40-67	1,389	2,360	114	105	118	946,532	1,549,028	7,841	14,323	12,762
17	40-67	40-67	1,477	2,528	117	108	121	1,003,876	1,655,689	8,130	14,944	13,330

Note: Costs and benefits are undiscounted

D.2. Impact of dwell-times (inverse of progression rates)

As our model is deterministic, to test the impact of uncertainty in dwell times (inverse of progression rates between preclinical disease stages) on results for optimal screening schedules, we generated 100 runs of the model by sampling for different values of dwell times between the ranges presented in the Table S11. To ensure that progression in advanced stages of cancer are more aggressive than earlier stages, we sample using the following equation. Dwell time in stage $i = LR_i + \mathrm{x} * (UR_i - LR_i)$ where $\mathrm{x} = 1, 2, 3, 4, 5 \ldots \ldots \ldots 100$

We present the results in Table S12 and S13 below and discuss the findings in Results and Discussion sections of the main manuscript.

Table S11: Data assumptions for sensitivity analysis on dwell-times (inverse of progression rates)

Stage, *i*	**Lower range (LR)** (14)	**Upper range (UR)** (14)
DCIS	4.50	5.50
Local	2.50	3.76
Regional	1.54	3.10
Distant	1.50	2.50

Table S12: Sensitivity analysis on dwell-times- Comparing age interval, false positives per 1000 women, life-years (LY) saved per 1000 women and QALYs saved per 1000 women between basecase and sensitivity analysis case

	Age interval for screening			False positives per 1000 women			Life years saved per 1000 women			QALYs saved per 1000 women		
Dwell time range	Basecase	Lower	Upper	Basecase	Lower range	Upper range	Basecase	Lower range	Upper range	Basecase	Lower range	Upper range
No screening	–	–	–	reference	reference	reference	reference	reference	reference	reference	reference	reference
1	51-51	53-53	51-51	129	130	131	19	20	16	22	23	19
2	50-56	50-57	45-53	220	220	227	31	35	28	37	40	32
3	46-57	47-60	46-56	315	303	316	44	48	38	51	55	44
4	45-60	45-61	42-57	395	393	414	54	61	47	62	69	53
5	44-61	44-62	42-58	489	487	504	62	71	54	72	80	62
6	44-62	46-64	43-60	579	556	587	70	78	61	80	88	69
7	41-62	44-64	40-61	676	655	680	78	89	67	89	100	75
8	42-63	44-65	42-62	766	742	765	84	96	71	95	107	80
9	42-64	42-65	41-62	842	838	862	88	104	76	101	115	86
10	41-64	42-66	40-63	939	925	957	95	110	80	107	122	91
11	42-65	43-67	41-64	978	954	977	99	114	82	112	126	93
12	41-66	42-67	40-65	1,055	1,041	1,062	101	120	86	115	132	97
13	40-66	41-68	39-65	1,132	1,124	11,44	105	125	89	119	137	99
14	41-67	41-69	38-65	1,213	11,89	1,215	107	129	91	122	140	103
15	40-67	40-70	40-66	1,302	1,273	1,308	112	132	93	129	146	105

Note: Costs and benefits are undiscounted

Table S13: Sensitivity analysis on dwell-times- Comparing cost per LY saved, cost per QALY saved and cost per 1000 women between basecase and sensitivity analysis case

	Total cost per 1000 women, USD			Cost per life years saved per 1000 women, USD			Cost per QALY saved per 1000 women, USD		
Dwell time range	Basecase	Lower range	Upper range	Basecase	Lower range	Upper range	Basecase	Lower range	Upper range
No screening	reference	reference	reference	reference	reference	reference	reference	reference	reference
1	135,788	138,523	132,950	4,376	4,077	5,234	3,825	3,565	4,451
2	192,122	195,377	194,078	4,499	3,919	5,093	3,805	3,433	4,452
3	252,467	248,101	250,589	4,541	3,990	5,340	3,940	3,506	4,613
4	303,885	306,188	313,215	4,653	4,100	5,640	4,056	3,644	4,955
5	363,927	365,741	371,392	5,021	4,333	5,906	4,334	3,866	5,161
6	421,727	410,840	424,724	5,273	4,559	6,165	4,593	4,024	5,407
7	484,359	473,560	484,411	5,535	4,674	6,527	4,877	4,184	5,762
8	542,360	529,842	539,647	5,830	4,942	6,928	5,158	4,425	6,089
9	591,269	591,351	601,758	6,121	5,151	7,291	5,342	4,644	6,440
10	654,089	647,777	663,087	6,331	5,386	7,637	5,609	4,861	6,772
11	680,196	667,839	677,671	6,339	5,379	7,617	5,627	4,851	6,737
12	730,554	723,637	732,651	6,712	5,569	7,939	5,896	5,042	7,040
13	780,415	777,678	787,631	6,931	5,791	8,261	6,099	5,254	7,453
14	832,693	831,718	832,472	7,290	6,012	8,556	6,412	5,538	7,613
15	890,442	874,706	892,507	7,480	6,191	9,045	6,640	5,623	8,032

Note: Costs and benefits are undiscounted

D.3. Impact of uncertainty in carcinoma in-situ (CIS) pathways

Recent studies, under the context of over-diagnosis of cancers, have highlighted the uncertainty around pathways of CIS stage and mammography sensitivity for diagnosis at this stage, and their corresponding impact on the CIS progression rate estimates. (30), (31), (32), (33), (34), (35), (36)Therefore, as uncertainty analysis, we evaluated 5 CIS Cases (Table S14), each using different combinations of values for CIS progression rate, proportion of invasive cancers initiating directly in local stage, and mammography sensitivity in CIS stage. We present results for Peru and the US below in Figures S8 and S9. We present the observations and discuss the findings in Results and Discussion sections of the main manuscript.

Table S14: Scenarios for uncertainty analysis of CIS pathway

Uncertainty analysis case	Dwell-time for progressive CIS	Proportion of cancers initiating directly as local invasive cancer	Mammography sensitivity for CIS	References
1	5.22 years	0%	40% for ages over 50 years. 28% for ages less than 50	(14)
2	3 months	0%	88%	(37)
3	5 months	0%	88%	(37)
4	2 years	18.9%	40%	(30)
5	15 years	18.9%	40%	(30)

Table S15: Uncertainty analysis on CIS pathways- Comparing LY saved per 1000 women, and false positives per 1000 women for Peru between basecase and CIS uncertainty cases 1 to 5

Number of lifetime screens	Life years (LY) saved per 1000 women (base case)	Life years saved per 1000 women (case 1)	Life years saved per 1000 women (case 2)	Life years saved per 1000 women (case 3)	Life years saved per 1000 women (case 4)	Life years saved per 1000 women (case 5)	False positives per 1000 women (base case)	False positives per 1000 women (case 1)	False positives per 1000 women (case 2)	False positives per 1000 women (case 3)	False positives per 1000 women (case 4)	False positives per 1000 women (case 5)
1	19	12	6	6	8	12	129	131	114	115	208	136
2	31	20	12	12	14	22	220	221	206	207	295	234
3	44	28	16	17	20	30	315	310	296	295	385	325
4	54	33	21	22	25	36	395	387	368	372	461	420
5	62	38	25	26	30	42	489	480	455	455	546	513
6	70	44	28	29	34	47	579	568	514	518	640	576
7	78	49	32	33	38	52	676	656	589	610	718	661
8	84	53	35	36	42	56	766	741	686	685	790	749
9	88	57	39	40	45	60	842	792	778	778	875	839
10	95	62	42	43	49	63	939	881	860	868	963	925
11	99	66	45	47	52	66	978	970	945	945	1,060	1,011
12	101	69	48	49	56	67	1,055	1,054	1,018	1,018	1,078	1,095
13	105	73	50	51	58	71	1,132	1,129	1,046	1,060	1,166	1,176
14	107	75	52	54	61	72	1,213	1,210	1,133	1,135	1,248	1,259
15	112	78	54	56	63	75	1,302	1,295	1,215	1,216	1,334	1,331

Note: Costs and benefits are undiscounted

Table S16: Uncertainty analysis on CIS pathways- Comparing cost per 1000 women, cost per LY saved saved for Peru between basecase and CIS uncertainty cases 1 to 5

Number of lifetime screens	Cost per 1000 women (basecase)	Cost per LY saved (basecase)	Cost per 1000 women (case 1)	Cost per LY saved (case 1)	Cost per 1000 women (case 2)	Cost per LY saved (case 2)	Cost per 1000 women (case 3)	Cost per LY saved (case 3)	Cost per 1000 women (case 4)	Cost per LY saved (case 4)	Cost per 1000 women (case 5)	Cost per LY saved (case 5)
1	135,788	4,376	135,830	7,149	129,024	12,526	129,062	12,276	138,641	10,466	131,019	7,090
2	192,122	4,499	193,458	6,875	188,063	11,258	188,358	10,865	188,370	9,290	193,832	6,883
3	252,467	4,541	250,243	7,113	246,328	11,826	244,601	11,196	244,033	9,409	252,667	6,981
4	303,885	4,653	299,973	7,474	292,776	11,213	293,991	10,909	301,988	9,765	314,130	7,401
5	363,927	5,021	360,229	7,993	348,899	11,562	347,972	11,262	351,312	9,847	373,950	7,776
6	421,727	5,273	416,699	8,362	387,080	11,733	388,623	11,418	405,595	10,300	415,552	7,827
7	484,359	5,535	473,416	8,600	435,981	11,965	447,694	11,918	466,612	10,710	471,040	8,240
8	542,360	5,830	528,402	8,979	498,311	12,581	496,824	12,212	516,815	10,883	527,786	8,664
9	591,269	6,121	561,846	8,901	557,611	12,762	556,120	12,382	563,390	11,248	586,271	9,097
10	654,089	6,331	619,480	9,129	610,658	13,050	614,314	12,903	618,502	11,516	642,001	9,530
11	680,196	6,339	676,791	9,470	665,225	13,397	664,028	13,055	674,917	11,878	697,797	9,962
12	730,554	6,712	731,399	9,782	713,079	13,762	711,823	13,407	737,388	12,281	752,753	10,528
13	780,415	6,931	780,252	9,980	731,854	13,564	740,038	13,306	750,856	12,054	805,382	10,788
14	832,693	7,290	832,282	10,359	788,329	14,006	788,392	13,516	807,695	12,398	859,501	11,259
15	890,442	7,480	887,648	10,725	841,214	14,406	840,664	14,011	860,801	12,817	906,150	11,459

Note: Costs and benefits are undiscounted

Table S17: Uncertainty analysis on CIS pathways- Comparing QALYs saved per 1000 women and cost per QALY saved for Peru between basecase and CIS uncertainty cases 1 to 5

Number of lifetime screens	QALY saved per 1000 women (base case)	QALY saved per 1000 women (case 1)	QALY saved per 1000 women (case 2)	QALY saved per 1000 women (case 3)	QALY saved per 1000 women (case 4)	QALY saved per 1000 women (case 5)	Cost per QALY saved (basecase)	Cost per QALY saved (case 1)	Cost per QALY saved (case 2)	Cost per QALY saved (case 3)	Cost per QALY saved (case 4)	Cost per QALY saved (case 5)
1	22	13	7	7	9	14	3,825	6,184	10,125	10,044	10,466	6,301
2	37	24	14	14	17	24	3,805	5,928	9,558	9,342	9,290	6,206
3	51	32	19	20	24	33	3,940	6,105	10,176	9,582	9,409	6,256
4	62	39	25	26	29	40	4,056	6,352	9,431	9,264	9,765	6,638
5	72	45	30	31	35	47	4,334	6,857	9,757	9,505	9,847	7,010
6	80	51	34	35	40	52	4,593	7,173	9,694	9,519	10,300	7,080
7	89	57	38	39	45	57	4,877	7,398	9,899	10,018	10,710	7,448
8	95	61	42	43	49	61	5,158	7,752	10,562	10,258	10,883	7,851
9	101	66	46	48	53	65	5,342	7,697	10,836	10,521	11,248	8,283
10	107	71	50	51	57	68	5,609	7,954	11,051	10,993	11,516	8,675
11	112	75	54	55	61	71	5,627	8,292	11,384	11,110	11,878	9,069
12	115	79	56	58	64	73	5,896	8,582	11,686	11,407	12,281	9,561
13	119	83	59	60	67	77	6,099	8,758	11,479	11,349	12,054	9,833
14	122	86	61	63	70	79	6,412	9,076	11,946	11,551	12,398	10,245
15	129	89	64	66	73	82	6,640	9,405	12,286	11,986	12,817	10,470

Note: Costs and benefits are undiscounted

Table S18: Uncertainty analysis on CIS pathways- Comparing number of lifetime screens under different scenarios for both Peru and the US

Number of lifetime screens	Basecase		Case 1		Case 2		Case 3		Case 4		Case 5	
	Peru	US	Peru	US	Peru	US	Peru	US	Peru	US	Peru	US
1	51-51	_	52-52	_	61-61	_	60-60	_	53-53	_	47-47	_
2	50-56	_	50-55	_	53-62	61-66	52-61	_	52-60	_	44-48	_
3	46-57	_	50-57	_	50-63	60-70	51-63	61-70	52-62	60-71	44-51	_
4	45-60	_	50-60	_	52-65	57-71	51-64	57-71	50-63	56-71	44-52	_
5	44-61	_	46-61	_	51-66	55-71	51-66	54-71	50-64	56-72	43-53	_
6	44-62	_	46-62	_	52-70	53-72	51-70	54-72	50-66	53-72	42-54	_
7	41-62	_	46-63	50-68	51-71	53-72	48-70	52-72	47-66	52-73	42-55	_
8	40-63	_	45-63	50-69	46-71	52-73	47-71	52-73	46-67	51-73	41-56	45-61
9	42-64	_	45-64	46-69	45-71	51-73	46-71	51-73	46-70	51-73	40-56	43-61
10	41-64	47-68	44-64	45-70	47-72	51-74	45-72	48-74	45-70	50-73	40-57	41-61
11	40-64	45-68	43-64	46-70	45-72	49-74	45-72	47-74	44-70	48-74	40-58	42-62
12	41-66	44-68	41-65	43-70	44-73	50-75	44-73	47-74	42-71	48-74	41-59	42-62
13	40-66	43-68	42-65	44-70	44-73	46-75	44-73	47-75	44-71	47-75	40-60	41-62
14	41-67	41-68	42-66	43-70	42-74	45-75	43-73	46-75	42-72	46-75	41-61	41-62
15	40-67	41-69	42-66	44-71	42-74	45-76	42-74	45-76	41-72		38-61	40-62

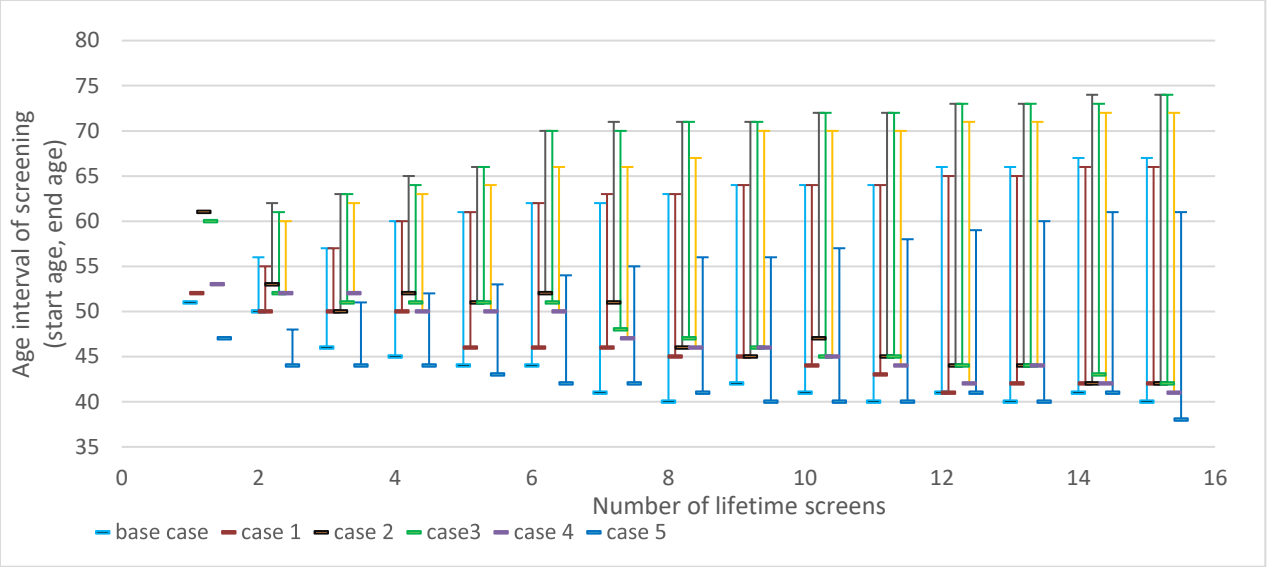


Figure S8: Uncertainty analysis on CIS pathways- Comparing screening age intervals for Peru between basecase and CIS uncertainty analysis cases 1 to 5

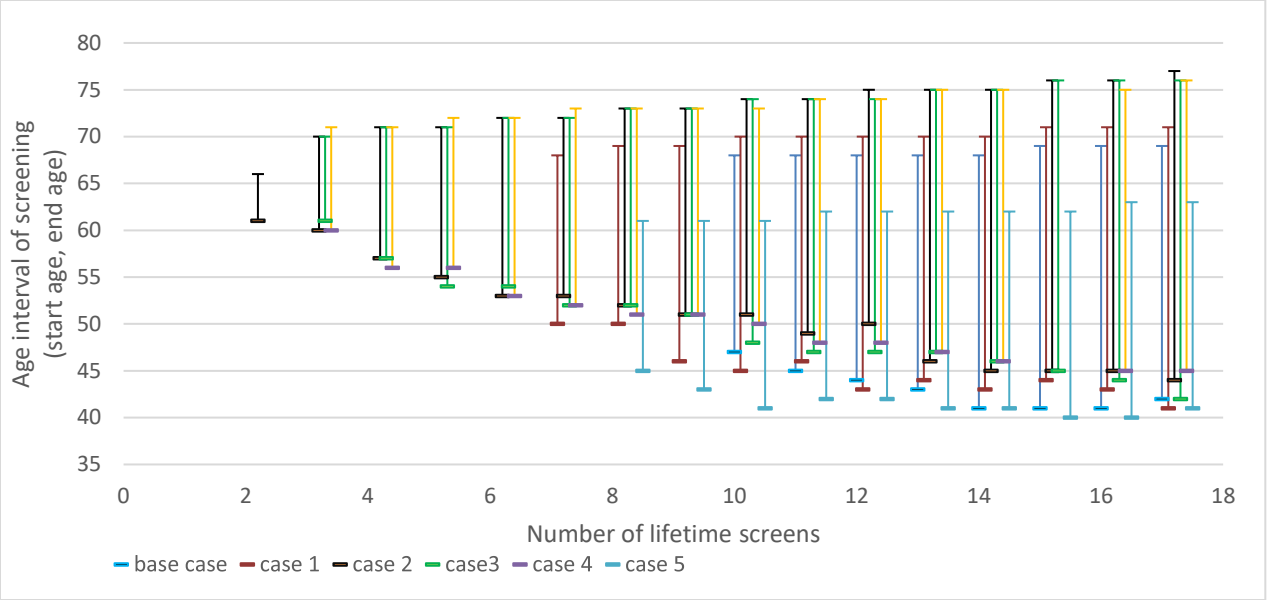


Figure S9: Uncertainty analysis on CIS pathways- Comparing screening age intervals for the US between basecase and CIS uncertainty analysis cases 1 to 5

References

1. Mandelblatt J, Schechter C, Lawrence W, Yi B, Cullen J. Chapter 8: The spectrum population model of the impact of screening and treatment on US breast cancer trends from 1975 to 2000: Principles and practice of the model methods. JNCI Monographs. 2006 Oct 1; 36: p. 47-55.

2. Tan SY, Oortmarssen GJv, Koning HJd, Boer R, Habbema JDF. The MISCAN-Fadia Continuous Tumor Growth Model for Breast Cancer. 2006; 36.

3. Tan K, Simonella L, Wee H, Roellin A, Lim Y, Lim W, et al. Quantifying the natural history of breast cancer. British journal of cancer. 2013 Oct; 109(8): p. 2035.

4. Fryback D, Stout N, Rosenberg M, Trentham-Dietz A, Kuruchittham V, Remington P. Chapter 7: The Wisconsin breast cancer epidemiology simulation model. JNCI Monographs. 2006 Oct 1; 36: p. 37-47.

5. Duffy S, Day N, Tabár L, Chen H, Smith T. Markov models of breast tumor progression: some age-specific results. JNCI Monographs. 1997 Jan 1; 22: p. 93-7.

6. Tabar , Vitak , Chen , Prevost , Duffy. Update of the Swedish Two-County Trial of breast cancer screening: histologic grade-specific and age-specific results. Swiss surgery. 1999 Oct 1; 5(5): p. 199-204.

7. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012. Int J Cancer. 2012; 136:E359-86.

8. Torre L, Bray F, Siegel R, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015 March 01; 65(2): p. 87-108.

9. Gopalappa C, Guo J, Meckoni P, Munkhbat B, Pretorius C, Lauer J, et al. A Two-Step Markov Processes Approach for Parameterization of Cancer State-Transition Models for Low-and Middle-Income Countries. Medical Decision Making. 2018; 38(4): p. 520-530.

10. World Health Organization. Tackling NCDs: 'best buys' and other recommended interventions for the prevention and control of noncommunicable diseases. [Online].: WHO; 2017 [cited 2019 March 18. Available from: https://apps.who.int/iris/handle/10665/259232.

11. World Health organization. CANCER INTERVENTIONS TECHNICAL BRIEFING. [Online].: WHO; 2017 [cited 2019 March 18. Available from: https://www.who.int/ncds/governance/Cancers-FINAL-18May.pdf?ua=1.

12. Ralaidovy AH, Gopalappa C, Ilbawi A, Pretorius C, Lauer JA. Cost-effective interventions for breast cancer, cervical cancer, and colorectal cancer: new results from WHO-CHOICE. Cost Effectiveness and Resource Allocation. 2018; 16(1): p. 38.

13. IARC WHO. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide. [Online].; 2012 [cited 2018 February 12. Available from: http://globocan.iarc.fr/Pages/age-specific_table_sel.aspx.

14. Okonkwo QL, Draisma G, der Kinderen A, Brown ML, de Koning HJ. Breast cancer screening policies in developing countries: a cost-effectiveness analysis for India. Journal of the National Cancer Institute. 2008; 100(18): p. 1290-1300.

15. Groot MT, Baltussen R, Uyl-de Groot CA, Anderson BO, Hortobágyi GN. Costs and health effects of breast cancer interventions in epidemiologically different regions of Africa, North America, and Asia. The Breast Journal. 2006; 12(s1).

16. Zelle SG, Baltussen R, Otten JD, Heijnsdijk EA, van Schoor G, Broeders MJ. Predicting the stage shift as a result of breast cancer screening in low-and middle-income countries: a proof of concept. J Med Screen. 2015 March 01; 22(1): p. 8-19.

17. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics. CA: a cancer journal for clinicians. 2015 March 01; 65(2): p. 87-108.

18. Cronin KA, Mariotto AB, Clarke LD, Feuer EJ. Additional Common Inputs for Analyzing Impact of Adjuvant Therapy and Mammography on U.S. Mortality. Journal of the National Cancer Institute Monographs. 2006.

19. Vallejos CS, Gómez HL, Cruz WR, Pinto JA, Dyer RR, Velarde R, et al. Breast cancer classification according to immunohistochemistry markers: subtypes and association with clinicopathologic variables in a peruvian hospital database. 2010; 10(4).

20. Zelle SG, Vidaurre T, Abugattas JE, Manrique JE, Sarria G, Jeronimo J, et al. Cost-effectiveness analysis of breast cancer control interventions in Peru. PLoS One. 2013; 8(12), e82575.

21. Rojas TV. INEN Global Perspectives, Natonal Cancer Control Plan: "Plan Esperanza". Lima, Peru:; 2016.

22. Zhang J, Mason JE, Denton BT, Pierskalla WP. Applications of operations research to the prevention, detection, and treatment of disease. Wiley Encyclopedia of Operations Research and Management. Wiley. 2011.

23. Rehani M, Vassileva J. Survey of imaging technology and patient dose recording practice in developing countries. Radiation protection dosimetry. 2018 Feb 9; 181(3): p. 240-5.

24. Stout NK, Lee SJ, Schechter CB, Kerlikowske K, Alagoz O, Berry D, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. JNCI: Journal of the National Cancer Institute. 2014; 106(6).

25. International Agency for Research on Cancer. Breast cancer screening Volume 15. Lyon Cedex 08, France:; 2016.

26. World Health Organization. WHO position paper on mammography screening. [Online].; 2014 [cited 2019 March 27. Available from: http://apps.who.int/iris/bitstream/10665/137339/1/9789241507936_eng.pdf?ua=1&ua=1.

27. Yankaskas BC, Haneuse S, Kapp JM, Kerlikowske K, Geller B, Buist DSM. Performance of First Mammography Examination in Women Younger Than 40 Years. Journal of National Cancer Insitute. 2010; 102(10): p. 692-701.

28. BCSC. BREAST CANCER SURVEILLANCE CONSORTIUM. [Online]. [cited 2019 March 4. Available from: http://www.bcsc-research.org/statistics/performance/screening/2008/perf_age_time.html.

29. Lee S, Zelen M. Chapter 11: A Stochastic Model for Predicting the Mortality of Breast Cancer. JNCI Monographs. 2006; 2006(36): p. 79–86.

30. Ryser MD, Worni M, Turner EL, Marks JR, Durrett R, Hwang ES. Outcomes of Active Surveillance for Ductal Carcinoma in Situ: A Computational Risk Analysis. JNCI: Journal of the National Cancer Institute. 2016 May; 108(5): p. djv372.

31. De Gelder R, Heijnsdijk EA, Van Ravesteyn NT, Fracheboud J, Draisma G, De Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiologic reviews. 2011; 33(1): p. 111-121.

32. Lee SJ, Li X, Huang H, Zelen M. The Dana-Farber CISNET model for breast cancer screening strategies: An update. Medical Decision Making. 2018; 38((1_suppl)): p. 44S-53S.

33. Schechter CB, Near AM, Jayasekera J, Chandler Y, Mandelblatt JS. Structure, function, and applications of the Georgetown–Einstein (GE) breast cancer simulation model. Medical Decision Making. 2018; 38(1_suppl): p. 66S-77S.

34. Draisma G, Fracheboud J. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. Breast cancer research: BCR. 2006; 8(1): p. 202-202.

35. Gunsoy NB, Garcia-Closas M, Moss SM. Modelling the overdiagnosis of breast cancer due to mammography screening in womenaged 40 to 49 in the United Kingdom. Breast cancer research. 2012; 14(6): p. R152.

36. van Ravesteyn NT, Van Den Broek JJ, Li X, Weedon-Fekjær H, Schechter CB, Alagoz O, et al. Medical Decision Making. Modeling ductal carcinoma in situ (DCIS): An overview of CISNET model approaches. 2018; 38(1_suppl): p. 126S-139S.

37. Yen MF, Tabar L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. European journal of cancer. 2003 Aug 1; 39(12): p. 1746-54.