University of Cambridge & University College London
Version: 1.0.00
Released: 2025-09-30



# MSM-CanProg: Model Profile

## University of Cambridge & University College London

### Contact

Professor Nora Pashayan (np275@medschl.cam.ac.uk)

### Funding

The development of this model was supported by the NIH/NCI CISNET Prostate Cancer Grant (U01CA253915).

### Suggested Citation

Pashayan N, van den Hout A,. MSM-CanProg: Model Profile. [Internet] Sep 30, 2025. Cancer Intervention and Surveillance Modeling Network (CISNET). Available from: https://cisnet.cancer.gov/resources/files/mpd/prostate/CISNET-prostate-msm-canprog-model-profile-1.0.00-2025-09-30.pdf

### Version Table

| Version | Date | Notes |
|---|---|---|
| 1.0.00 | 2025-09-30 | Initial release |

University of Cambridge
& University College
London
Readers Guide



# Reader's Guide

**Core Profile Documentation**

These topics will provide an overview of the model without the burden of detail. Each can be read in about 5-10 minutes. Each contains links to more detailed information if required.

**Model Purpose**
This document describes the primary purpose of the model.

**Model Overview**
This document describes the primary aims and general purposes of this modeling effort.

**Assumption Overview**
An overview of the basic assumptions inherent in this model.

**Parameter Overview**
Describes the basic parameter set used to inform the model, more detailed information is available for each specific parameter.

**Component Overview**
A description of the basic computational building blocks (components) of the model.

**Output Overview**
Definitions and methodologies for the basic model outputs.

**Results Overview**
A guide to the results obtained from the model.

**Key References**
A list of references used in the development of the model.

University of Cambridge
& University College
London




# Model Purpose

The MSM-CanProg (UK) encompasses continuous-time multistate survival models (MSM) and a microsimulation based on the MSM framework. The main aims of our models are:

1. To estimate natural history parameters such as sojourn time, sensitivity, age of entry into screen-detectable state, and the effect of covariates, including time-varying covariates, on transition parameters.[1,2]
2. To simulate life trajectories and evaluate outcomes of screening strategies, including screen-detected cancers, interval cancers, clinically diagnosed cancers, overdiagnoses, and cancer deaths, across varying screening frequencies and age ranges.[1-4]
   Further work is in progress to incorporate time varying PSA into MSM to inform PSA-tailored screening frequency and time to next investigation in active surveillance[5].

# References

1. Rikesh Bhatt, Ardo Van Den Hout, Nora Pashayan. A multistate survival model of the natural history of cancer using data from screened and unscreened population. Statistics in Medicine. 2021;40(16):3791–3807.
2. Mai Ngoc Bui, Ardo van den Hout, Rikesh Bhatt, Nora Pashayan. Non-homogeneous multistate partial Markov models: A simulation scheme for evaluating cancer screening strategies. Under review.
3. Rikesh Bhatt, Ardo van den Hout, Antonis C Antoniou, Mitul Shah, Lorenzo Ficorella, Emily Steggall, et al. Estimation of age of onset and progression of breast cancer by absolute risk dependent on polygenic risk score and other risk factors. Cancer. 2024;130(9):1590.
4. Richard M. Martin, Emma L. Turner, Grace J. Young, Chris Metcalfe, Eleanor I. Walsh, J. Athene Lane, et al. Prostate-Specific Antigen Screening and 15-Year Prostate Cancer Mortality: A Secondary Analysis of the CAP Randomized Clinical Trial. JAMA. 2024;331(17):1460.
5. Cuevas Andrade, Nora Pashayan. Parsimonious multistate models using longitudinal data and time-dependent covariates: applications to a liver cirrhosis clinical trial. Under review.

University of Cambridge & University College London





# Model Overview

We use a non-homogeneous multistate model with time varying transition probabilities to estimate natural history parameters, and a microsimulation based on multistate framework to investigate how often and when to offer screening to optimise the benefit-harm trade-off of a screening programme.

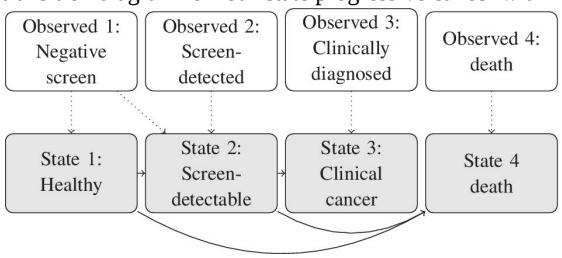
The MSM-CanProg (UK) is a multistate model with six states: 'healthy'/no-detectable cancer (S1), screen-detectable (S2-a), screen-detected (S2-b), clinically diagnosed (S3), death from other causes (S4), and cancer-specific death (S5) (see Figure 1 in Component Overview).

The transition to S2-a is interval censored, while the transitions to S3, S4 and S5 occur at exact times, with death acting as a competing risk on the transition between S1 and S2-a, and between S2-a and S3. The first observation in non-screened individuals is considered left censored. All individuals are assumed to be truly in state S1 at some initial age. An individual with negative screening test may truly be in S1 or could be in S2-a but misclassified as being in S1 (see Figure 2). Sojourn time is the length of time spent in S2-a given that an individual transitions to S3.

Key features of the multistate survival model are:

- It allows using interval censored, left censored, right censored, and left truncated panel data.
- Data can be used from both screened and unscreened populations to derive transition parameters to clinical states.
- Each transition can be specified by a variety of parametric models including exponential, Weibull, and Gompertz hazard models.
- With a parametric approach, the model can be used for prediction and out-of-sample extrapolation.
- Age varying transition hazards and age-varying misclassification are estimated simultaneously.
- The transition parameters are estimated by optimisation process to maximise the likelihood function built from the transition intensity function (see Transition Intensity and Likelihood).

Figure 2: State transition diagram for four-state progressive cancer with misclassification




Source: Bhatt et al, 2021

The microsimulation is built on the multistate framework, projecting individuals' life trajectories from birth to death while embedding the natural history of cancer (see Figure 1 in Component Overview). The screening process and follow-up period are integrated into the simulation scheme.

Key features of our microsimulation are:

- A flexible simulation scheme that allows for the superimposition of different screening strategies by varying calendar years, age range, frequency, and uptake of screening. This enables projection and comparison of screening outcomes between no screening and various screening strategies, including screen-detected cancers, interval cancers, overdiagnoses, and cancer-specific deaths.
- Estimates of transition parameters and misclassification, derived from the multistate survival model, are used as input parameters for the microsimulation model.
- When panel data are not available, the microsimulation model could be used as a tool to derive estimates of transition parameters. We can use cross-sectional data with screening outcomes to estimate natural history parameters. Here, the outputs of the microsimulation model are calibrated against observed data (see Calibration), and the multistate model is fitted to the simulated data to estimate the transition parameters and misclassification (see Microsimulation).

## Current Contributors

Nora Pashayan

Ardo van den Hout

Sangyu Lee

Zhuang Xu

Ilse Cuevas Andrade

## Past Contributors

Rikesh Bhatt

Mai Ngoc Bui





# Assumption Overview

We have used the following assumptions:

- Progressive disease – all individuals with screen-detectable cancer if they live long enough will eventually present with clinical cancer. The disease progression is not reversible, that means there is no regression back to an earlier state.
- Screen-detectable state is a latent state that precedes progression to clinical diagnosis.
- The transition rates from S1 (healthy) and S2-a (screen-detectable states) to S4 (death from other causes) are the same.
- Misclassification is independent of misclassification in previous screening rounds.
- We set the age (age 40) at which we assumed all men will be in S1 (healthy / no prostate cancer) state.

University of Cambridge & University College London



# Parameter Overview

This document outlines the list of parameters used in our MSM-CanProg (UK) multistate model and microsimulation.

## List of parameters

A. For the multistate model

- Age ($t$) : age at each observation
- States ($S_t$) : the true state of subject at age ($t$)
    - State 1 (Healthy), State 2-a (Screen-detectable cancer), State 2-b (Screen-detected cancer), State 3 (Clinically diagnosed), State 4 (Death from other-causes), and State 5 (Cancer-specific death)
- Observed States ($O_t$) : the observed state at age ($t$)
    - State 1 (Negative Screen – which could be true negative or false negative), State 2-b (Screen-detected cancer), State 3 (Clinically diagnosed cancer), State 4 (Death from other-causes), and State 5 (Cancer-specific death)
- Censored state : Here an individual could be in S1 or S2-a. This occurs when an individual has not attended screening or was not offered screening (such as in the control arm, and after the screening period ends)
- $Q$: Intensity matrix
    - $q_{ij}(t)$ : the transition intensity from State $i$ to $j$, where $i \neq j$
    - $q_{ii}(t) = -\sum_{j \neq i} q_{ij}(t)$
- Gompertz distribution: $q_{ij}(t) = exp(\lambda_{ij} + \beta_{ij}t + \alpha_{ij}Z)$ for $i \neq j$, when $\beta_{ij} > 0$
- Exponential distribution: $q_{ij}(t) = exp(\lambda_{ij} + \alpha_{ij}Z)$ for $i \neq j$
    - $\lambda_{ij}$: the intercept term of transition hazards from State $i$ to State $j$
    - $\beta_{ij}$: the coefficient of age-varying hazard from State $i$ to State $j$ when the transition hazard has Gompertz distribution
    - $Z$: the vector of time-constant variables
    - $\alpha_{ij}$: the coefficient of $Z$ from State $i$ to State $j$ in the hazard function
- Misclassification probability: $e_{ij}(t) = P(O_t = j | S_t = i)$
- $v_m$: the intercept term in the misclassification model
- $r_m$: the coefficients of $t$ (age) in the age-varying misclassification model
- $P(t_0, t_1)$: the transition probability matrix between age, $t_0$ and $t_1$
- $L(\theta | t_1, \ldots, t_n)$ : the likelihood contribution across $n$ observations for an individual

B. For the microsimulation

- $t_0$: the left truncated age (40)
- $Y_0$: the calendar year at the initial age
- $S_0$: the initial state at the preset initial age. We assumed that all individuals are truly in the State 1 at $t_0$.

- $Y_i$: the calendar year at the $i$th observation
- $S_i$: the true state at the $i$th observation
- $O_i$: the observed state at the $i$th observation
- $t_{ij}$: the time to stay at State $i$ before transitioning to State $j$
- $\lambda_{ij}$: the intercept term of transition hazards from State $i$ to State $j$
- $\beta_{ij}$: the coefficient of age-varying hazard from State $i$ to State $j$ when the transition hazard has Gompertz distribution
- $\alpha_{ij}$: the coefficient of $Z$ from State $i$ to State $j$ in the hazard function
- $e_{ij}(t) = P(O_t = j|S_t = i)$: the misclassification probability
- The frequency of screening
- The total number of screening episodes
- The age-range of screening
- The range of years during which screening is offered
- Uptake of screening (compliance rate)

University of Cambridge & University College London





# Component Overview

## The estimation component

Our multistate model consists of three subcomponents (see Figure 1):

1. Natural history component (see Figure 1a – blue arrows)
2. Progression of disease from clinical diagnosis to death (see Figure 1a – red arrows)[1],
3. Progression of disease from screen-detection to death (see Figure 1b)

Figure 1: State transition diagram for the natural history model including the progression of disease from clinical diagnosis to death (Figure 1a) and the transition model from the screen detected prostate cancer to death (Figure 1b).

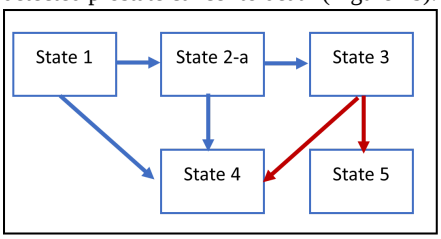



*Figure 1a*

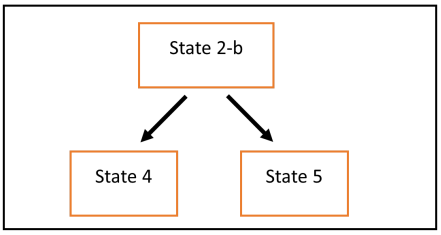



*Figure 1b*

*State 1 (healthy), State 2-a (screen detectable cancer), State 2-b (screen detected cancer), State 3 (clinically diagnosed cancer), State 4 (death from other causes) and State 5 (cancer-specific death).*

## The simulation component

Our microsimulation consists of three subcomponents:

1. Projection of individual life histories from birth to death
2. An embedded natural history subcomponent
3. Overlaid screening processes

The details are described in the Model Overview (see Microsimulation)

# References

1. Rikesh Bhatt, Ardo Van Den Hout, Nora Pashayan. A multistate survival model of the natural history of cancer using data from screened and unscreened population. Statistics in Medicine. 2021;40(16):3791–3807.

University of Cambridge & University College London





# Output Overview

This document describes the output of our multistate models and microsimulation scheme.

Table1: Outputs from each subcomponent of our multistate model and microsimulation

| **Multistate model** | | | **Microsimulation** | | |
|---|---|---|---|---|---|
| **Natural History** | **Progression after screen-detection** | **Progression after clinical diagnosis** | **Life trajectory** | **Natural History** | **Screen Scenarios** |
| Age-varying state transition parameters: S1 to S2-a, S2-a to S3, S1 to S4, S2-a to S4 | Age-varying state transition parameters from screen-detection to death: S2-b to S4 and S2-b to S5 | Age-varying state transition parameters from clinical diagnosis to death:S3 to S4 and S3 to S5 | Time span from birth to death | Screen- detectable cancer | Screen-detected Cancer |
| Misclassification (1-sensitivity) | | | Age at death from non-cancer causes | Age of entry to the screen-detectable state | Age at screen detection |
| Age of entry to the screen-detectable state | | | Deaths from non-cancer causes | Sojourn time | Misclassified state |
| Sojourn time | | | | Symptomatic / clinically diagnosed cancers | Interval cancer |
| | | | | Age at the diagnosis of clinical/symptomatic cancer | Overdiagnosed cancer |
| | | | | Age at death from cancer cause | Sensitivity of the test / screening episode |
| S1: Healthy, S2-a: Screen-detectable prostate cancer, S2-b: Screen-detected prostate cancer, S3: Clinically diagnosed prostate cancer, S4: Other-cause Death, S5: Cancer-specific death | | | | | |

University of Cambridge & University College London





# Results Overview

As an example, we present here the results of applying our microsimulation and the calibration approach to the Cluster Randomized Trial of PSA Testing for Prostate (CAP). The results are detailed in the published paper (see Martin et al, 2024).[1]

## Overview

We aimed to estimate the sojourn time, PSA screening episode sensitivity, and probability of overdiagnosis by age in the CAP trial.

CAP is a cluster-randomized trial to evaluate the effect of single invitation for PSA screening for men 50 to 69 years, from 2002 to 2009 and followed up until March 2021, on cancer specific mortality as compared to no invitation for screening.

We applied the multistate survival model with parametric hazards and the following states: healthy, screen-detectable, screen-detected, clinically diagnosed, cancer-specific death, and death from other causes, to estimate the natural history parameters and time to death after cancer diagnosis [see Figure 1 in Component Overview]. We assumed Weibull distribution for the transition between the healthy and screen-detectable states, and Gompertz distribution for all other transitions. We estimated the transition parameters and the misclassification of states (i.e. 1- sensitivity) by maximising the likelihood functions (see Transition intensity).

We simulated a cohort of three million men aged 50-69 years and followed to death using the multistate model framework with the estimates of transition parameters, and calibrated them against CAP data – prostate cancer incidence rate, cancer-specific and all-other cause mortality rates (Figure A) and age at death (Figure B). We assumed that there was no screening in the control arm.

We derived the mean sojourn time and overdiagnosis from microsimulation using the estimated transition parameters and one-off screening between ages 50 to 69 and assuming 85% of men with elevated PSA level undertake biopsy. We calculated the sojourn time as the length of time in screen-detectable state given a transition to clinically diagnosed state.

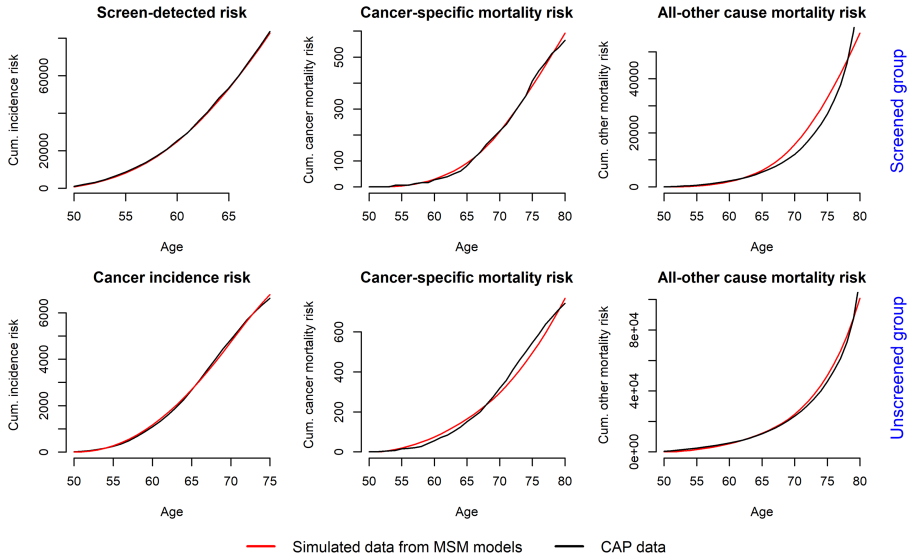
We estimated the extent of overdiagnosis as the difference in the cumulative prostate cancer incidence between screened and unscreened groups over lifetime. The probability of overdiagnosis is the fraction overdiagnosed among screen-detected cases.

---

## Results

Figures A and B are the outputs of the microsimulation model calibrated against the CAP data.

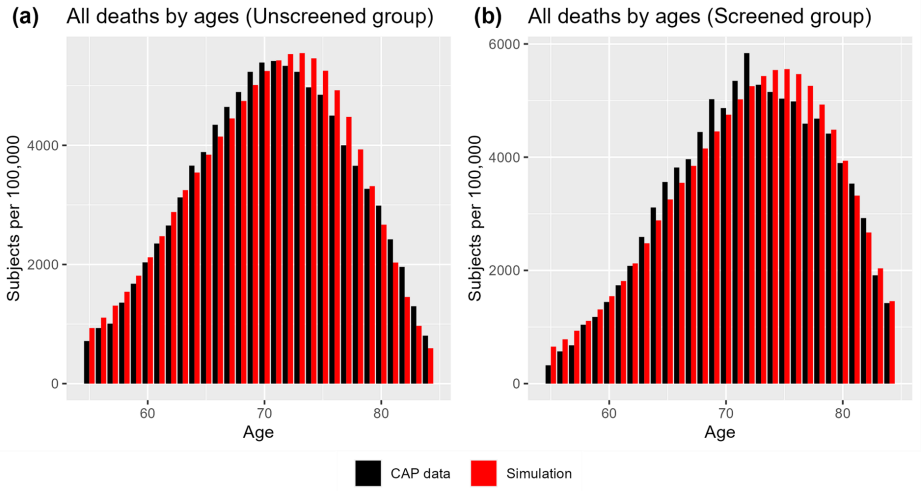
Figure A: Comparing simulated data to empirical data from CAP for the cumulative prostate cancer incidence and cancer-specific and all-other cause mortality risk among the screened men and the unscreened group.




*The estimated curves for the simulation were derived by averaging the estimates across 200 simulations of 3 million subjects aged 50-69 years, with one screening in the screening group.*

Source: Martin et al, 2024

Figure B: Comparing the number of all-cause deaths per 100,000 subjects by age between the simulated data and empirical data in the screened and unscreened groups.




*The estimated curves for the simulation were derived by averaging the estimates across 200 simulations of 3 million subjects aged 50-69 years, with one screening in the screening group.*

Source: Martin et al, 2024

The estimated curves for the simulation were derived by averaging the estimates across 200 simulations of 3 million subjects aged 50-69 years, with one screening in the screening group.

For the age group 50 to 69 years, the mean sojourn time was 13.4 years, increasing from 12.1 years in the 50-54 age group to 15.3 years in the 65-69 age group. For the age group 50 to 69 years, the probability of overdiagnosis was 15%, increasing from 9% in the 50-54 age group to 21% in the 65-69 age group. (See Table A.) The episode sensitivity (the ability of the full diagnostic process – testing and biopsy – to find cancer in the detectable preclinical phase) increased from 50.0% to 85.3% for ages 50 to 69.

Table A: Estimated mean sojourn time and probability of overdiagnosis.

| Age group | Mean Sojourn time (years) | 95% Confidence Interval (years) |
|---|---|---|

| | | |
|---|---|---|
| 50-54 | 12.1 | 12.1-12.2 |
| 55-59 | 13.2 | 13.1-13.2 |
| 60-64 | 14.2 | 14.2-14.3 |
| 65-69 | 15.3 | 15.2-15.3 |
| 50-69 | 13.4 | 13.4-13.4 |
| **Age group** | **Mean Overdiagnosis (%)** | **95% Confidence Interval (%)** |
| 50-54 | 9.2 | 8.9-9.4 |
| 55-59 | 13.3 | 13.1-13.5 |
| 60-64 | 17.1 | 17.0-17.3 |
| 65-69 | 20.8 | 20.6-21.0 |
| 50-69 | 15.0 | 14.4-15.5 |

*Overdiagnosis estimates are based on simulation of 200 cohorts of 3 million men aged 50-69 years followed to death.*

Source: Martin et al, 2024

# References

1. Richard M. Martin, Emma L. Turner, Grace J. Young, Chris Metcalfe, Eleanor I. Walsh, J. Athene Lane, et al. Prostate-Specific Antigen Screening and 15-Year Prostate Cancer Mortality: A Secondary Analysis of the CAP Randomized Clinical Trial. JAMA. 2024;331(17):1460.

University of Cambridge & University College London





# Transition Intensity

## Transition Intensity Matrix and Likelihood

### Transition intensity and probability matrix

The transition intensity matrix at the age $t$, $Q_t$, is defined as

$$Q = \begin{bmatrix} -(q_{12}(t) + q_{14}(t)) & q_{12}(t) & 0 & q_{14}(t) & 0 \\ 0 & -(q_{23}(t) + q_{24}(t)) & q_{23}(t) & q_{24}(t) & 0 \\ 0 & 0 & -(q_{34}(t) + q_{35}(t)) & q_{34}(t) & q_{35}(t) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where $q_{ij}(t)$ is the transition rate from State $i$ to State $j$ at age, $t$. Each transition hazard is parametric with exponential or Gompertz distribution. The hazard function is built in a log hazard framework with age-varying elements and potentially other variables. The Gompertz hazard function for the transition between State $i$ and $j$ is

$$q_{ij}(t) = exp(\lambda_{ij} + \beta_{ij}t + \alpha_{ij}Z) \text{ for } i \neq j,$$

where $\beta_{ij} > 0$ is the coefficient of age, $Z$ is the time-constant variable and $\alpha_{ij}$ is the coefficient of $Z$. The exponential hazard has only an intercept and time-constant terms when $\beta_{ij} = 0$.

Additionally, misclassification is considered only at the screening phase for un-detected cancers (false negatives) in the model. The misclassification rate, $e_{ij} = P(O_t = i \mid S_t = j)$, is defined

$$\begin{aligned} e_{12}(t) &= \frac{exp(v + rt)}{(1 + exp(v + rt))} \\ e_{22}(t) &= 1 - e_{21} \\ e_{ii}(t) &= 1 \text{ and } e_{ij} = 0, \text{ otherwise.} \end{aligned}$$

In a similar spirit, we constructed another multi-state model to explore the transition process after screen detection (see Figure 1-b). The transition intensity matrix is defined as

$$Q^* = \begin{bmatrix} (-(q^*_{24}(t) + q^*_{25}(t)) & q^*_{24}(t) & q^*_{25}(t) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where each transition rate is derived equivalently to the previous five-state model.

The transition probability matrix in the time interval between $t_0$ and $t_1$ for $t_0 < t_1$ is defined as

$$P(t_0, t_1) = exp(Q_{t0}(t_1 - t_0))$$

where $Q_{t0}$ is the transition matrix derived at $t_0$.

We applied a piecewise constant approximation using a grid with the regular length, $l$, and the approximate likelihood can be written as

$$P(t_0, t_1) = P(t_0, t_0 + l)P(t_0 + l, t_0 + 2l) \cdots P(t_0 + nl, t_1)$$

where $0 \leq n < \frac{(t_1 - t_0)}{l}$.

---

### Likelihood function

The data for our model include left truncated, censored, interval-censored, right censored, and misclassified observations.

The likelihood contribution for an individual, who has N observations, at ages $t_1$ to $t_N$, is given by Bhatt et al, 2021[1]:

$$L(\theta \mid t_1, \ldots, t_n) = \epsilon^T (\prod_{\pi=1}^{n} L_i(\theta | t_{i-1}, t_i))\ 1,$$

where $\theta$ is the set of parameters of the model and $t_0$ is the left truncated age (40) at which we assumed all mean are in the healthy state (S1). Additionally, $\epsilon^T = (1, 0, 0, 0)$, which implies that all men are healthy at the initial age, $t_0$, and $1^T = (1, 1, 1, 1)$ to sum up the likelihood contribution of possible states at the last observation, including censored states.

Misclassification can be incorporated into the model, when State 2 (Screen detectable state) can be misclassified as State 1 (Healthy). The probability of transition from State 1 or 2 to observing state $o_n$ is

$$P(O_n = o_n | S_{n-1} = 1 \text{ or } 2) = \sum_{s_n=1,2} P(S_n = s_n | S_{n-1} = 1 \text{ or } 2) P(O_n = o_n | S_n = s_n).$$

The likelihood matrix, $L_n$, when the observed state is 1 or 2, can be written as

$$L_n = P(t_n, t_{n-1}) E_(t_n)$$

where $P(t_{n-1}, t_n)$ is the transition probability matrix in the interval, $(t_{n-1}, t_n)$. Additionally, the misclassification matrix is

$$E_{t_n} = \begin{bmatrix} e_{1o_n}(t) & 0 & 0 & 0 & 0 \\ 0 & e_{2o_n}(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where $e_{ij}(t) = P(O_j = o_j | S_i = s_i)$.

In our model, we assumed that all men are healthy at $t_0$ and individuals who had already clinical cancer or had died before entry to the study are censored from the analysis. Therefore, the probability of the first observation with left truncation is

$$\begin{aligned} P(O_1 = o_1 | S_0 = 1, S_1 \notin \{3, 4, 5\}) &= \frac{P(O_1 = o_1 \cap S_1 \notin \{3, 4, 5\} | S_0 = 1)}{P(S_1 \notin 3, 4, 5 | S_0 = 1)} \\ &= \sum_{i=1,2} \frac{P(S_1 = s_1 | S_0 = 1) P(O_1 = o_1 | S_1 = s_1)}{P(S_1 \notin 3, 4, 5 | S_0 = 1)}. \end{aligned}$$

Therefore, the contribution of the first observations, $L_1$, is

$$L_1 = \begin{bmatrix} \frac{1}{P_{11}(t_0, t_1) + P_{12}(t_0, t_1)} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{P_{22}(t_0, t_1)} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} P_0 E_1,$$

where $P_0$ is the transition probability matrix at $t_0$ and $E_1$ is the misclassification matrix at the first observation, $t_1$.

Clinical cancers and deaths are events whose times are observed exactly, and the likelihood contribution of state $d$ is

$$L_n = P(t_n, t_{n-1}) Q(t_{n-1}) \delta_d$$

where

$$\delta_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\delta_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$\delta_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The likelihood function, $L_K$, of $K$ individuals is a product of individual likelihood contributions:

$$L_K(\theta; t_1, \ldots, t_N) = \prod_{i=1}^{K} L^i(\theta; t_1, \ldots, t_{n_i})$$

where $L^i(t_1, \ldots, t_{n_i})$ is the likelihood contribution from the ith individual.

# References

1. Rikesh Bhatt, Ardo Van Den Hout, Nora Pashayan. A multistate survival model of the natural history of cancer using data from screened and unscreened population. Statistics in Medicine. 2021;40(16):3791–3807.

University of Cambridge
& University College
London





# Microsimulation

## Model Overview – Microsimulation

We simulated the life trajectory of individuals from bith to death while embedding the natural history of cancer, following the steps given by Bui et al. (Under Review):

1. Set the initial calendar year ($Y_0$) at the pre-set initial age ($t_0$=40) when we assume that all men are in State 1 (healthy).

2. Men in State 1 can transition to either State 2 or 4. The transition time to event can be simulated from the transition distributions respectively, $t_{12} \sim Gomp(t_0; \lambda_{12}, \beta_{12})$ and $t_{14} \sim Gomp(t_0; \lambda_{14}, \beta_{14})$. If $t_{12} < t_{14}$, $t_1 = t_0 + t_{12}$ and proceed to Step 3. If $t_{12} > t_{14}$, $t_1 = t_0 + t_{14}$ and proceed to Step 5

3. The next state from State 2 can be either 3 or 4. Equivalently, the time to event is calculated as $t_{23} \sim Exp(t_1; \lambda_{12})$ and $t_{24} \sim Gomp(t_1; \lambda_{14}, \beta_{14})$. If $t_{23} < t_{24}$, $t_2 = t_1 + t_{23}$ and go to Step 4. If $t_{23} > t_{24}$, $t_2 = t_1 + t_{24}$ and proceed to Step 5.

4. The next state from State 3 can be either 4 or 5. The time to event is calculated as $t_{34} \sim Gomp(t_2; \lambda_{34}, \beta_{34})$ and $t_{35} \sim Gomp(t_2; \lambda_{35}, \beta_{35})$. If $t_{34} < t_{35}, t_3 = t_2 + t_{34}$. Otherwise, $t_3 = t_2 + t_{35}$. In either case, process to Step 5.

5. End the simulation and record the transition times along with the corresponding states.

The transition time to event is simulated from the assumed hazard distribution conditioned at time $t$. For example, when the transition has the Gompertz distribution with the parameters $\alpha$ and $\beta$ at $t_n$ : $Gomp(t; \alpha, \beta)$ , the simulated survival time to event can be written

$$t_{n+1} = (\log(exp(\alpha + \beta t_n) - \beta \log U) - \alpha)/\beta$$

where $U \sim Unif(0, 1)$.

We superimposed a screening strategy onto the simulated life-trajectory and natural history of prostate cancer. This included defining age range of screening, frequency of screening, and calendar year of starting screening. We then compared the outcomes of different screening strategies with each other and with no screening.

University of Cambridge & University College London





# Calibration

## Model Overview – Calibration

For calibration of the simulated data against empirical data, we have used calibration, by comparing the simulated estimates of cumulative incidence of clinically diagnosed cancers, all-other cause mortality, cancer-specific mortality rate and distribution of age at death (see Box 1). Using the calibration approach we aimed to estimate the set of parameters related to the natural history of prostate cancer, transitions following screen detection and clinical diagnosis, as well as misclassification, that align with empirical data.

Box 1: Steps of the calibration process (This box will be released online following the publication of Bui et al. (Under review))

*University of Cambridge & University College London*





# Key References

Rikesh Bhatt, Ardo Van Den Hout, Nora Pashayan. A multistate survival model of the natural history of cancer using data from screened and unscreened population. Statistics in Medicine. 2021;40(16):3791–3807.

Rikesh Bhatt, Ardo van den Hout, Antonis C Antoniou, Mitul Shah, Lorenzo Ficorella, Emily Steggall, et al. Estimation of age of onset and progression of breast cancer by absolute risk dependent on polygenic risk score and other risk factors. Cancer. 2024;130(9):1590.

Cuevas Andrade, Nora Pashayan. Parsimonious multistate models using longitudinal data and time-dependent covariates: applications to a liver cirrhosis clinical trial. Under review.

Richard M. Martin, Emma L. Turner, Grace J. Young, Chris Metcalfe, Eleanor I. Walsh, J. Athene Lane, et al. Prostate-Specific Antigen Screening and 15-Year Prostate Cancer Mortality: A Secondary Analysis of the CAP Randomized Clinical Trial. JAMA. 2024;331(17):1460.

Mai Ngoc Bui, Ardo van den Hout, Rikesh Bhatt, Nora Pashayan. Non-homogeneous multistate partial Markov models: A simulation scheme for evaluating cancer screening strategies. Under review.