



# CISNET PROSTATE CANCER COLLABORATORS

**Important note:** This document will be updated periodically. The most current version is available at <http://cisnet.cancer.gov/profiles>. Note that unlike most PDF documents, the CISNET model profiles are not suitable for printing as they are not typically written or read in sequential fashion.

We recommend you let your interests guide you through this document, using the navigation tree as a general guide to the content available.

The intent of this document is to provide the interested reader with insight into ongoing research. Model parameters, structure, and results contained herein should be considered representative but preliminary in nature.

We encourage interested readers to contact the contributors for further information.

|   |   |
|---|---|
|    | Erasmus MC  |
|   | Fred Hutchinson<br>Cancer Research Center (PSAPC) |
|  | Fred Hutchinson<br>Cancer Research Center (PCSIM) |
|  | University of Michigan                            |



FLEXKB DOCUMENT  
Version: HI.001.12212009.86039  
Document generated: 12/21/2009

# ERASMUS MC (PROSTATE)

**Important note:** This document will be updated periodically. The most current version is available at <http://cisnet.cancer.gov/profiles>. Note that unlike most PDF documents, the CISNET model profiles are not suitable for printing as they are not typically written or read in sequential fashion.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

We recommend you let your interests guide you through this document, using the navigation tree as a general guide to the content available.

The intent of this document is to provide the interested reader with insight into ongoing research. Model parameters, structure, and results contained herein should be considered representative but preliminary in nature.

We encourage interested readers to contact the contributors for further information.

Go directly to the: [Reader's Guide](#).



# READERS 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



# MODEL PURPOSE

## SUMMARY

The MISCAN micro-simulation model is used to analyze the effect of PSA screening on prostate cancer incidence and mortality. This document summarizes the objectives in developing a prostate cancer simulation model.

## PURPOSE

The MISCAN computer simulation model has been developed for estimating the effect of cancer screening in a dynamic population, to explain results of cancer screening trials, to predict and compare the (cost-) effectiveness of different screening policies, and to monitor the results of population screening programs.

The objective of the prostate cancer model is to quantify the role of PSA screening in prostate cancer incidence and mortality. The prostate cancer screening model is used to simulate the results of the Rotterdam section of the ERSPC trial as the incidence and mortality in the US population. By calibrating the model to the trial data and baseline incidence, parameters for the natural history have been estimated. Using the MISCAN model, based on the results of ERSPC Rotterdam, we try to understand the trends in the US and how they differ from European or Dutch conditions.

The models are used to determine optimal screening ages and test intervals and to calculate cost-effectiveness of various screening policies, compared with a situation without screening. Also, the models are used to estimate unobservable processes and variables (natural history of the disease, the amount of overdiagnosis and lead time) in the ERSPC trial as well as the US population.



# MODEL OVERVIEW

## SUMMARY

This document provides an overview of the modeling effort, and describes the model in general terms.



Readers Guide  
Model Overview  
Assumption Overview  
Parameter Overview  
Component Overview  
Output Overview  
Results Overview  
Key References

## PURPOSE

In the MISCAN model knowledge on natural history of prostate cancer, screening and treatment obtained from randomized controlled trials and observational studies are integrated. In this way MISCAN can be helpful in analyzing and explaining results of cancer screening trials, predicting the (cost-) effectiveness of different screening policies and predicting the potential of present and new interventions on future national trends. See also [Model Purpose](#).

## BACKGROUND

The MISCAN computer program has been used for building screening models for cancers of breast, prostate, cervix, colon and lung<sup>1234</sup>. The MISCAN prostate cancer model has been used to model trends of prostate cancer incidence and mortality in the ERSPC-trial Rotterdam, and in the ERSPC-trial Sweden, the Dutch population and in the US population. With these models it is possible to compare trends of prostate cancer with and without treatment and screening.

The ERSPC model has been used to predict mean lead times and overdiagnosis rates, associated with different screening programs<sup>15</sup>. It has also been used to provide epidemiological evidence of dedifferentiation as a mechanism of progression in prostate cancer<sup>6</sup>.

## MODEL DESCRIPTION

MISCAN model is a micro-simulation model. Using the model inputs, independent life histories are generated including a possible cancer history, the effects of treatment and the effects of early detection by screening. The MISCAN-prostate model contains four primary components:

1. Demography component
2. Natural History component
3. Treatment component
4. Screening component

First the demography component simulates a population of individual life histories, according to the demography parameters. Each individual in the population consists of a date of birth and age of death.

Subsequently the Natural history component simulates prostate cancer histories for each individual life history separately. Some individuals will have no prostate cancer in their life and others will have prostate cancer in their life. Once the individual has prostate cancer the cancer can progress to different preclinical states. In the preclinical phase the tumor is asymptomatic, but can be detected by screening. In this definition,

the preclinical phase does not only depend on biological processes, but also on the state of medical technology. Eighteen preclinical detectable states are defined in combinations of clinical T-stage (T1, T2 and T3), Gleason grade (well, moderately, and poorly differentiated) and metastatic stage (local-regional and distant). From each preclinical detectable state the cancer can progress to the clinical disease state, which implies that cancer is diagnosed because of symptoms.

In the third part the treatment component simulates the life history after clinical diagnosis. Detection with cancer is followed by treatment and possibly prostate cancer death. Different treatments have their treatment-specific survival of prostate cancer death.

The screening component super-imposes screening on the life histories in the absence of screening. Screening tests applied to a person in a preclinical disease state may result in detection and alter the life history of this individual. We assume that the consequences of early detection by screening are that a part of the screen-detected men is cured of prostate cancer and will die from other causes. For the other part of the screen-detected men early detection does not alter the life history.

See [Component Overview](#) for a more elaborated description of these components.

## CONTRIBUTORS

Rob Boer  
Gerrit Draisma  
Eveline Heijnsdijk  
Harry de Koning  
Elisabeth Wever

## REFERENCES:

- <sup>1</sup> Draisma, G, Boer, R, Otto, SJ, van der Crujisen, IW, Damhuis, RA, Schröder, FH, de Koning, HJ. "Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer" in *J Natl Cancer Inst* 2003; 95: : 868-878
- <sup>2</sup> Tan, SYGL, van Oortmarssen, GJ, Piersma, N. "Estimating Parameters of a Microsimulation Model for Breast Cancer Screening" in *Annals of Operations Research* 2003; 119: : 43-61
- <sup>3</sup> Loeve, F, Boer, R, van Oortmarssen, GJ, van Ballegooijen, M, Habbema, JD. "The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening." in *Comput Biomed Res* 1999; 32: : 13-33
- <sup>4</sup> Akker, ME van den, van Ballegooijen, M, van Oortmarssen, GJ, Boer, R, Habbema JD. "Cost-effectiveness of cervical cancer screening: comparison of screening policies." in *J Natl Cancer Inst* 2002; 94: : 193-204
- <sup>5</sup> Draisma, G, Etzioni, R, Tsodikov, A, Mariotto, A, Wever, E, Gulati, R, Feuer, E, de Koning, HJ "Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context" in *J Natl Cancer Inst* 2009; 101: : 374-383
- <sup>6</sup> Draisma, G, Postma, R, Schröder, FH, van der Kwast, TH, de Koning, HJ. "Gleason score, age and screening: modeling dedifferentiation in prostate cancer" in *Int J Cancer* 2006; 119: : 2366-2371

# ASSUMPTION OVERVIEW

## SUMMARY

The assumptions made for the MISCAN model are described in this section.

## BACKGROUND

The MISCAN prostate cancer model can be used to simulate prostate cancer screening and treatment policies in a dynamic population (see [Model Purpose](#)), based on assumptions on demography, natural history of prostate cancer, treatment and screening. Most of the assumptions arise from the unobservable part in the screening and treatment of prostate cancer, the natural history of the disease and the effect of screening on improvement of survival

## ASSUMPTION LISTING

The MISCAN ERSPC and US population model use the following assumptions, categorized by model component (see [Component Overview](#)):

### Demography

- a. The (country specific) life table is the same for all men in the same birth cohort
- b. Death from prostate cancer and death from other causes are independent
- c. The life time prostate cancer risk is the same for all men in a the same birth cohort

### Natural history

- a. Tumor onset:

Tumors are assumed to initiate with the same age specific initiation rate for all men.

- b. Progression of disease:

The tumor starts in the preclinical phase. Progression is defined by a matrix of transition probabilities between states, and dwelling time distributions for the time spent in each state. The dwelling times are determined by Weibull distributions. Transition probabilities and dwelling time distributions are age-dependent. A correlation between duration in subsequent states is assumed.

In the preclinical phase the cancer can be detected by screening. There are eighteen preclinical detectable states which are derived from combinations of clinical T-stage (T1, T2 and T3), Gleason grade (well, moderately, and poorly differentiated) and metastatic stage (local-regional and distant).

- c. Clinical detection:

From each preclinical detectable state the cancer can progress to the clinical disease state, which implies that cancer is diagnosed because of symptoms. The progression to the clinical state is defined by the matrix of transition probabilities between states, and dwelling time distributions determined by Weibull distributions. To explain a higher incidence and a more favorable stage distribution in the control arm of the trial compared to the base population, the population in 1991, it is assumed that in the trial population (during trial period) prostate cancer was clinically diagnosed earlier than in the baseline situation in 1991. Specifically, it is assumed that the hazard of being clinically diagnosed given that you are in the preclinical disease state in the trial population compared to the baseline situation is larger. This difference can for instance be attributed to contamination (screening in the control arm) or to changes in clinical practice leading to earlier diagnosis e.g to the use of PSA testing for symptomatic disease in a clinical setting.

### **Treatment**

After prostate cancer diagnosis the treatments radical prostatectomy, radiation therapy and active surveillance can be assigned.

#### a. Treatment dissemination:

ERSPC model: treatment is modeled as a multinomial logit model with covariates age, T-stage and Gleason score at diagnosis. The categories of the multinomial model are radical prostatectomy, radiation therapy and active surveillance. The parameter estimates are based on data of the ERSPC trial section Rotterdam from the year 2000.  
US model: treatment is modeled as a multinomial logit model with covariates age, year and grade at diagnosis. The categories of the multinomial model are radical prostatectomy, radiation therapy and active surveillance. The parameter estimates are based on SEER and Ca PSURE data. Conditional on patient's characteristics available at diagnosis and primary therapy hormone therapy is assigned by a logistic model.

#### b. Survival after treatment:

Baseline survival:

The baseline survival has been estimated from SEER (Surveillance, Epidemiology and

End Results) data in the pre-PSA era, specifically of cases diagnosed between 1983 and 1986. The survival curves were modeled using Poisson regression with grade, stage, age and treatment type as explanatory variables. To assign the survival curves in our model we assumed that Gleason score 7 or less than 7 corresponds to grade well/moderately differentiated and that Gleason score more than 7 corresponds to grade poor/undifferentiated.

Treatment effect:

The time of death of prostate cancer is defined by the survival curve of the corresponding treatment. Bill-Axelsson et al.<sup>1</sup> showed for men with clinically diagnosed localized prostate cancer a relative risk of 0.65 for the efficacy of radical prostatectomy compared to the efficacy of watchful waiting. Considering this result, we assume that men receiving watchful waiting have the baseline survival and that those men receiving radical prostatectomy and radiation therapy have a relative risk of 0.65 compared to watchful waiting for local-regional cancers. For distant prostate cancer it is assumed that treatment has no effect on the survival, implying that irrespective of the treatment type all men diagnosed with prostate cancer in the distant stage have a survival generated from the corresponding baseline survival curve.

### Screening

a. Attendance to screening:

In the model, men can only be screened when they are still alive at the moment of the screen and when they have not already been diagnosed with prostate cancer.

ERSPC model: Data of the Rotterdam section of the ERSPC trial have been used to simulate the age and year specific attendance rate. Also, the attendance to the screening is dependent on whether or not the person attended the last screening.

US model: For the dissemination of PSA screening, we used the results of Mariotto et al.<sup>2</sup>, who retrospectively constructed PSA screening histories in the population by use of survey data from the 2000 National Health Interview Survey and claims data from the linked SEER-Medicare database (<http://healthservices.cancer.gov/seermedicare/>).

b. Sensitivity of the test:

PSA screening and subsequent biopsy are modeled as one single test. The test has a T-stage-dependent sensitivity. These parameters are estimated using data of the ERSPC trial Rotterdam and the US population. We do not model digital rectal exam (DRE) explicitly.

c. Effect on survival because of early detection by screening:

We assume that a part of the screen-detected men is cured from cancer and that for the other part detection does not alter the life history. The cure rate is estimated by assuming a mortality reduction of 27% in the ERSPC model after a follow-up of 9 years for men who were actually screened. The mortality reduction of 27 % was observed in

the ERSPC-trial<sup>3</sup>.

### Parameter estimation

Model parameters for the natural history component and the test-sensitivity are estimated as follows: A model is constructed for a specific situation, such as prostate cancer incidence in the US or both arms of the ERSPC trial Rotterdam. Parameters are then estimated by numerical minimization of the deviance between observed numbers of cases and the corresponding numbers predicted by the model. Deviances are calculated assuming Poisson likelihood for incidence data or a multinomial likelihood for stage distribution data. For the minimization an adapted version of the simplex optimization method of Nelder and Mead is used<sup>4</sup>. Optimization is initiated with small sample sizes and repeated with larger sample sizes (up to 1 million) when optimization progress is no longer statistically significant.

ERSPC model: Estimates of natural history parameters and test sensitivities were obtained using observed detection rates and interval cancer rates and stage distributions in the ERSPC-trial Rotterdam.

US model: US-specific estimates of test-sensitivities were obtained using observed age-specific incidence and age-specific stage distribution (local/regional vs distant). For parameter estimation data of men 50 to 84 years old diagnosed in 1975 to 2000 from the SEER registry were used.

### REFERENCES:

- 1 Bill-Axelsson A, Holmberg L, Filen F, Ruutu M, Garmo H, Busch C, et al. "Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial" in *J Natl Cancer Inst* 2008; 100: : 1144-1154
- 2 Mariotto, AB, Etzioni, R, Krapcho, M, Feuer, EJ. "Reconstructing PSA testing patterns between black and white men in the US from Medicare claims and the National Health Interview Survey" in *Cancer* 2007; 109: : 1877-1886
- 3 Schröder, FH, Hugosson, J, Roobol, MJ, et al "Screening and prostate-cancer mortality in a randomized European study" in *N Engl J Med* 2009; 360: : 1320-1328
- 4 Bazaraa, MS, Sherali, HD, Shetty, CM. "Nonlinear programming: theory and algorithms" 1993; : 353

# PARAMETER OVERVIEW

## SUMMARY

Provides a complete overview of the parameters used to quantify the MISCAN-Prostate model.

## BACKGROUND

The MISCAN-Prostate model consists of four basic components: The demography component, the natural history component, the treatment component and the screening component. Each component has its own set of parameters.

## PARAMETER LISTING OVERVIEW

### Demography Parameters

1. number of birth cohorts
2. proportion of the population in each birth cohort
3. for each birth cohort parameters of its birth table
4. for each birth cohort the parameters of its life table

### Natural history Parameters

1. parameters for the age specific distribution of onset of the first screen detectable state
2. for each birth cohort the life time prostate cancer risk
3. parameters for the duration distribution in each preclinical state
4. parameters for the transition probability in each preclinical state
5. parameters for additional clinical diagnosis
6. correlation between duration in subsequent states
7. parameters for survival after clinical diagnosis by age at diagnosis, year of diagnosis, grade and stage of disease at diagnosis

### Screening Test Parameters

1. parameters for the dissemination of PSA screening by age and year
2. test-sensitivity parameters
3. cure rate parameters defining the benefit because of early detection

### Treatment parameters

1. parameters for the dissemination of treatment by age at diagnosis, year of diagnosis, grade and stage of disease at diagnosis
2. hazard ratios associated with initial treatments i.e. radical prostatectomy, radiation therapy and radiation therapy combined with hormones



# COMPONENT OVERVIEW

## SUMMARY

An overview of the major components in the MISCAN-Prostate model.

## OVERVIEW

As described in the Model overview document, the MISCAN-prostate model contains four primary components: Demography, Natural History, Treatment and Screening.

## COMPONENT LISTING

### **Demography Component**

The demography component simulates a population of individual life histories, according to the demography parameters. The demography parameters are:

1. birth table parameters
2. life table parameters

Each individual in the population consists of a date of birth and age of death. It is possible to define a dynamic population of all ages, which can be adjusted for different countries. Also it is possible to define a cohort of people with the same age or age range.

### **Natural History Component**

The cancer related event history is defined by a sequence of disease states and the ages at which these states are entered. The life histories are generated by a semi-Markov process, defined by a matrix of transition probabilities between states, and dwelling



time distributions for the time spent in each state. The disease history is divided in a preclinical phase and a clinical phase. The preclinical phase corresponds to the asymptomatic states, that do not lead to clinical diagnosis, but can be detected by screening. In this definition, the preclinical phase does not only depend on biological processes, but also on the state of medical technology. Its parameters have to be estimated from indirect evidence. In the Miscal prostate cancers model there are eighteen preclinical detectable states which are derived from combinations of clinical T-stage (T1, T2 and T3), Gleason grade (well, moderately, and poorly differentiated) and metastatic stage (local-regional and distant). The progression through these states is illustrated in Figure 1. From each preclinical detectable state the cancer can progress to the clinical disease state, which implies that cancer is diagnosed because of symptoms.

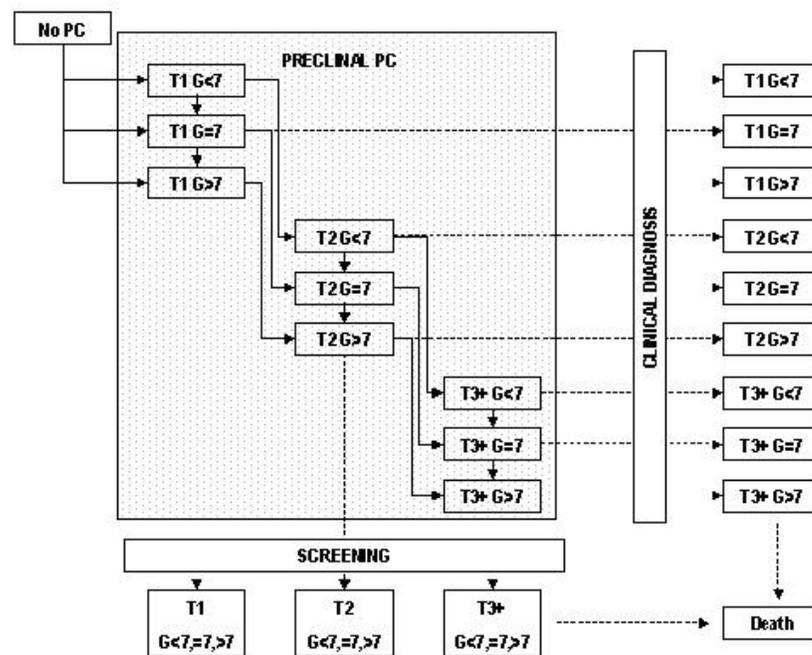


Figure 1: The MISCAN prostate cancer model. Prostate cancer develops from no prostate cancer via 1 or more screen-detectable preclinical stages to a clinically diagnosed cancer. There is also a distinction between local en metastatic stage, but for simplicity not illustrated. Screening is superimposed on the life histories in the absence of screening. Screening may detect cancers earlier in one of the preclinical screen-detectable states.



### **Screening Component**

Screening is super-imposed on the life histories in the absence of screening. Screening tests applied to a person in a preclinical disease state may result in detection and alter his life history. A screening test is defined by its stage-specific sensitivity. A screening policy, is defined by the tests used, attendance rate and screening ages. Screening ages may be selected at regular intervals, or stochastically, allowing the modeling of both regular screening as in trials or screening programs and opportunistic screening. Screen detection may alter the cause of events. We assume that the consequences of early detection by screening is that a part of the screen-detected men is cured of prostate cancer and that for the other part detection does not alter the life history.

### **Treatment Component**

The life history after clinical diagnosis is defined by stage-specific survival functions. Detection with cancer is followed by treatment and a survival of prostate cancer death. Different treatments can be assigned and the different treatments have their treatment-specific survival of prostate cancer death.

# OUTPUT OVERVIEW

## SUMMARY

This document describes the main outputs of the Miscan microsimulation model.

## OUTPUT LISTING

The main outputs of the model are:

1. Projected incidence by age, year, clinical T-stage, Gleason score, metastatic state and mode of detection.
2. Treatment assignment by age, year, clinical T-stage, Gleason score, metastatic state and mode of detection.
3. Five-, 10-, 15-, and 20-year survival by age and stage at diagnosis and by treatment.
4. Mortality by age and year at death and cause of death and by clinical T-stage, Gleason score, metastatic state and mode of detection.
5. Number of PSA tests performed by age and year and total number of men screened.
6. Detection rate by age, year and screen round (first screen or subsequent screen). Detection rate is defined as cancers detected / # of men screened.
7. Overdiagnosis rates by age, year of diagnosis, clinical T-stage, Gleason score. An individual is overdiagnosed if he is screen detected but would not have been diagnosed in his lifetime in the absence of screening.
8. Mean lead time. Lead time is defined as the amount of time, in years, between prostate cancer detection and either clinical diagnosis in the absence of screening or death by other causes. Lead time is calculated for all screen-detected cancers and for the screen-detected relevant (non-overdiagnosed) cases only.
9. Mean sojourn time: Time from disease onset to clinical diagnosis.
10. Total life years of the population and total life years in a particular interval (e.g. from birth to diagnosis, from diagnosis till death)

All outputs (except for overdiagnosis and lead time) are projected in the presence and in the absence of screening.



# RESULTS OVERVIEW

## SUMMARY

This document lists various results generated by the model.

## OVERVIEW

First a model was made for the screen arm and control arm of the European Randomized Study of Screening for Prostate Cancer (ERSPC) trial section Rotterdam. This model was adjusted to project the US population, by adjusting some input parameters and fitting the model to the US incidence and stage distribution. With both models lead time and overdiagnosis were estimated.

## RESULTS LIST

[Model ERSPC trial Rotterdam](#)

[Model US population](#)

[Lead Time](#)

[Overdiagnosis](#)

**Erasmus MC**



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

# MODEL ERSPC

## Model for ERSPC trial Rotterdam

### Summary

This document describes the model made for the European Randomized Study of Screening for Prostate Cancer (ERSPC) trial.

### Overview

A prostate cancer model has been made and validated using the results of the ERSPC trial Rotterdam and baseline incidence and stage distribution data of the Netherlands.

### Methods

Based on the results of the Rotterdam section of the ERSPC trial and baseline incidence and stage distribution a model was made which could accurately predict the results of the first two screening rounds<sup>1,2</sup>. The trial started in Rotterdam in 1994 and included 21166 men aged 55-74 in the control arm and 21210 men in the screen arm. In the first years a PSA cut-off of 4 ng / ml was used as an indication for biopsy, later this was changed to 3 ng / ml. The model was also validated with the baseline incidence in the Netherlands (1991) and the stage distribution of clinically diagnosed cancers (1991-1993), of the Rotterdam Cancer Registry.

### Results

After fitting the parameters (transition probabilities, dwelling times, test sensitivities), the model could predict the observed baseline values accurately (Table 1):

Table 1: Baseline incidence in the Netherlands 1991 and stage distribution 1991-1993, compared with the model predictions.

#### Incidence per 1000 men years

| Age group                | Observed | Model prediction |
|--------------------------|----------|------------------|
| 50-55                    | 0.14     | 0.16             |
| 55-60                    | 0.36     | 0.46             |
| 60-65                    | 1.19     | 1.13             |
| 65-70                    | 2.59     | 2.45             |
| 70-75                    | 4.50     | 4.49             |
| 75-80                    | 6.57     | 6.68             |
| 80-85                    | 7.98     | 8.33             |
| 85+                      | 8.52     | 7.71             |
| 55-75 (trial population) | 1.86     | 1.91             |

#### Stage distribution (%)

| Stage     | Observed | Model prediction |
|-----------|----------|------------------|
| Localized | 58.03    | 57.75            |
| Regional  | 18.81    | 19.44            |
| Distant   | 23.15    | 22.85            |

The detection rate in the screen arm per round is compared with the model predictions (Figure 1)

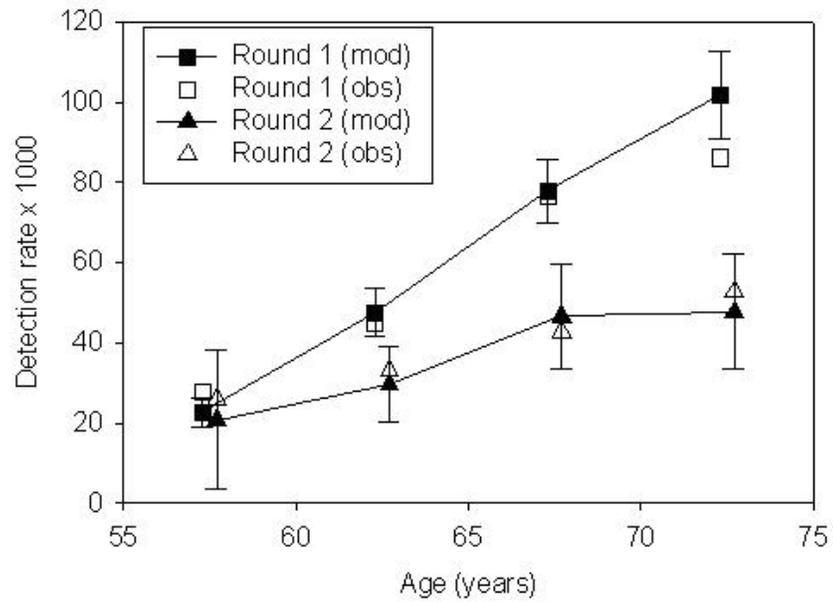


Figure 1: Detection rate in the screen arm, observed and predicted by the model.

The stage distribution and Gleason score compared with the model are presented in Figure 2.

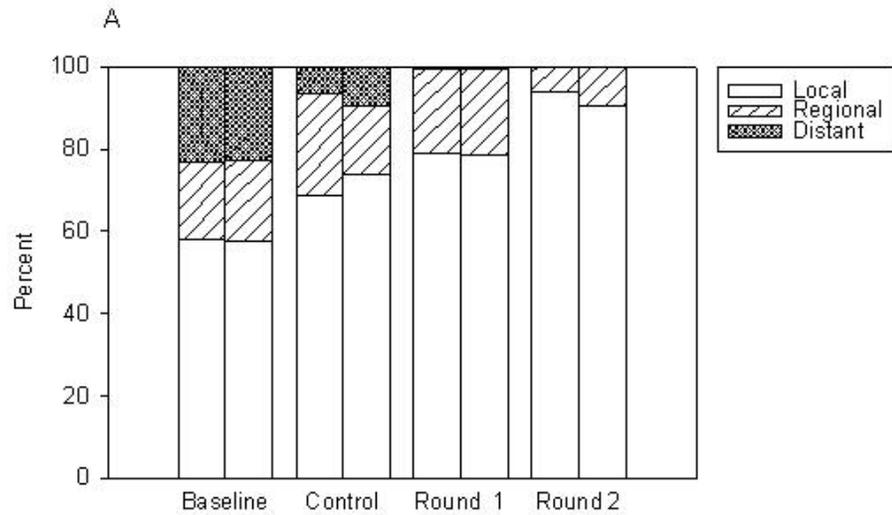


Figure 2A: Stage distribution of baseline (1991) level, in the control arm and in the first and second round of the trial. Left bar of each pair is the observed value, right bar the model prediction.

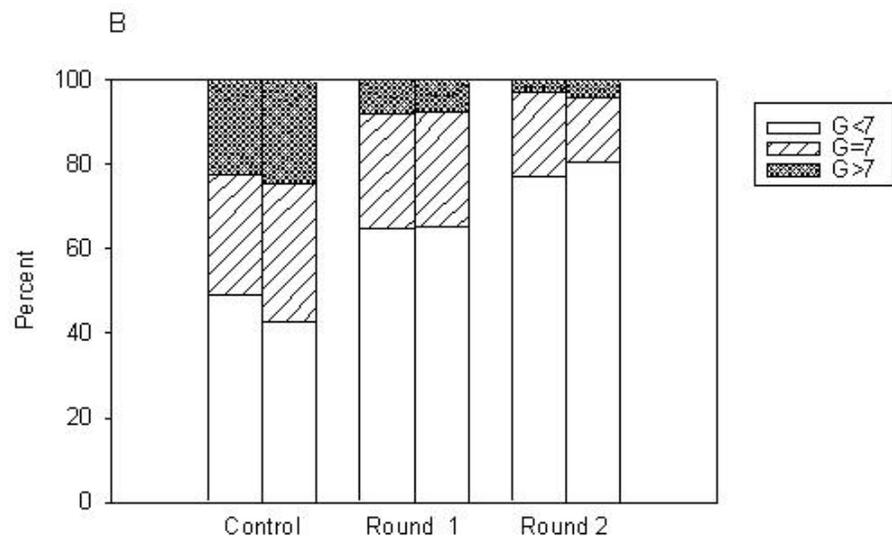


Figure 2B: Gleason score distribution in the control arm and in the first and second round of the trial. Left bar of each pair is the observed value, right bar the model prediction.



### **Discussion**

The model, fitted to the baseline and trial data reproduced the essential characteristics of the observed data on clinical incidence, detection rates and tumor stage and Gleason score distributions. However, observed incidence and detection rates in the older age groups in the trial were significantly lower than predicted by the model. These results suggest a selection effect in the older age groups (older participants in the trial could be healthier than average).

### **Conclusion**

The model could acceptably well project the observed baseline incidence and stage distribution and the results of the first two rounds of the ERSPC trial section Rotterdam.

### **REFERENCES:**

- <sup>1</sup> Draisma, G, Boer, R, Otto, SJ, van der Cruijsen, IW, Damhuis, RA, Schröder, FH, de Koning, HJ. "Lead times and overdetected due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer" in J Natl Cancer Inst 2003; 95: : 868-878
- <sup>2</sup> Draisma, G, Postma, R, Schröder, FH, van der Kwast, TH, de Koning, HJ. "Gleason score, age and screening: modeling dedifferentiation in prostate cancer" in Int J Cancer 2006; 119: : 2366-2371



# MODEL US

## Summary

This document describes the model made for the US population.

## Overview

The model of the ERSPC trial has been modified to a model for the US population.

## Methods

The validated MISCAN-model developed for the progression of prostate cancer and screening in the ERSPC-trial Rotterdam is adjusted for the US situation by adapting the population and the PSA testing practice<sup>1</sup>. Also, an estimated extra stage-specific risk of clinical diagnosis has been added, implying an earlier diagnosis of prostate cancer in the absence of screening in the United States. The model is calibrated to the SEER 9 incidence from 1985 to 2000, as well as stage distribution data.

## Results



## REFERENCES:

- <sup>1</sup> Draisma, G, Etzioni, R, Tsodikov, A, Mariotto, A, Wever, E, Gulati, R, Feuer, E, de Koning, HJ "Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context" in J Natl Cancer Inst 2009; 101: : 374-383

# LEAD TIME

## Lead time

### Summary

This document describes the estimates of lead time using the model based on the ERSPC Rotterdam trial and the model for the US population.

### Overview

Lead time, the time that screening advances cancer diagnosis, is estimated with the use of the validated models for the ERSPC trial and the US population.

### Methods

For this study lead time is defined as the amount of time, in years, between prostate cancer detection and either clinical diagnosis in the absence of screening or death by other causes<sup>12</sup>. The model for the ERSPC trial as well as the model for the US population has been used to estimate lead time. The lead time was calculated for various screen programs, for all screen detected cancers and for screen-detected relevant (non-overdiagnosed) cancers only.

### Results

The lead time is dependent on the screening program (Table 1).

Table 1. Mean lead time for various screening programs using the ERSPC model.

| screen program       | mean lead time (years) |                |      |
|----------------------|------------------------|----------------|------|
|                      | all cases              | relevant cases |      |
| age                  |                        |                |      |
| single               | 55                     | 12.3           | 12.8 |
| 60                   | 11.0                   | 11.5           |      |
| 65                   | 9.5                    | 10.0           |      |
| 70                   | 7.7                    | 8.1            |      |
| 75                   | 6.0                    | 6.2            |      |
| interval             | every year, 55-67      | 12.3           | 13.7 |
| every year, 55-75    | 11.6                   | 13.4           |      |
| every 4 years, 55-67 | 11.2                   | 12.3           |      |
| every 4 years, 55-75 | 10.3                   | 11.7           |      |

For the US population, the estimated lead times were lower: 6.9 years for all cases and 7.8 years for the relevant cases.

### Discussion

The results suggest that regular screening as in the ERSPC trial for prostate cancer may advance diagnosis by approximately 10 years when assuming 100% attendance. For screening in the US population, the estimated lead times were lower.

### Conclusion

Due to differences in PSA testing between US and the ERSPC trial, there is a small difference in estimated lead time. However, the lead time is long in comparison with other cancers.



Erasmus MC (Prostate)  
Lead Time  
References:

## REFERENCES:

- <sup>1</sup> Draisma, G, Boer, R, Otto, SJ, van der Crujisen, IW, Damhuis, RA, Schröder, FH, de Koning, HJ. "Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer" in J Natl Cancer Inst 2003; 95: : 868-878
- <sup>2</sup> Draisma, G, Etzioni, R, Tsodikov, A, Mariotto, A, Wever, E, Gulati, R, Feuer, E, de Koning, HJ "Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context" in J Natl Cancer Inst 2009; 101: : 374-383

# OVERDIAGNOSIS

## Overdiagnosis

### Summary

This document describes the estimates of overdiagnosis using the model based on the ERSPC Rotterdam trial and the model for the US population.

### Overview

Overdiagnosis, the detection by screening of cancers that would not be detected in the absence of screening, is estimated with the use of the validated models for the ERSPC trial and the US population.

### Methods

For this study overdiagnosis is defined as cancers that would not have been diagnosed within the person's life time in the absence of screening<sup>12</sup>. The model for the ERSPC trial as well as the model for the US population has been used to estimate overdiagnosis. In the ERSPC model 100% attendance to screening is assumed. Overdiagnosis was calculated for various screen programs and expressed as percentage irrelevant cancers of screen detected cancers.

### Results

The amount of overdiagnosis is dependent on the screening program (Table 1).

Table 1. Percentage of overdiagnosis for various screening programs using the ERSPC model.

| screen program       | age               | % overdiagnosis |
|----------------------|-------------------|-----------------|
| single               | 55                | 27              |
| 60                   | 38                |                 |
| 65                   | 47                |                 |
| 70                   | 53                |                 |
| 75                   | 56                |                 |
| interval             | every year, 55-67 | 50              |
| every year, 55-75    | 56                |                 |
| every 4 years, 55-67 | 48                |                 |
| every 4 years, 55-75 | 54                |                 |

For the US situation the estimated overdiagnosis is 44%.

### Discussion

The introduction of regular PSA screening in the Netherlands would lead to a substantial increase in prostate cancer incidence. In the model prediction, approximately half of the screen-detected cancers would not have been diagnosed in the absence of screening.

### Conclusion

Screening is associated with a considerable amount of overdiagnosis.



Erasmus MC (Prostate)  
Overdiagnosis  
References:

## REFERENCES:

- <sup>1</sup> Draisma, G, Boer, R, Otto, SJ, van der Crujisen, IW, Damhuis, RA, Schröder, FH, de Koning, HJ. "Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer" in J Natl Cancer Inst 2003; 95: : 868-878
- <sup>2</sup> Draisma, G, Etzioni, R, Tsodikov, A, Mariotto, A, Wever, E, Gulati, R, Feuer, E, de Koning, HJ "Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context" in J Natl Cancer Inst 2009; 101: : 374-383



FLEXKB DOCUMENT  
Version: HI.001.12212009.86040  
Document generated: 12/21/2009

# FRED HUTCHINSON CANCER RESEARCH CENTER (PSAPC)

[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

**Important note:** This document will be updated periodically. The most current version is available at <http://cisnet.cancer.gov/profiles>. Note that unlike most PDF documents, the CISNET model profiles are not suitable for printing as they are not typically written or read in sequential fashion.

We recommend you let your interests guide you through this document, using the navigation tree as a general guide to the content available.

The intent of this document is to provide the interested reader with insight into ongoing research. Model parameters, structure, and results contained herein should be considered representative but preliminary in nature.

We encourage interested readers to contact the contributors for further information.

Go directly to the: [Reader's Guide](#).



# READERS 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.

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

### 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.

**Validations Overview** A discussion of the major calibration and validation exercises performed throughout model development.

### Key References

A list of references used in the development of the model.



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

# MODEL PURPOSE

## SUMMARY

The PSAPC microsimulation model extends our earlier modeling studies of prostate cancer natural history, prostate-specific antigen (PSA) screening, and disease-specific and other-cause mortality in the US population. The extension involves a new modeling approach and an additional component that models the effects of trends in primary treatment on disease-specific mortality. This document describes the main objective of the PSAPC model.

## PURPOSE

Our primary objective behind modeling prostate cancer trends is to disentangle the roles of PSA screening and changes in primary treatment patterns in US prostate cancer incidence and mortality trends. While both prostate cancer incidence and mortality rates have continued to fall since the early 1990s, the relative contributions of screening and treatment to the observed declines remain intensely debated.

Early results of two randomized clinical trials of PSA screening were recently released, and unfortunately their findings may have only added to the confusion. The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial in the US found no difference in the rates of death from prostate cancer in men who underwent annual PSA screening compared with men who were assigned usual care. The European Randomized Study of Screening for Prostate Cancer (ERSPC) trial involving eight European countries found that PSA screening every 4 years (every 2 years in the Swedish study center) reduced the rate of death from prostate cancer by 20% compared with men randomized to no screening; an even greater benefit is observed among men who actually underwent screening. Reconciling the results of these studies will be an important area of future research.

In contrast, only limited information is available concerning the comparative efficacy of primary treatments--conservative management, radical prostatectomy, and radiation therapy with or without androgen deprivation therapy.

In this context, drawing inference about the value of screening versus treatment from observed trends is very challenging. However, the number of people whose lives are directly or indirectly affected by prostate cancer screening and/or diagnosis every day underscores the potential value to be gained from modeling efforts.



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

# MODEL OVERVIEW

## SUMMARY

This document reviews the motivation for developing a new model of prostate cancer natural history, PSA screening, and treatment practices in the US population. A brief model description is also included.

## BACKGROUND

The original FHCRC CISNET prostate model ([PCSIM](#)) provided a direct link between prostate cancer progression and PSA growth. However, while intuitively reasonable, the link could not be tested empirically. In addition, the cross-model dependence of its components and the large number of parameters (over 30) made systematic estimation intractable. While univariate estimation and informal experimentation provided important information about prostate cancer progression and helped us to understand ways to improve our modeling efforts, we recognized the imperative of a more coherent modeling approach.

The deficiencies of the original FHCRC CISNET motivated an overhaul and the adoption of a new, simpler, unified, statistically coherent model framework. At its core, the new PSAPC model continues to exploit a linkage between prostate cancer progression and PSA growth. In contrast with the original model formulation, this link can now be examined via formal statistical methods since model parameters that determine disease natural history explicitly depend on PSA levels. In other words, the link between progression and PSA growth is now captured through model parameters instead of representing an inflexible assumption buried deep in the internal model structure.

## MODEL DESCRIPTION

Data from the Prostate Cancer Prevention Trial<sup>1</sup> and Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial<sup>2</sup> inform the model about individual PSA growth rates. These rates determine individual PSA trajectories and are linked to hazards of cancer progression events in a simulated population. The hazards of cancer progression represent a natural history model that accounts for clinical diagnosis, while the PSA trajectories together with screening dissemination and biopsy patterns account for screen detections. By comparing the total projected number of new cases to observed incidence, we simultaneously estimate the natural history parameters linking PSA with event hazards and calibrate the model to the US population. Once calibrated, we then systematically remove an intervention (or combination of interventions) and compare projected mortality in its presence and its absence to quantify its impact on mortality.

## CONTRIBUTORS

Ruth Etzioni  
Roman Gulati  
Lurdes Inoue  
Jeffrey Katcher  
Bill Hazelton



## REFERENCES:

- <sup>1</sup> Thompson, IM, Goodman, PJ, Tangen, CM, Lucia, MS, Miller, GJ, Ford, LG, Lieber, MM, Cespedes, RD, Atkins, JN, Lippman, SM, Carlin, SM, Ryan, A, Szczepanek, CM, Crowley, JJ, Coltman, CA Jr. "The influence of finasteride on the development of prostate cancer" in *N Engl J Med* 2003; 349: 3: 214-24
- <sup>2</sup> Andriole, GL, Levin, DL, Crawford, ED, Gelmann, EP, Pinsky, PF, Chia, D, Kramer, BS, Reding, D, Church, TR, Grubb, RL, Grubb, RL, Izmirlian, G, Ragard, LR, Clapp, JD, Prorok, PC, Gohagan, JK, PLCO Project Team "Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: Findings from the initial screening round of a randomized trial" in *J Natl Cancer Inst* 2005; 97: : 433-8

# ASSUMPTION OVERVIEW

## SUMMARY

This document describes the core assumptions of prostate cancer natural history in the PSAPC model.

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

## BACKGROUND

The main idea behind the PSAPC model is to link PSA growth with prostate cancer progression. The model is similar to models linking disease progression with tumor growth, but the PSAPC model replaces tumor volume with an observable biomarker, namely PSA. The model consists of two main components: longitudinal PSA growth and transitions between natural history disease states (i.e., healthy, preclinical, clinical, localized, metastatic). The hazards of transitioning from one state to the next are dependent on age or PSA growth.

## ASSUMPTION LISTING

### PSA GROWTH

We assume:

- PSA growth is log-linear in age
- A changepoint occurs at onset
- PSA growth rates are heterogeneous across individuals

More precisely, we assume PSA grows as follows:

$$\log(y_i(t)) = \beta_{0i} + \beta_{1i}t + \beta_{2i}(t - t_{oi})I(t > t_{oi}) + \varepsilon$$

where

- $i$  indexes subjects
- $y_i(t)$  is PSA at age  $t$
- $I(A)$  is 1 if  $A$  is true and 0 otherwise
- $\beta_{0i} \sim N(\mu_0, \sigma_0^2)$
- $\beta_{ki} \sim N(\mu_k, \sigma_k^2)I(\beta_{ki} > 0), k = 1, 2$
- $\varepsilon \sim N(0, \tau^2)$

Note that  $N(\mu_k, \sigma_k^2)I(\beta_k > 0)$  represents a truncated normal distribution disallowing negative PSA growth. Estimated PSA growth rates together with between-individual truncated normal distributions are illustrated below. These plots are based on parameters estimated from the control group of the Prostate Cancer Prevention Trial<sup>1</sup> and tuned to validate against results of the initial screening round of the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial.<sup>2</sup>

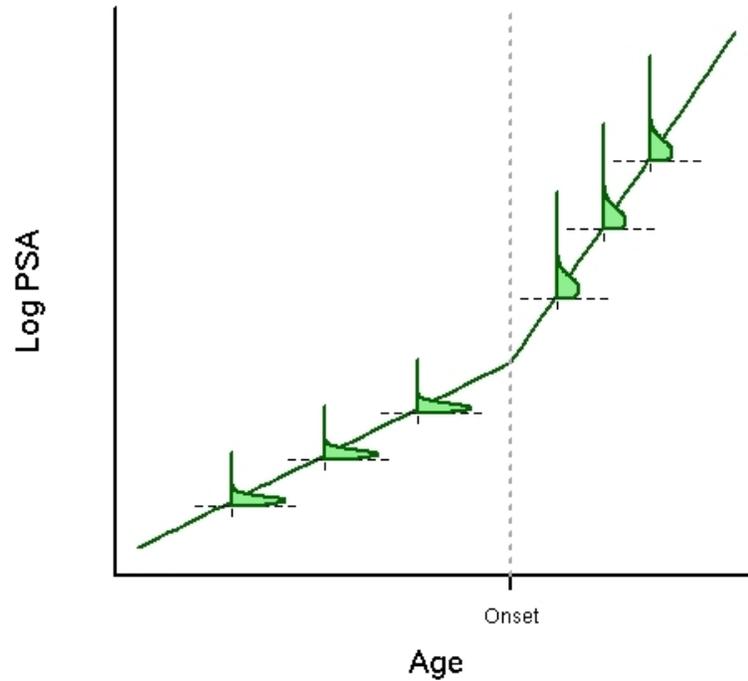


Figure 1. Log-linear PSA growth with truncated normal slopes and changepoint at disease onset.

## NATURAL AND CLINICAL HISTORY EVENT HAZARDS

### DISEASE ONSET

The hazard of prostate cancer onset is proportional to age:

$$\lambda_o(t) = \gamma_o t$$

A variant of the model allows this hazard to increase exponentially with age.

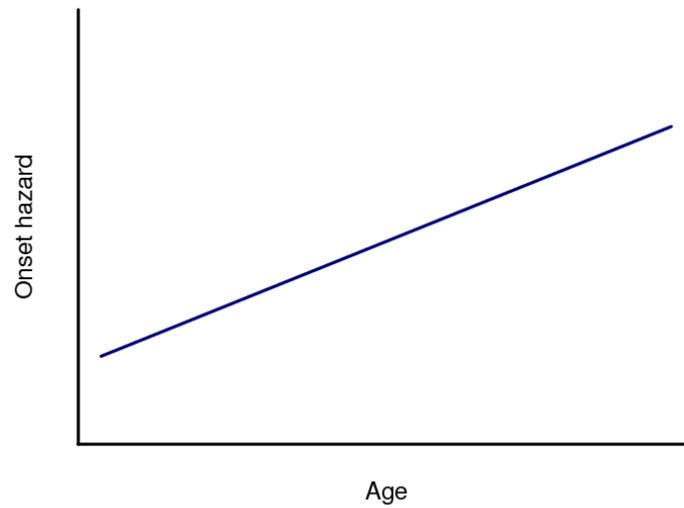


Figure 2. Hazard of disease onset.

### DISEASE METASTASIS

The hazard of transition from localized to metastatic cancer is:

$$\lambda_m(t) = \gamma_m \tilde{y}_i(t)$$

where  $\tilde{y}_i(t) = \exp \{ \beta_{0i} + \beta_{1i}t + \beta_{2i}(t - t_{oi})I(t > t_{oi}) \}$  denotes the individual-specific mean PSA trajectory.

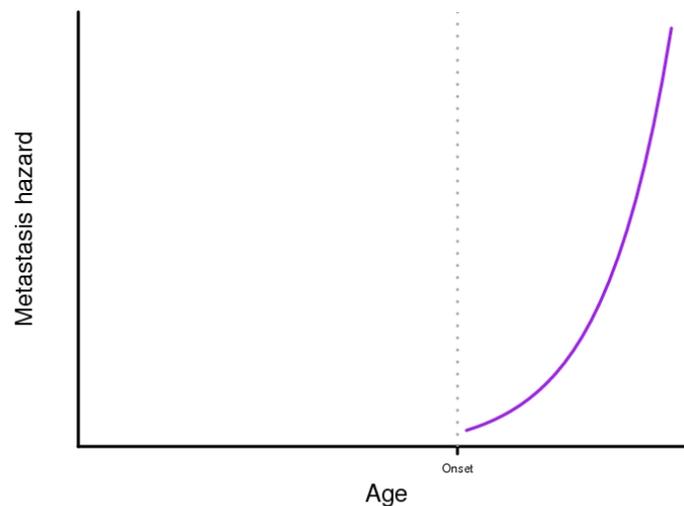


Figure 3. Hazard of progression to advanced stage.

### CLINICAL DIAGNOSIS

The hazard of clinical diagnosis before metastasis is:

$$\lambda_c(t) = \gamma_c \tilde{y}_i(t)$$

and after metastasis is:

$$\lambda_c(t) = \theta_c \gamma_c \tilde{y}_i(t)$$

This specification allows for a greater chance that an individual with metastatic cancer will present symptoms and be diagnosed than one with localized disease.

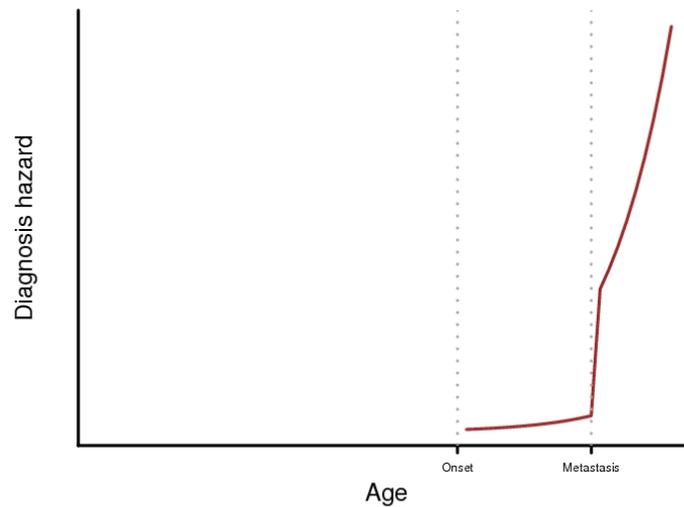


Figure 4. Hazard of clinical diagnosis.

## ADDITIONAL VARIANTS OF NATURAL AND CLINICAL HISTORY EVENT HAZARDS

An extended version of the model incorporates disease grade, categorized as low-moderate (Gleason score 2-7) versus high (Gleason score 8-10). This version has the following additional assumptions:

- PSA growth after disease onset differs for cases with high-grade versus low-grade disease, i.e., the distribution of individual-specific PSA growth rates differs for high- versus low-grade cases
- Disease grade is determined at onset and does not change over time
- The transition rate from localized to metastatic disease, given PSA level, depends on grade category. Thus, the hazard of transition from localized to metastatic cancer for low-grade tumors is:

$$\lambda_m(t) = \gamma_m \tilde{y}_i(t)$$

and for high-grade tumors is:

$$\lambda_m(t) = \theta_m \gamma_m \tilde{y}_i(t)$$



- The transition rate from preclinical disease to clinical diagnosis given PSA depends on both grade and stage. Thus, the hazard of clinical diagnosis for low-grade tumors before metastasis is:

$$\lambda_c(t) = \gamma_{c_1} \tilde{y}_i(t)$$

for high-grade tumors before metastasis is:

$$\lambda_c(t) = \gamma_{c_2} \gamma_{c_1} \tilde{y}_i(t)$$

for low-grade tumors after metastasis is:

$$\lambda_c(t) = \theta_{c_1} \tilde{y}_i(t)$$

and for high-grade tumors after metastasis is:

$$\lambda_c(t) = \theta_{c_2} \theta_{c_1} \tilde{y}_i(t)$$

## REFERENCES:

- <sup>1</sup> Thompson, IM, Goodman, PJ, Tangen, CM, Lucia, MS, Miller, GJ, Ford, LG, Lieber, MM, Cespedes, RD, Atkins, JN, Lippman, SM, Carlin, SM, Ryan, A, Szczepanek, CM, Crowley, JJ, Coltman, CA Jr. "The influence of finasteride on the development of prostate cancer" in N Engl J Med 2003; 349: 3: 214-24
- <sup>2</sup> Andriole, GL, Levin, DL, Crawford, ED, Gelmann, EP, Pinsky, PF, Chia, D, Kramer, BS, Reding, D, Church, TR, Grubb, RL, Grubb, RL, Izmirlan, G, Ragard, LR, Clapp, JD, Prorok, PC, Gohagan, JK, PLCO Project Team "Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: Findings from the initial screening round of a randomized trial" in J Natl Cancer Inst 2005; 97: : 433-8



# PARAMETER OVERVIEW

## SUMMARY

This document describes parameters in the PSAPC model.

## BACKGROUND

In compiling data for estimating model parameters, our main goal was to obtain data that reflects the US population. For this reason, PSA growth rate parameters are input based on data from the PCPT<sup>1</sup> and PLCO<sup>2</sup>, natural and clinical history parameters are estimated via calibration to SEER incidence data, screening dissemination parameters are input based on the NHIS-Medicare PSA data, treatment dissemination data are based on SEER, and biopsy compliance is based on data from the PLCO. All of these data sources reflect either large, population-based surveys or registries or large, population-based trials. Since we do not have large trials in the US comparing initial treatments for prostate cancer, we use data from the Scandinavian trial<sup>3</sup> on radical prostatectomy and selected observational studies to set cause-specific hazard ratios associated with different initial treatment choices. Finally, we base our estimates of biopsy accuracy on a review of relevant literature (see [Biopsy Compliance And Accuracy](#)).

## PARAMETER LISTING OVERVIEW

## PARAMETERS

Parameters in the PSAPC model are listed below. Each set of parameters is identified either as input (i.e., provided to the model based on external sources or model assumptions) or fitted (i.e., estimated via calibration to observed prostate cancer incidence).

- PSA growth parameters (input; based on analysis of longitudinal PSA data from the PCPT and PLCO)
  - PSA growth intercept (value at age 35) mean and variance across individuals ( $\mu_0, \sigma_0^2$ )
  - Pre-onset PSA growth slope mean and variance ( $\mu_1, \sigma_1^2$ )
  - Post-onset PSA growth slope mean and variance ( $\mu_2, \sigma_2^2$ )
  - PSA noise or within-individual error ( $\tau^2$ )
- Natural and clinical history parameters (fitted)
  - Onset hazard ( $\gamma_o$ )
  - Metastasis hazard ( $\gamma_m$ )
  - Pre-metastasis clinical diagnosis hazard ( $\gamma_c$ )
  - Post-metastasis clinical diagnosis hazard ( $\theta_c$ )

- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References



- Grade-based model: Additional PSA growth parameters (input)
  - Post-onset PSA growth slope mean and variance for low-grade cases
  - Post-onset PSA growth slope mean and variance for high-grade cases
  
- Grade-based model: Additional parameters (fitted)
  - Probability a tumor is low grade at onset
  - Metastasis hazard for low-grade cases
  - Metastasis hazard for high-grade cases
  - Pre-metastasis clinical diagnosis hazard for low-grade cases ( $\gamma_{c1}$ )
  - Pre-metastasis clinical diagnosis hazard for high-grade cases ( $\gamma_{c2}$ )
  - Post-metastasis clinical diagnosis hazard for low-grade cases ( $\theta_{c1}$ )
  - Post-metastasis clinical diagnosis hazard for high-grade cases ( $\theta_{c2}$ )
  
- Biopsy parameters (input)
  - Likelihood of referral to biopsy if PSA is below 4.0 ng/ml
  - Biopsy compliance rate, i.e., probability a biopsy is performed if referred; frequencies depend on PSA level and age
  - Biopsy accuracy rate, i.e., probability that a biopsy will detect a tumor if it is present; increases across calendar years
  - Biopsy compliance and accuracy increase to 100% for individuals within  $\delta$  years of transitioning to metastatic disease
  
- Survival parameters (input)
  - Hazard of non-prostate cancer death
  - Baseline prostate cancer survival in the absence of treatment
  - Hazard ratios associated with initial treatments, i.e., radical prostatectomy, radiation therapy, and radiation therapy combined with hormones
  
- Dissemination parameters (input)
  - Screening dissemination: Annual probability of having a PSA test
  - Treatment dissemination: Annual probability of initial treatment choice

## REFERENCES:

- <sup>1</sup> Thompson, IM, Goodman, PJ, Tangen, CM, Lucia, MS, Miller, GJ, Ford, LG, Lieber, MM, Cespedes, RD, Atkins, JN, Lippman, SM, Carlin, SM, Ryan, A, Szczepanek, CM, Crowley, JJ, Coltman, CA Jr. "The influence of finasteride on the development of prostate cancer" in N Engl J Med 2003; 349: 3: 214-24



- <sup>2</sup> Andriole, GL, Levin, DL, Crawford, ED, Gelmann, EP, Pinsky, PF, Chia, D, Kramer, BS, Reding, D, Church, TR, Grubb, RL, Grubb, RL, Izmirlian, G, Ragard, LR, Clapp, JD, Prorok, PC, Gohagan, JK, PLCO Project Team "Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: Findings from the initial screening round of a randomized trial" in *J Natl Cancer Inst* 2005; 97: : 433-8
- <sup>3</sup> Bill-Axelson, A, Holmberg, L, Ruutu, M, Häggman, M, Andersson, SO, Bratell, S, Spångberg, A, Busch, C, Nordling, S, Garmo, H, Palmgren, J, Adami, HO, Norlén, BJ, Johansson, JE, Scandinavian Prostate Cancer Group Study No. 4 "Radical prostatectomy versus watchful waiting in early prostate cancer" in *N Engl J Med* 2005; 352: 19: 1977-84



# COMPONENT OVERVIEW

## SUMMARY

This document describes the main components of the PSAPC simulation model in detail.

## OVERVIEW

The general steps in estimating the natural history parameters and calibrating the model to the US population are as follows.

- A simulated population of individuals is generated to match observed male population counts by age and year. As a consequence of the generation scheme, each simulated individual has a date of birth and a date of all-cause death. Simulated individuals are then randomly assigned PSA growth rates, ages at natural and clinical history events, ages at which PSA screening occurs, and screen-specific biopsy compliance and sensitivity indicators.
- Simulated individual natural and clinical history time courses are followed to determine whether they are screen detected, clinically diagnosed, or neither. In other words, individuals are aged forward and undergo disease progression and screening with each event determining future possible event paths (so that, for example, individuals that are clinically diagnosed do not undergo subsequent screening). Screened individuals are recommended to biopsy if their PSA exceeds 4.0 ng/ml; biopsy occurs based on a biopsy compliance indicator, and the biopsy detects cancer in individuals who have had disease onset based on a biopsy sensitivity indicator. Diagnosed individuals are assigned an initial treatment and, as a consequence of the treatment assignment, a new age at death due to prostate cancer is generated. The earlier of the individual's ages at all-cause and cause-specific death is taken as the true age at death.
- Counts of individuals that are screen detected or clinically diagnosed are then tallied by age, year, and stage at diagnosis. Similarly, counts of prostate cancer death and all-cause death are tallied by age and year at death.
- Projected counts are compared with observed incidence counts by age, year, and stage at diagnosis in a Poisson likelihood. A variant of the Nelder-Mead algorithm for stochastic maximum likelihood<sup>1</sup> is used to estimate model parameters and to calibrate the model to observed incidence data. To account for Monte Carlo error, model parameters are re-estimated for multiple random number seeds.

## COMPONENT LISTING

### POPULATION GENERATION, PSA GROWTH, AND NATURAL/CLINICAL HISTORY

- [Population Generation](#)
- [All Cause Mortality](#)
- [Psa Growth](#)
- [Natural And Clinical History](#)

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)



Fred Hutchinson CRC (PSAPC)  
Component Overview  
References:

## PSA AND DRE SCREENING

- Biopsy Compliance And Accuracy
- Dre Detections

## CLINICAL PRESENTATION AND SCREEN DETECTION

- Treatment Distributions
- Cause Specific Mortality
- Treatment Efficacy

## CLINICAL PRESENTATION AND SCREEN DETECTION

- Model Estimation

## REFERENCES:

- <sup>1</sup> Spall, J "Introduction to stochastic search and optimization: Estimation, simulation, and control" 2003;



# OUTPUT OVERVIEW

## SUMMARY

This document describes the main outputs of the PSAPC microsimulation model.

## OVERVIEW

The main outputs of the PSAPC model are as follows:

- Projected incidence by age, year, stage, grade, and mode of detection.
- Overdiagnosis rates by age and year of diagnosis. An individual is overdiagnosed if he is screen detected but would not have been clinically diagnosed in his lifetime.
- Mean lead time (time from screen detection to clinical diagnosis). We calculate three definitions of lead times:
  - **Relevant** lead times are calculated only for non-overdiagnosed individuals, i.e., individuals for which age at clinical diagnosis precedes age at death.
  - **Censored** lead times are calculated for both non-overdiagnosed individuals and for overdiagnosed individuals, with lead times for overdiagnosed individuals censored at death from other causes.
  - **Uncensored** lead times are calculated for both non-overdiagnosed individuals and for overdiagnosed individuals. The lead times for overdiagnosed individuals are not censored at death from other causes.
- Mean sojourn time (time from disease onset to clinical diagnosis) for the three corresponding definitions.
- Five-, 10-, 15-, and 20-year survival by age and stage at diagnosis for men diagnosed in 2000.
- Mortality by age and year at death and cause of death. Mortality projected under basecase settings compared with that under a given intervention (or combination of interventions) is the main way in which we quantify the intervention's impact.

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)



# RESULTS OVERVIEW

## SUMMARY

This document outlines PSAPC results.

## RESULTS LIST

- [Projected Incidence](#)
- [Lead And Sojourn Times](#)
- [Over Diagnosis](#)

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

# BIOPSY COMPLIANCE AND ACCURACY

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

## BIOPSY COMPLIANCE

Each subject is assigned a profile of discrete uniform random draws that indicate whether he will comply with referral to biopsy and whether a biopsy is sensitive enough to detect existing cancer at each screen.

Biopsy compliance rates vary by age and PSA level based on PLCO trial data illustrated below.<sup>1</sup> Note that to reflect the use of diagnostic PSA testing for metastatic and symptomatic cases, we force biopsy compliance to be 100% when an individual is within  $\delta = 2$  years of transitioning to metastatic disease.

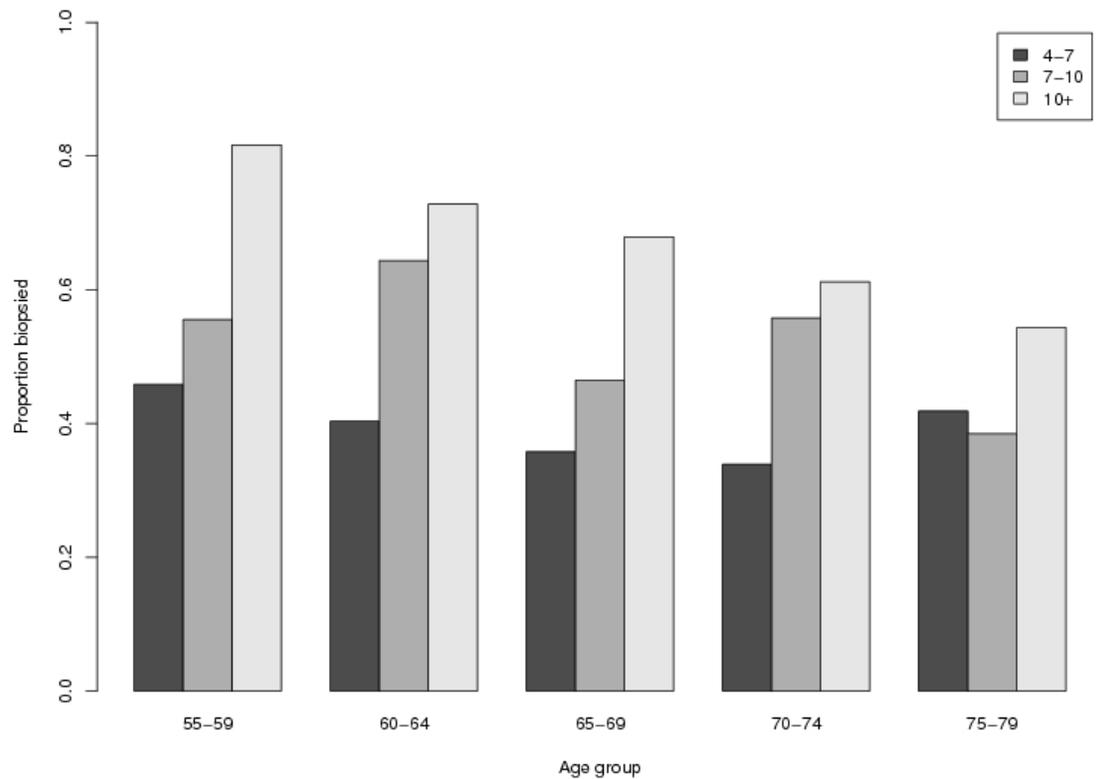


Figure 7. Biopsy compliance rates by age and PSA level.

## BIOPSY SENSITIVITY

Biopsy sensitivity is based on a literature review of how biopsy schemes have changed over the time period considered.<sup>8</sup> Based on these studies we assume:

- Sensitivity increases linearly with the number of cores

- 6-core sensitivity is 80% sensitive
- 8+ cores are 100% sensitive
- The proportion of 6-core scheme decreases linearly after 1995 in favor of 8+ cores

The middle blue line pictured below reflects average biopsy sensitivity rates. (The other lines represent alternative sensitivity patterns to be considered when investigating the robustness of model projections. As for biopsy compliance, to reflect the use of diagnostic PSA testing for metastatic and symptomatic cases, we force biopsy sensitivity to be 100% when an individual is within  $\delta = 2$  years of transitioning to metastatic disease.

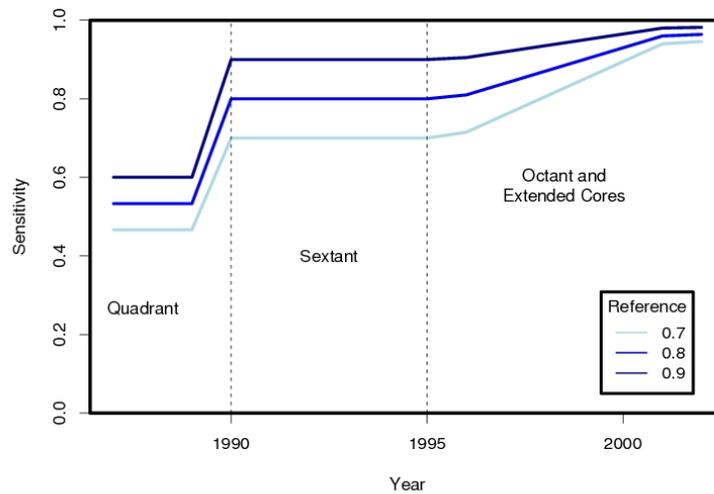


Figure 8. Biopsy sensitivity rates by calendar year.

## REFERENCES:

- 1 Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK "Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial" in J Urol 2000; 173: 3: 746-50
- 2 Hodge KK, McNeal JE, Terris MK, Stamey TA "Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate" in J Urol 1989; 142: : 71-5
- 3 Stamey TA "Making the most out of six systematic biopsies" in Urology 1995; 45: 1: 2-12
- 4 Norberg M, Egevad L, Holmberg L, Sparen P, Norlen BJ, Busch C "The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer" in Urology 1997; 50: : 562-6
- 5 Babaian RJ, Toi A, Kamo K, Troncso P, Sweet J, Evans R, Johnston D, Chen M "A comparative analysis of sextant and an extended 11-core multisite directed biopsy strategy" in J Urol 2000; 163: 152-7
- 6 Klein EA, Zippe CD "Editorial: Transrectal ultrasound guided prostate biopsy – defining a new standard" in J Urol 2000; 163: 179-80
- 7 Presti JC, Chang JJ, Bhargava V, Shinohara K "The optimal systematic prostate biopsy scheme should include 8 rather than 6 biopsies: Results of a prospective clinical trial" in J Urol 2000; 163: 1: 163-6
- 8 Eichler K, Hempel S, Wilby J, Myers L, Bachmann LM, Kleijnen J "Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: A systematic review" in J Urol 2006; 175: 1605-12

# POPULATION GENERATION

## POPULATION GENERATION

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

The target population for drawing inference is SEER 9 men (all races) aged 50-84 in 1975-2000 by single-year age group and calendar year. However, we also model younger ages (i.e., 20-49) and earlier years (i.e., 1950-1974) in order to improve the quality of the model calibration to the target population trends.

The population is generated by creating simulated individuals to populate observed male counts in the observed age-year table one birth year cohort at a time. For each individual in each cohort, we generate a cohort-specific age at all-cause death derived from US life tables. While alive, the individual ages along the cohort-specific diagonal strip of the table contributing to the counts in those cells (birth year 1895 is shown in the figure below). This generation process is repeated until the count in the first calendar year matches the observed total. The process continues along the diagonal with deficits between generated and observed totals filled by new individuals. In practice, we observe only deficiencies and no surpluses, reflecting net immigration into the SEER 9 catchment areas.

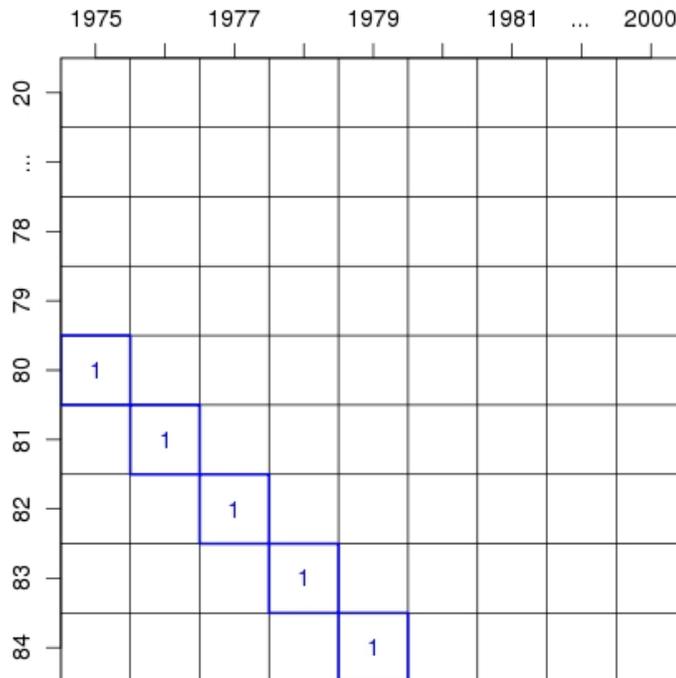


Figure 5. Generating individuals to match observed population counts.

In practice, observed counts are partitioned into many (typically 100) subpopulations that sum to the observed counts. This multi-subpopulation representation allows us to simulate the full SEER 9 population while constraining the number of simulated individuals in memory at any point in time.

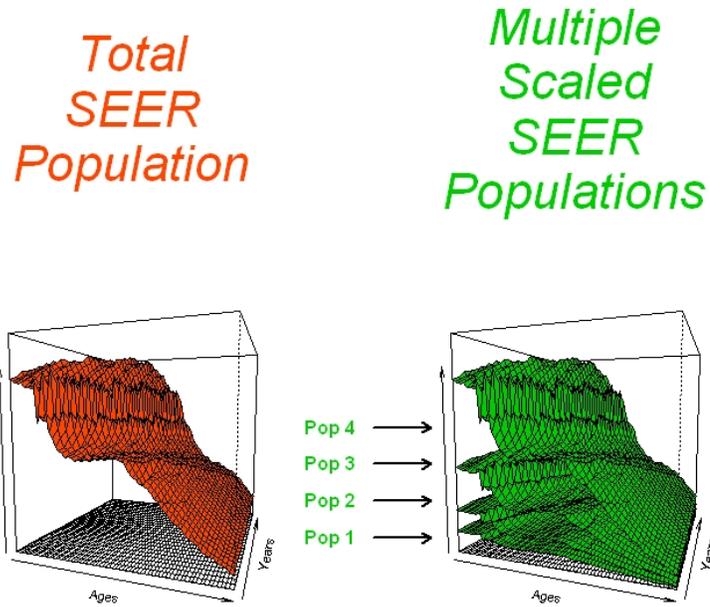


Figure 6. Partitioning full population into sub-populations.



# ALL CAUSE MORTALITY

## ALL-CAUSE MORTALITY

All-cause annual mortality from NCI (based on US life tables from the Berkeley Mortality Database) for ages 0-119 and years 1950-2000 were converted to cohort tables for birth years 1900-2000. Birth-year-specific annual hazards ( $h_y(t)$ ) were then converted to cumulative distribution functions ( $F_y(t)$ ) using the standard relationship:

$$F_y(t) = 1 - \exp \left\{ - \sum_{k=0}^t h_y(k) \right\}$$

where  $y = 1900, \dots, 2000$ . The CDF for year 1900 ( $F_{1900}(t)$ ) was then assumed for years 1865-1899. To ensure death by age 120 we set  $F_y(120) = 1$ .

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)



# PSA GROWTH

## PSA GROWTH

Each subject is assigned a profile of normal and truncated normal random draws that determine his PSA at a reference age, PSA growth rates, and PSA noise at screen and natural and clinical history events.

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

### PSA growth

One normally distributed random draw is used to generate PSA at age 35; this serves as the intercept for log-linear PSA growth over his lifetime; mean PSA at this age is 0.2 ng/ml. A series of normally distributed random draws are used to generate PSA noise at each screen and at each natural and clinical history event.

Long lists ( $n = 50000$ ) of truncated normally distributed draws are randomly sampled and assigned to represent individual-specific PSA growth rates. The means and variances of these random variates are based on a Bayesian mixed model fit to longitudinal PSA growth curves from the Prostate Cancer Prevention Trial (PCPT).<sup>1</sup> We use PCPT interim case data for individuals with at least 3 PSA tests.

### REFERENCES:

- <sup>1</sup> Inoue, L, Etzioni, R, Morrell, C, Muller, P “Modeling disease progression with longitudinal markers” in Journal of the American Statistical Association 2008; 103: : 259-70

# NATURAL AND CLINICAL HISTORY

## NATURAL AND CLINICAL HISTORY

Each subject is assigned a profile of continuous uniform random draws that determine ages at natural and clinical history events.

[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

Ages at onset, at transition to metastasis, and at clinical presentation are generated using random uniform draws evaluated in inverted survivor functions corresponding to each hazard function (a standard analogue of the well-known inverse CDF method). For example, to generate age at onset, the survivor function is:

$$S_o(t) = \exp\left(-\int_0^t \gamma_o s ds\right) = \exp\left(-\frac{\gamma_o}{2} t^2\right)$$

We obtain age at onset for individual  $i$  by evaluating the inverted survivor function at random uniform draw  $u_{oi}$ :

$$t_{oi} = S_o^{-1}(u_{oi}) = \sqrt{-\frac{2}{\gamma_o} \log(u_{oi})}$$

Given his age at onset, we obtain his PSA at onset using his PSA growth rate parameters and random noise:

$$\tilde{y}_i(t_{oi}) = \exp(\beta_{0i} + \beta_{1i} t_{oi} + \varepsilon)$$

where the  $\beta_{ki}$  are the individual-specific PSA growth rates and  $\varepsilon$  is PSA noise at age  $t_{oi}$ . Similarly, to generate his age at metastasis, the survivor function is:

$$S_m(t) = \exp\left(-\int_{t_{oi}}^t \gamma_m \tilde{y}_i(s) ds\right) = \exp\left\{-\frac{\gamma_m}{\beta_{1i} + \beta_{2i}} [\tilde{y}_i(t) - \tilde{y}_i(t_{oi})]\right\}$$

and we obtain an age at metastasis corresponding to random uniform draw  $u_{mi}$  as:

$$t_{mi} = S_m^{-1}(u_{mi}) = \frac{1}{\beta_{1i} + \beta_{2i}} \left\{ \log \left[ \tilde{y}_i(t_{oi}) - \frac{\beta_{1i} + \beta_{2i}}{\gamma_m} \log(u_{mi}) \right] - (\beta_{0i} + \beta_{2i} t_{oi}) \right\}$$

Age and PSA at clinical presentation are generated analogously.



# DRE DETECTIONS

## DRE DETECTIONS

In one variant of the model, we account for DRE detections by randomly assigning individuals with negative PSA test results to biopsy. The frequency of referral to biopsy among men with PSA below 4 is based on a study by Schröder et al (1998) which found that the sensitivity of DRE is approximately 20% for PSA below 3.0 ng/ml and 40% for PSA from 3.0 to 3.9 ng/ml.<sup>1</sup> Men with a negative PSA who are referred to biopsy are assumed to comply with a frequency that is similar to that among men with a moderately elevated PSA (PSA between 4.0 and 7.0 ng/ml).

[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

## REFERENCES:

- <sup>1</sup> Schroder, FH, van der Maas, P, Beemsterboer, P, Kruger, AB, Hoedemaeker, R, Rietbergen, J, Kranse R "Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer" in J Natl Cancer Inst 1998; 90: 23: 1817-23



# TREATMENT DISTRIBUTIONS

## TREATMENT DISTRIBUTIONS

Empirical distributions for treatment choices conservative management (None), radical prostatectomy (RP), and radiation therapy (RT) provide the basis for multinomial random assignment of treatments among individuals diagnosed with local-regional stage disease by grade at diagnosis (Gleason score 2-7 and Gleason 8-10). Similarly, empirical proportions of men receiving androgen deprivation therapy (ADT) form the basis for binomial random assignment by age, year, and grade at diagnosis.<sup>1</sup>

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

## REFERENCES:

- <sup>1</sup> Tsodikov, A, Solomon, C “Generalized self-consistency: Multinomial logit model and Poisson likelihood” in J Stat Plan Inference 2008; 138: 8: 2380-97



# CAUSE SPECIFIC MORTALITY

## CAUSE-SPECIFIC MORTALITY

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

We used Poisson regression models to estimate survival curves for untreated cases. Cause-specific SEER 9 actuarial survival data from SEER\*Stat for men diagnosed at ages 50-84 in 1983-1986 were considered as representative of pre-PSA-era survival. Model covariates included age, treatment decisions (None, RT, or RP), stage (local-regional or distant), grade (SEER categories I-II, III-IV, or unknown) at diagnosis, and selected interactions. The models provide reasonable agreement with observed survival and the projected survival curves for men treated conservatively (i.e., not with RP or RT) agree closely with the curves of Albertsen et al. (2005).<sup>1</sup> These survival curves are used as the baseline cause-specific survival for untreated cases. This baseline survival is adjusted using hazard ratios that reflect treatment-specific efficacy for treated cases.

## REFERENCES:

- <sup>1</sup> Albertsen, PC, Hanley, JA, Fine J "20-year outcomes following conservative management of clinically localized prostate cancer" in JAMA 2005; 293: 17: 2095-101

# TREATMENT EFFICACY

## TREATMENT EFFICACY HAZARD RATIOS

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

To reflect the benefits of treatment, survival for untreated cases is inflated by hazard ratios to obtain survival for treated cases. For RP we assume a hazard ratio of 0.56 both with and without androgen deprivation therapy.<sup>1</sup> For RT, we conducted an informal survey of expert clinicians. We found general agreement that RT+ADT is believed to be similarly efficacious as RP. RT alone, however, remains worse than RP despite improvements in the early 1990s.

Based on these results, we fix treatment basecase RT+ADT efficacy at 0.56 and set RT efficacy at 0.9 before the 1990s and linear decrease to 0.7 by 1995, where it remains to 2000. RT efficacy trends are summarized in the figure below.

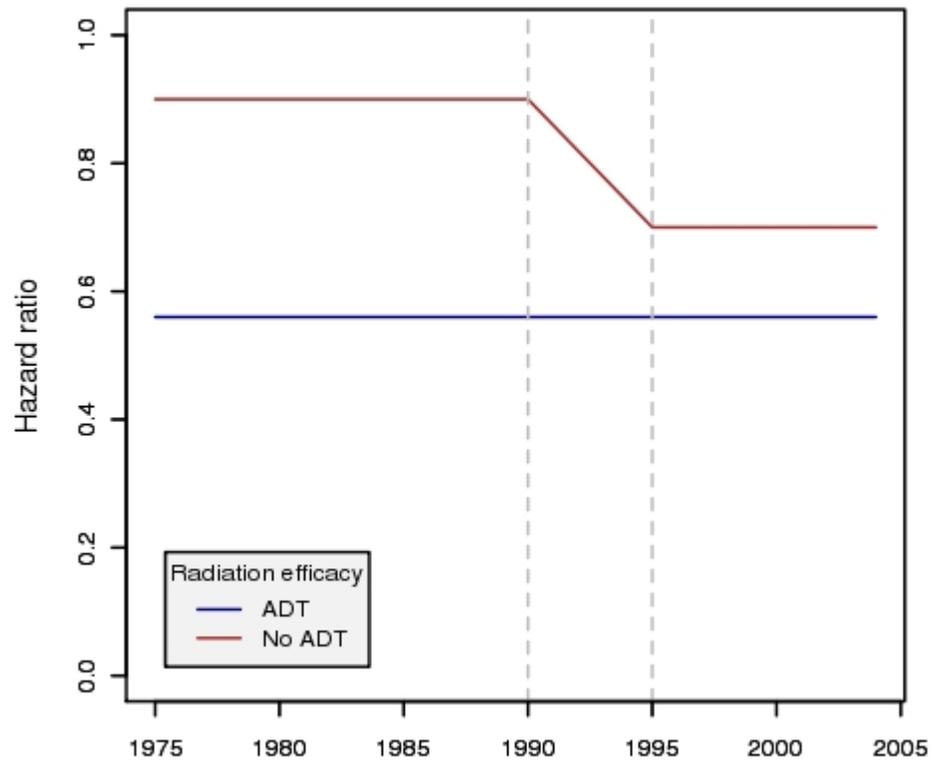


Figure 9. Efficacy of radiation therapy with and without androgen deprivation therapy by calendar year.



Fred Hutchinson CRC (PSAPC)  
Treatment Efficacy  
References:

## REFERENCES:

- <sup>1</sup> Bill-Axelsson, A, Holmberg, L, Ruutu, M, Häggman, M, Andersson, SO, Bratell, S, Spångberg, A, Busch, C, Nordling, S, Garmo, H, Palmgren, J, Adami, HO, Norlén, BJ, Johansson, JE, Scandinavian Prostate Cancer Group Study No. 4 "Radical prostatectomy versus watchful waiting in early prostate cancer" in N Engl J Med 2005; 352: 19: 1977-84

# MODEL ESTIMATION

## MODEL ESTIMATION

Subjects' PSA trajectories, disease natural histories, and screening experience yield projected incidence counts by age, year, and stage. Comparing with corresponding observed counts, we estimate parameters by maximizing the log Poisson likelihood:

$$\log L(\Theta | \mathbf{O}, \mathbf{E}) = \sum_{ays} \left[ O_{ays} \times \left( 1 + \log \frac{E_{ays}}{O_{ays}} \right) - E_{ays} \right]$$

where

- $\Theta = (\mu_0, \mu_1, \mu_2, \sigma_0^2, \sigma_1^2, \sigma_2^2, \tau^2, \gamma_O, \gamma_A, \gamma_C, \theta_C) \cup \Omega$
- $\Omega$  = grade-specific or other model-variant-specific parameters
- $\mathbf{O}$  = SEER 9 observed counts
- $\mathbf{E}$  = model-projected counts
- $a = 20, \dots, 84$  indexes ages
- $y = 1975, \dots, 2000$  indexes calendar years
- $s$  = local-regional or distant stages

Maximization is performed using the Nelder-Mead algorithm adapted for stochastic likelihoods<sup>1</sup> based on Bhat, a suite of optimization routines generously provided by Dr. Georg Luebeck.<sup>2</sup>

Note that observed local-regional and distant stage incidence counts are inflated to account for cases with unknown stage. Such unstaged cases are allocated to local-regional or distant stage according to their relative proportions in each age group and calendar year.

## REFERENCES:

- <sup>1</sup> Spall, J "Introduction to stochastic search and optimization: Estimation, simulation, and control" 2003;
- <sup>2</sup> Luebeck, G "Bhat: General likelihood exploration"

# PROJECTED INCIDENCE

## SUMMARY

This documents summarizes projected incidence from the basecase PSAPC model.

## RESULT TYPE

Target Simulation

## OVERVIEW

Incidence projections reflect the successfulness of calibration of the model to the US population. Our goal is to match as closely as possible observed incidence patterns.

## METHODS

Age-adjusted observed and projected incidence trends are presented by stage. Results are based on averages across 20 random seeds and re-estimated natural and clinical history parameters.

## RESULT

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

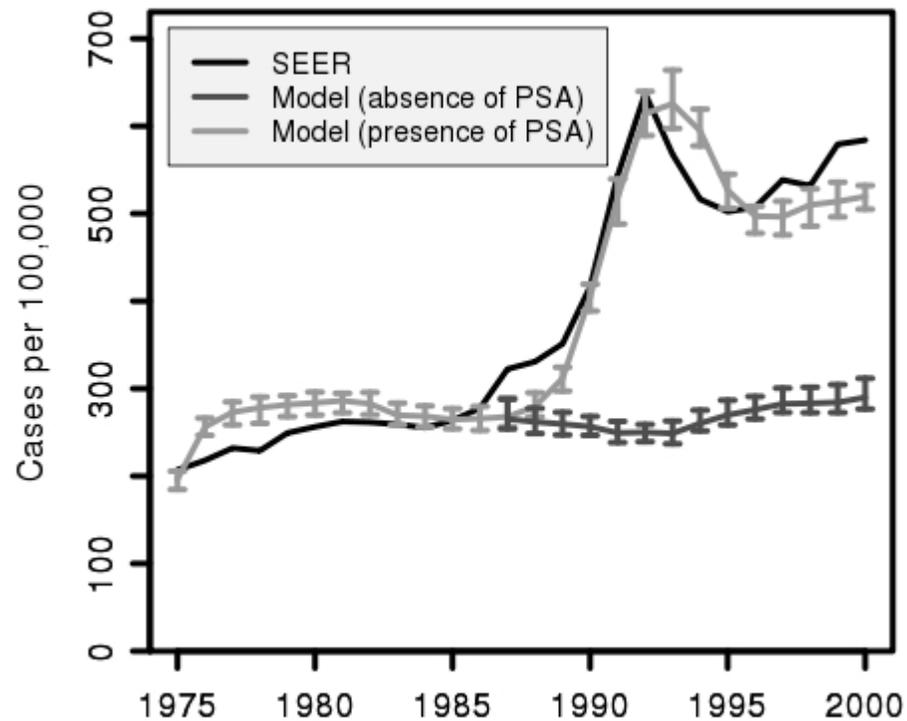


Figure 10. Age-adjusted observed and projected local-regional stage incidence.

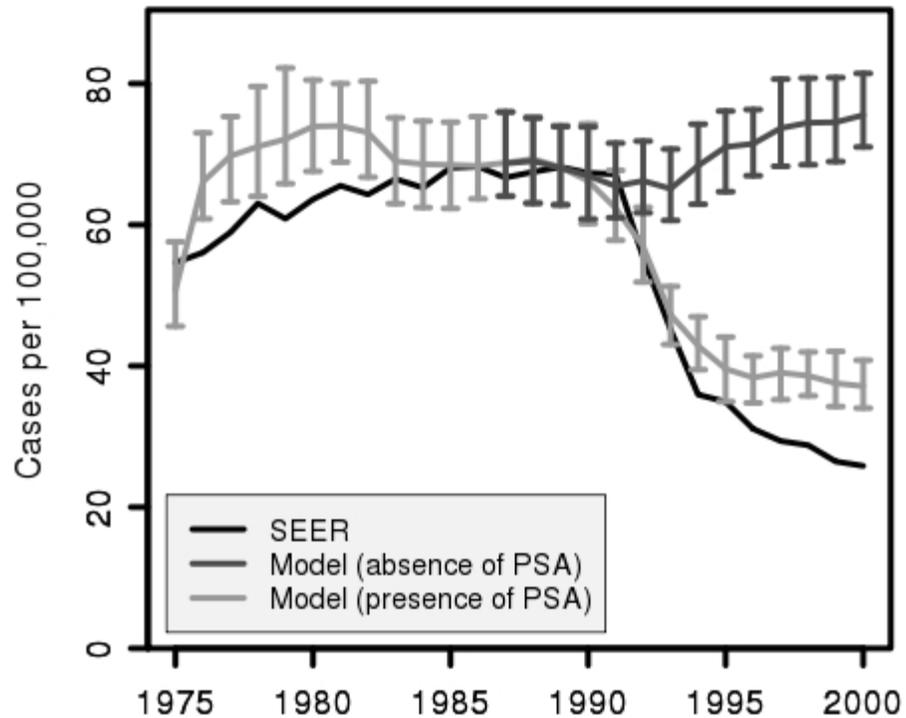


Figure 11. Age-adjusted observed and projected distant stage incidence.

## DISCUSSION

Projected incidence matches the general shape of observed incidence rather well. The model overprojects local-regional stage incidence in the pre-PSA era, and the spike following early PSA dissemination is less peaked than observed. The model underprojects distant stage incidence in the late 1980s, then fails to fall as quickly as observed. Difficulty attaining the observed decline in distant stage incidence has been experienced in other modeling frameworks as well.<sup>1</sup>

## CONCLUSION

Model projections are imperfect but reasonable considering the simplicity of its assumptions.

## REFERENCES:

- <sup>1</sup> Etzioni R, Gulati R, Falcon S, Penson DF "Impact of PSA screening on the incidence of advanced stage prostate cancer in the United States: A surveillance modeling approach" in *Med Decis Making* 2008; 28: 3: 323-31



# LEAD AND SOJOURN TIMES

## SUMMARY

This documents summarizes projected lead and sojourn time results from the basecase PSAPC model.

## RESULT TYPE

Validation

## OVERVIEW

Lead time represents an important measure of the benefit of screening since it represents the time by which diagnosis is advanced by screening. Since detection of cancer at an earlier stage confers a survival benefit (this is the main argument behind early detection programs), lead times quantify the potential benefit in the context of the cancer's natural history. Sojourn time, reflecting duration of pre-diagnosis disease progression in the absence of screening, provides valuable information concerning this natural history.

## METHODS

Note that mean lead and sojourn times are projected based on the basic model variant that excludes DRE screening to avoid confounding PSA screening with DRE screening. *Lead time* is defined as the time interval from screen detection to clinical diagnosis. *Sojourn time* is defined as the time interval from disease onset to clinical diagnosis. Results reported here are averages over 10 runs.

## RESULT

Mean lead times by age at PSA detection

Mean lead times by age at PSA detection

| Age      | Censored | Relevant | Uncensored |
|----------|----------|----------|------------|
| 50-54    | 7.77     | 7.88     | 8.17       |
| 55-59    | 7.90     | 8.00     | 8.67       |
| 60-64    | 7.62     | 7.70     | 8.96       |
| 65-69    | 6.98     | 6.91     | 8.96       |
| 70-74    | 6.42     | 6.20     | 9.31       |
| 75-79    | 6.30     | 5.44     | 12.25      |
| 80-84    | 4.59     | 3.18     | 11.46      |
| Adjusted | 7.13     | 6.97     | 9.27       |

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

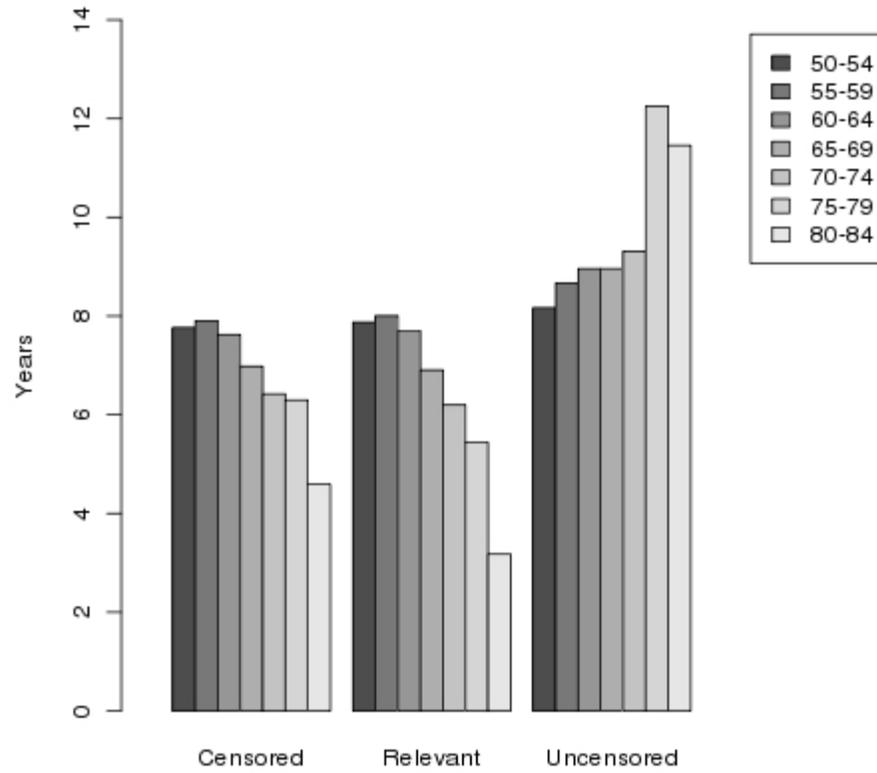


Figure 12. Mean lead times by age group and definition.

Mean sojourn times by age at onset

| Age      | Censored | Relevant | Uncensored |
|----------|----------|----------|------------|
| 50-54    | 19.18    | 18.29    | 16.84      |
| 55-59    | 16.74    | 16.27    | 14.67      |
| 60-64    | 14.30    | 14.12    | 12.18      |
| 65-69    | 11.99    | 11.91    | 9.66       |
| 70-74    | 9.79     | 9.62     | 7.97       |
| 75-79    | 7.69     | 7.33     | 6.15       |
| 80-84    | 5.78     | 5.11     | 4.51       |
| Adjusted | 13.80    | 13.36    | 11.77      |

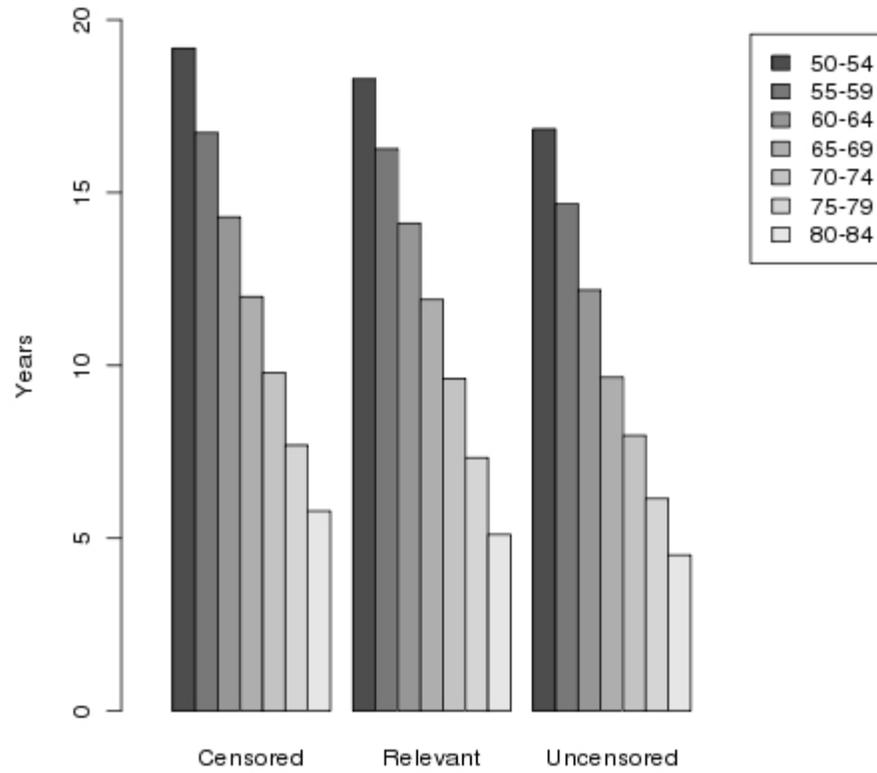


Figure 13. Mean sojourn times by age group and definition.

| Mean local-regional stage durations by age at onset |          |          |            |
|---|----------|----------|------------|
| Age   | Censored | Relevant | Uncensored |
| 50-54   | 12.20    | 21.53    | 22.52      |
| 55-59   | 9.70     | 19.33    | 19.86      |
| 60-64   | 7.23     | 16.90    | 16.70      |
| 65-69   | 5.05     | 14.41    | 13.41      |
| 70-74   | 3.58     | 11.76    | 11.21      |
| 75-79   | 2.28     | 8.97     | 8.77       |
| 80-84   | 1.32     | 6.17     | 6.51       |
| Adjusted  | 7.23     | 15.95    | 16.07      |

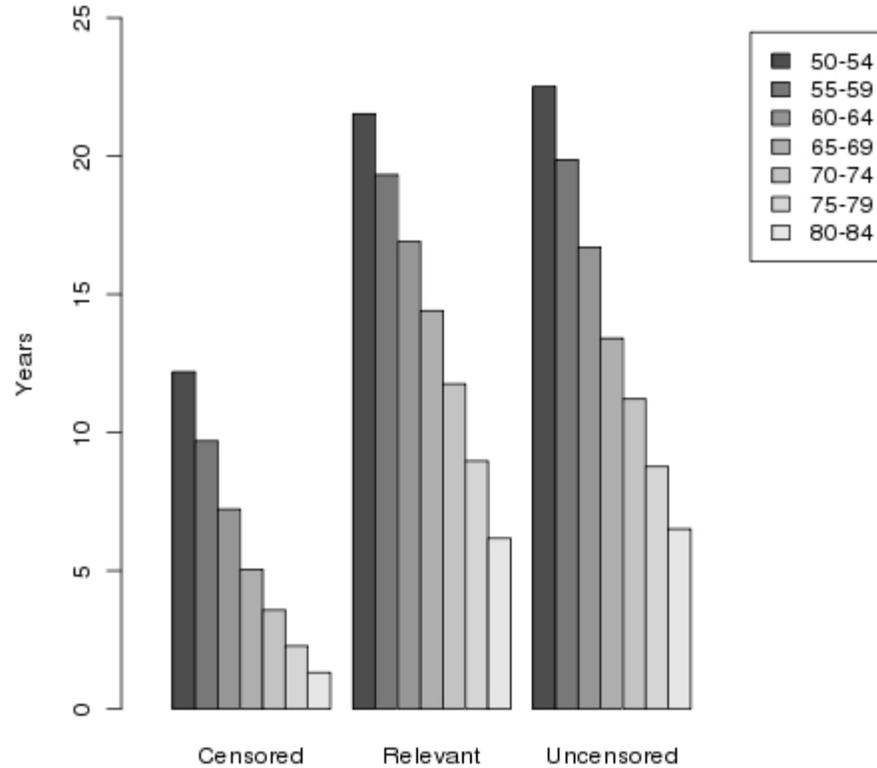


Figure 14. Mean local-regional stage durations by age group and definition.

## DISCUSSION

Relevant mean lead times are longer for younger men than for older men since the possible intervals until diagnosis narrows with age. Our projections are modestly higher than estimates reported by Gann et al.<sup>1</sup> and Telesca et al.<sup>2</sup> but considerably lower than those presented by Draisma et al.<sup>3</sup> However, we note that estimates based on data from a European screening trial differ in important ways from the US population setting.<sup>4</sup>

Relevant mean sojourn times around 12 years are consistent with earlier estimates obtain with the original version of the CISNET FHCRC model.<sup>5</sup>

Mean local-regional stage durations for relevant cases is estimated to be approximately 16 years. This estimate is difficult to validate since published literature tend to use a finer staging system than what is available in SEER. These results are nonetheless reported here for completeness.

## CONCLUSION

Model-projected mean lead and sojourn times are generally consistent with previously published studies.



## RELEVANT PARAMETERS

Validation of lead and sojourn times serves as a check of several model parameters, including the hazard of disease onset and the rate of transitioning to clinical disease.

## REFERENCES:

- <sup>1</sup> Gann PH, Hennekens CH, Stampfer MJ "A prospective evaluation of plasma prostate-specific antigen for detection of prostate cancer" in JAMA 1995; 273: 289-94
- <sup>2</sup> Telesca D, Etzioni R, Gulati R "Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends" in Biometrics 2008; 64: 1: 10-9
- <sup>3</sup> Draisma G, Boer R, Otto SJ, van der Crujisen IW, Damhuis RA, Schroder FH, de Koning HJ, "Lead times and overdiagnosis due to prostate-specific antigen screening: Estimates from the European Randomized Study of Screening for Prostate Cancer" in J Natl Cancer Inst 2003; 95: 12: 868-78
- <sup>4</sup> Draisma, G, Etzioni, R, Tsodikov, A, Mariotto, A, Wever, E, Gulati, R, Feuer, E, de Koning, H "Lead time and overdiagnosis in prostate-specific antigen screening: Importance of methods and context" in J Natl Cancer Inst 2009; 101: 6: 374-83
- <sup>5</sup> Etzioni R, Cha R, Feuer EJ, Davidov O "Asymptomatic incidence and duration in prostate cancer" in Am J Epidemiol 1998; 148: 775-85



# OVER DIAGNOSIS

## SUMMARY

This document summarizes projected overdiagnosis rates from the basecase PSAPC model.

## RESULT TYPE

Validation

## OVERVIEW

An individual is *overdiagnosed* if he is screen detected but would not have been diagnosed in the absence of PSA screening. Overdiagnosis rates represent one of the main drivers of costs associated with PSA screening.

## METHODS

The PSAPC counts simulated individuals who are screen-detected but whose date of clinical diagnosis exceeds his date of other-cause death. These overdiagnosis counts are recorded by age, year, and stage. Overdiagnosis rates are calculated by dividing these counts by all diagnoses or by screen detections in each age, year, and stage, aggregating across stages, then age-adjusted to the 2000 US standard million for ages 50-84. Reported overdiagnosis rates are averages over 20 random seeds with re-estimated natural and clinical history parameters.

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

## RESULT

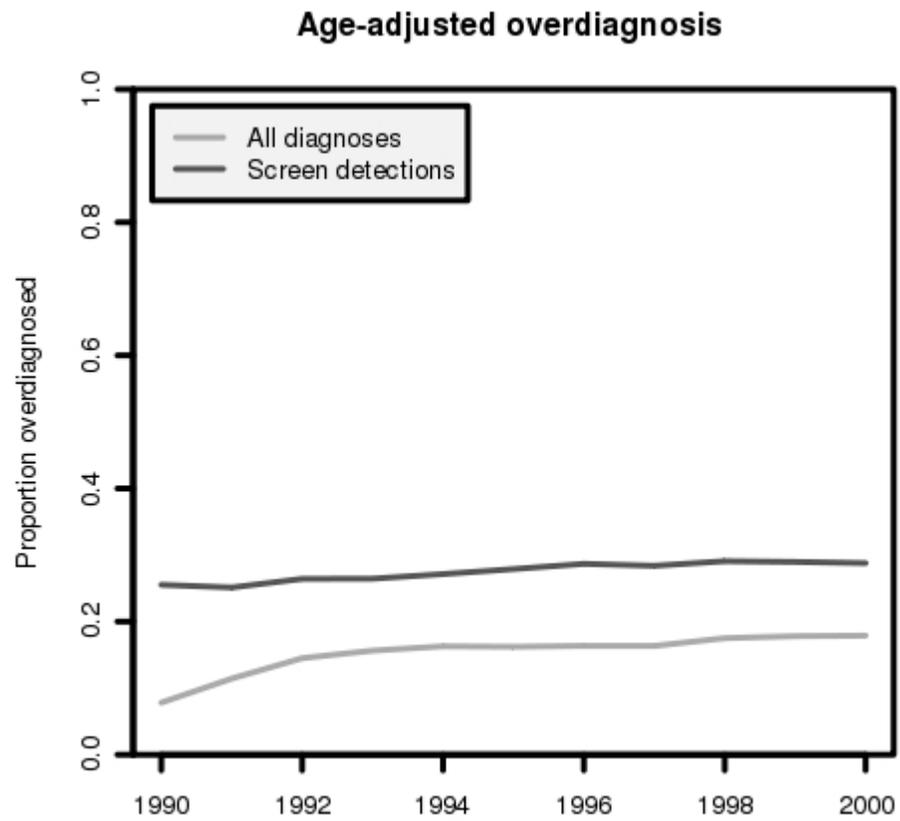


Figure 13. Age-adjusted overdiagnosis rates by calendar year.

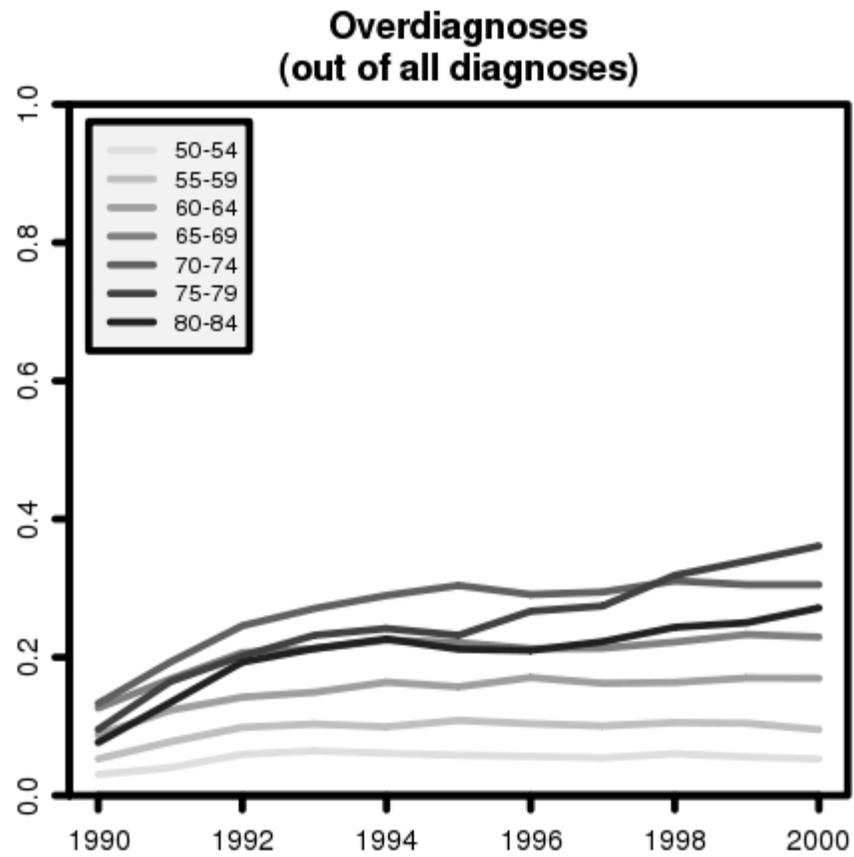


Figure 14. Overdiagnosis as fraction of all detections by age group and calendar year.

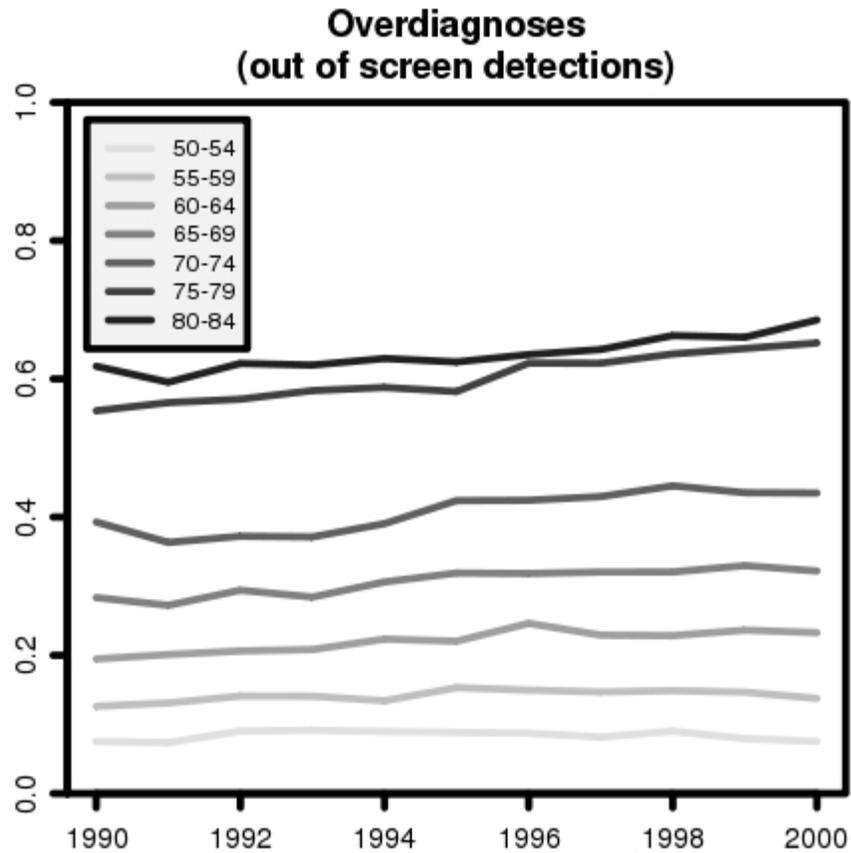


Figure 15. Overdiagnosis as fraction of screen detections by age group and calendar year.

## DISCUSSION

Age-adjusted overdiagnosis rates are relatively flat after 1992, about when PSA screening stabilized in the US population. The model projects that each year about 18% of new cases (29% of new screen-detected cases) are overdiagnosed.

The age-specific projections illustrate two intuitive patterns. First, overdiagnosis rates as fractions of screen detections are constant across years, while overdiagnosis rates as fractions of all diagnoses follow PSA dissemination trends, increasing as screening disseminates into the population in the early years then stabilizing in later years. Second, higher overdiagnosis rates are associated with older age groups; this is expected since older men face higher risk of other-cause death each year, so that when these men are detected by screening, it is more likely that other-cause death occurs before they would have presented clinically.

## CONCLUSION

Overdiagnosis results exhibit intuitive general features and are consistent with values reported in the literature for the US population.



Fred Hutchinson CRC (PSAPC)  
Over Diagnosis  
Conclusion

# KEY REFERENCES

- Albertsen PC** (1998) Editorial: Computer modeling—What should we look for? in *J Urol* 159:3, p 934
- Albertsen, PC, Hanley, JA, Fine J** (2005) 20-year outcomes following conservative management of clinically localized prostate cancer in *JAMA* 293:17, p 2095-101
- Andriole, GL, Levin, DL, Crawford, ED, Gelmann, EP, Pinsky, PF, Chia, D, Kramer, BS, Reding, D, Church, TR, Grubb, RL, Grubb, RL, Izmirlian, G, Ragard, LR, Clapp, JD, Prorok, PC, Gohagan, JK, PLCO Project Team** (2005) Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: Findings from the initial screening round of a randomized trial in *J Natl Cancer Inst* 97, p 433-8
- Babaian RJ,, Toi A,, Kamoi K,, Troncoso P,, Sweet J,, Evans R,, Johnston D,, Chen M** (2000) A comparative analysis of sextant and an extended 11-core multisite directed biopsy strategy in *J Urol* 163, p 152-7
- Babaian RJ, Kojima M, Ramirez EI, Johnston D** (1996) Comparative analysis of prostate specific antigen and its indexes in the detection of prostate cancer in *J Urol* 156:2, p 432-7
- Babaian, RJ, Fritsche, H, Ayala, A, Bhadkamkar, V, Johnston, DA, Naccarato, W, Zhang Z** (2000) Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL in *Urology* 56:6, p 1000-6
- Barry MJ, Fleming C, Coley CM, Wasson JH, Fahs MC, Oesterling JE** (1995) Should Medicare provide reimbursement for prostate-specific antigen testing for early detection of prostate cancer? Part IV: Estimating the risks and benefits of an early detection program in *Urology* 4:46, p 445-61
- Bill-Axelson, A, Holmberg, L, Ruutu, M, Häggman, M, Andersson, SO, Bratell, S, Spångberg, A, Busch, C, Nordling, S, Garmo, H, Palmgren, J, Adami, HO, Norlén, BJ, Johansson, JE, Scandinavian Prostate Cancer Group Study No. 4** (2005) Radical prostatectomy versus watchful waiting in early prostate cancer in *N Engl J Med* 352:19, p 1977-84
- Boring CC, Squires TS, Tong T, Montgomery S** (1994) Cancer Statistics, 1994 in *CA: A Cancer Journal for Clinicians*:44, p 7-26
- Brown ML, Riley GF, Potosky AL, Etzioni RD** (1999) Obtaining long-term disease-specific costs of care: Application to Medicare enrollees diagnosed with colorectal cancer in *Medical Care* 37:12, p 1249-59
- Carter HB, Morrell CH, Pearson JD, Brant LJ, Plato CC, Metter EJ, Chan DW, Fozard JL, Walsh PC** (1992) Estimation of prostatic growth using serial prostate-specific antigen measurements in men with and without prostate disease in *Cancer Res*:52, p 3323-8
- Carter HB, Piantadosi S, Isaacs JT** (1990) Clinical evidence for and implications of the multistep development of prostate cancer in *J Urol*:143, p 742-6
- Catalona WJ, Smith DS, Ratliff TL, Basler JW** (1993) Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening in *JAMA*, p 948-54
- Clark LC, Combs GF, Turnbull BW, Slate EH, Chalker DK, Chow J, Davis LS, Glover RA, Graham GF, Gross EG, Krongrad A, Leshner JL, Park HK, Sanders BB, Smith, CL, Taylor JR, and The Nutritional Prevention of Cancer Study Group** (1996) Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: A randomized clinical trial in *JAMA*, p 1957-63

- Coley CM, Barry MJ, Fleming C, Fahs MC, Mulley AG (1997)** Early detection of prostate cancer Part II: Estimating the risks, benefits, and costs in *Ann Intern Med* 126:6, p 468-79
- Crawford ED, DeAntoni EP, Etzioni R, Schaefer VC, Olson RM, Ross CA, and the Prostate Cancer Education Council (1996)** Serum prostate-specific antigen and digital rectal examination for the early detection of prostate cancer in a national community-based program in *Urology*:47, p 863-9
- Cronin KA, Slate EH, Turnbull BW, Wells MT (1994)** Using the Gibbs sampler to detect changepoints: Application to PSA as a longitudinal marker for prostate cancer in *Computing Science and Statistics*:26, p 314-8
- Draisma G, Boer R, Otto SJ, van der Crujisen IW, Damhuis RA, Schroder FH, de Koning HJ (2003)** Lead times and overdiagnosis due to prostate-specific antigen screening: Estimates from the European Randomized Study of Screening for Prostate Cancer in *J Natl Cancer Inst* 95:12, p 868-78
- Draisma, G, Etzioni, R, Tsodikov, A, Mariotto, A, Wever, E, Gulati, R, Feuer, E, de Koning, H (2009)** Lead time and overdiagnosis in prostate-specific antigen screening: Importance of methods and context in *J Natl Cancer Inst* 101:6, p 374-83
- Dugan JA, Bostwick DG, Myers RP, Qian J, Bergstrahl EJ, Oesterling JE (1996)** The definition and preoperative prediction of clinically insignificant prostate cancer in *JAMA* 275:4, p 288-94
- Eichler K,, Hempel S,, Wilby J,, Myers L,, Bachmann LM,, Kleijnen J (2006)** Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: A systematic review in *J Urol* 175, p 1605-12
- Etzioni R, Cha R, Cowen ME (1999)** Serial prostate specific antigen screening for prostate cancer: A computer model evaluates competing strategies in *J Urol* 162, p 741-8
- Etzioni R, Cha R, Feuer EJ, Davidov O (1998)** Asymptomatic incidence and duration in prostate cancer in *Am J Epidemiol*:148, p 775-85
- Etzioni R, Gulati R, Falcon S, Penson DF (2008)** Impact of PSA screening on the incidence of advanced stage prostate cancer in the United States: A surveillance modeling approach in *Med Decis Making* 28:3, p 323-31
- Etzioni R, Legler JM, Feuer EJ, Merrill RM, Cronin KA, Hankey BF (1999)** Cancer surveillance series: Interpreting trends in prostate cancer—Part III: Quantifying the link between population prostate-specific antigen testing and recent declines in prostate cancer mortality in *J Natl Cancer Inst*:91, p 1033-9
- Etzioni R, Pepe M, Longton G, Hu C, Goodman G (1999)** Incorporating the time dimension in receiver operating characteristic curves: A prostate cancer case study in *Med Decis Making*:19, p 242-51
- Etzioni R, Ramsey SD, Berry K, Brown M (2001)** The impact of including future medical care costs when estimating the costs attributable to a disease: A colorectal cancer case study in *Health Econ* 10:3, p 245-56
- Etzioni R, Urban N, Baker M (1996)** Estimating the costs attributable to a disease with application to ovarian cancer in *J Clin Epidemiol*:49, p 95-103
- Etzioni RD, Kadane JB (1995)** Bayesian statistical methods in public health and medicine in *Annu Rev Public Health*:16, p 23-41
- Fleming Craig, Wasson JH, Albertsen PC, Barry MJ, Wennberg JE (1993)** A decision analysis of alternative treatment strategies for clinically localized prostate cancer in *JAMA* 269:20, p 2650-8

- Gann PH, Hennekens CH, Stampfer MJ** (1995) A prospective evaluation of plasma prostate-specific antigen for detection of prostate cancer in *JAMA*:273, p 289-94
- Hodge KK,, McNeal JE,, Terris MK,, Stamey TA** (1989) Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate in *J Urol* 142, p 71-5
- Inoue LY, Etzioni R, Slate EH, Morrell C, Penson DF** (2004) Combining longitudinal studies of PSA in *Biostatistics* 5:3, p 483-500
- Inoue, L, Etzioni, R, Morrell, C, Muller, P** (2008) Modeling disease progression with longitudinal markers in *Journal of the American Statistical Association* 103, p 259-70
- Klein EA,, Zippe CD** (2000) Editorial: Transrectal ultrasound guided prostate biopsy—defining a new standard in *J Urol* 163:179-80,
- Luebeck, GBhat**: General likelihood exploration,
- Mariotto A, Etzioni R, Krapcho M, Feuer EJ** (2007) Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey in *Cancer* 109:9, p 1877-86
- Mettlin C, Murphy G, Babaian R, Chesley A, Kane RA, Littrup PJ, Mostofi FK, Ray PS, Shanberg AM, Toi A** (1996) The results of a five-year early prostate cancer detection intervention in *Cancer* 1:77, p 150-59
- National Hospital Discharge Survey Web Page** Hospital Discharge and Ambulatory Surgery Data (Latest update - July 21, 1999). <http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm>,
- Norberg M,, Egevad L,, Holmberg L,, Sparen P,, Norlen BJ,, Busch C** (1997) The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer in *Urology* 50, p 562-6
- Oesterling JE, Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM** (1993) Serum prostate-specific antigen in a community-based population of healthy men in *JAMA*:270, p 860-4
- Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH, Barnhart S, Hammar S** (1996) Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease in *N Engl J Med* 18:334, p 1150-5
- Pearson JD, Morrell CH, Landis PK, Carter HB, Brant LJ** (1994) Mixed-effects regression models for studying the natural history of prostate disease in *Stat Med*:13, p 587-601
- Pienta K, Goodson JA, Esper S** (1996) Epidemiology of prostate cancer: Molecular and environmental clues in *Urology* 5:48, p 676-83
- Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK** (2005) Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial in *J Urol* 173:3, p 746-50
- Potosky AL, Miller BS, Kramer BS, Albertsen PC** (1995) The role of increasing detection in the rising incidence of prostate cancer. in *JAMA*:273, p 548-552
- Potosky AL, Kessler L, Gridley G, Brown CC, Horm JW** (1990) Rise in prostatic cancer incidence associated with increased use of trans-urethral resection in *J Natl Cancer Inst*:82, p 1624-8
- Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG** (1993) Potential for cancer related health services research using a linked Medicare-tumor registry database in *Med Care* 8:31, p 732-48

- Presti JCJ, Chang JJ, Bhargava V, Shinohara K** (2000) The optimal systematic prostate biopsy scheme should include 8 rather than 6 biopsies: Results of a prospective clinical trial in *J Urol* 163:1, p 163-6
- Richie JP, Catalona WJ, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, deKernion JB, Ratliff TL, Kavoussi LR, Dalkin BL, et al** (1993) Effect of patient age on early detection of prostate cancer with serum Prostate-Specific Antigen and Digital Rectal Examination. in *Urology*:42, p 365-74
- Roberts RO, Bergstralh EJ, Katusic SK, Lieber MM, Jacobsen SJ** (1999) Decline in prostate cancer mortality from 1980 to 1997, and an update on incidence trends in Olmsted County, Minnesota in *J Urol* 161:2, p 529-33
- Schroder, FH, van der Maas, P, Beemsterboer, P, Kruger, AB, Hoedemaeker, R, Rietbergen, J, Kranse R** (1998) Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer in *J Natl Cancer Inst* 90:23, p 1817-23
- Slate EH, Clark LC** (1999) Using PSA to detect prostate cancer onset: An application of Bayesian retrospective and prospective changepoint identification in *Case Studies in Bayesian Statistics IV*, p 511-34
- Smith DS, Catalona WJ** (1994) The nature of prostate cancer detected through prostate-specific antigen based screening in *J Urol*:152, p 1732-6
- Spall, J** (2003) Introduction to stochastic search and optimization: Estimation, simulation, and control,
- Stamey TA** (1995) Making the most out of six systematic biopsies in *Urology* 45:1, p 2-12
- Stanford JL, Stephenson RA, Coyle LM, Cerhan J, Correa R, Eley JW, Gilliland F, Hankey B, Kolonel LN, Kosary C, Ross R, Severson R, West D** (1999) Prostate Cancer Trends 1973-1995, SEER Program in *National Cancer Institute*,
- Stanford JL, Wicklund KG, McKnight B, , Daling JR, Brawer MK** (1999) Vasectomy and risk of prostate cancer in *Cancer Epidemiol Biomarkers Prev*:8, p 881-6
- Surveillance Research Program, National Cancer Institute** (1997) SEER\*Stat Databases: Incidence SEER 9 Regs Public Use, November 2005 (1973-2003),
- Telesca D, Etzioni R, Gulati R** (2008) Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends in *Biometrics* 64:1, p 10-9
- Thompson, IM, Goodman, PJ, Tangen, CM, Lucia, MS, Miller, GJ, Ford, LG, Lieber, MM, Cespedes, RD, Atkins, JN, Lippman, SM, Carlin, SM, Ryan, A, Szczepanek, CM, Crowley, JJ, Coltman, CA Jr.** (2003) The influence of finasteride on the development of prostate cancer in *N Engl J Med* 349:3, p 214-24
- Tsodikov, A, Solomon, C** (2008) Generalized self-consistency: Multinomial logit model and Poisson likelihood in *J Stat Plan Inference* 138:8, p 2380-97
- Urban N, Drescher C, Etzioni R, Colby C** (1997) Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening in *Control Clin Trials*:18, p 251-70
- Weiss NS, Rossing MA** (1996) Healthy screenee bias in epidemiologic studies of cancer incidence in *Epidemiology* 3:7, p 319-22
- Whitmore WF** (1988) Background for screening: Natural history and treatment in *EORTC Genitourinary Group Monograph 5: Progress and Controversies in Oncological Urology II*, p 123-30
- Whitmore WF** (1990) Natural history of low-stage prostatic cancer and the impact of early detection in *Urol Clin N Am*:17, p 689-97



- Whittemore AS, Keller JB, Betensky R** (1991) Low grade latent prostate cancer volume: Predictor of clinical cancer incidence? in *J Natl Cancer Inst*:83, p 1231-5
- Whittemore AS, Lele C, Friedman GD, Stamey T, Vogelman JH, Orentreich N** (1995) Prostate-specific antigen as predictor of prostate cancer in black men and white men in *J Natl Cancer Inst*:87, p 354-60
- Wun LM, Merrill RM, Feuer EJ** (1998) Estimating lifetime and age-conditional probabilities of developing cancer in *Lifetime Data Anal*:4, p 169-86

---

CategoryCoreDocs



FLEXKB DOCUMENT  
Version: HI.001.12212009.86041  
Document generated: 12/21/2009

# FRED HUTCHINSON CANCER RESEARCH CENTER (PCSIM)

[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Validations Overview](#)  
[Key References](#)

**Important note:** This document will be updated periodically. The most current version is available at <http://cisnet.cancer.gov/profiles>. Note that unlike most PDF documents, the CISNET model profiles are not suitable for printing as they are not typically written or read in sequential fashion.

We recommend you let your interests guide you through this document, using the navigation tree as a general guide to the content available.

The intent of this document is to provide the interested reader with insight into ongoing research. Model parameters, structure, and results contained herein should be considered representative but preliminary in nature.

We encourage interested readers to contact the contributors for further information.

Go directly to the: [Reader's Guide](#).



# READERS 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.

- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Validations Overview
- Key References

### 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.

### Validations Overview

A discussion of the major calibration and validation exercises performed throughout model development to ensure (improve?) model correctness.

### Key References

A list of references used in the development of the model.

## Further Reading

---

These topics will provide an intermediate level view of the model. Consider these documents if you are interested gaining a working knowledge of the model, its inputs and outputs.

### Advanced Reading

---

These topics denote more detailed documentation about specific and important aspects of the model structure



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)

# MODEL PURPOSE

## SUMMARY

The FHCRC prostate cancer microsimulation is the product of extensive quantitative investigation into prostate cancer natural history, prostate-specific antigen (PSA) production, PSA testing, and disease-specific and other-cause mortality in the US population. This document summarizes FHCRC objectives in developing a prostate cancer microsimulation.

## PURPOSE

The objective of the FHCRC prostate cancer model is to quantify the role of PSA screening in US prostate cancer incidence and mortality trends. Prostate cancer incidence and mortality in the US have been declining since the early 1990s. The role of PSA screening in these trends is a subject of intense debate. Information on the efficacy of PSA testing from controlled clinical trials is lacking, and researchers and the public are divided about how much information about the test can be gleaned from the observed trends.

To address the need for a quantitative approach to linking population PSA testing and prostate cancer trends, our primary specific aim is to develop a computer microsimulation model to project the impact of PSA screening on US prostate cancer incidence and mortality. The model will first project population prostate cancer incidence and mortality in the absence of PSA screening. The model will then superimpose dissemination of PSA screening and the modeled population trends will be compared with those observed.

Early detection of prostate cancer is affected not only by the extent of screening but also by the ability of the test to identify latent cancers. This depends on the growth of PSA in prostate cancer cases which has been estimated in several studies. Since these studies yield somewhat inconsistent results, part of our modeling work will be to estimate PSA growth trajectories based on data from retrospective stored-serum studies. The results of this analysis will be used to inform the microsimulation model about PSA growth in men with prostate cancer.



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Validations Overview
- Key References

# MODEL OVERVIEW

## SUMMARY

This document describes the individual components of the microsimulation and discusses the insights to be gained by developing and using the FHCRC prostate cancer screening microsimulation.

## PURPOSE

Our primary aim is to estimate the impact of PSA screening on US prostate cancer incidence and mortality. Our approach is to generate disease and clinical histories for individual subjects in both the absence and presence of PSA screening. Comparison of these histories quantifies the impact of PSA screening on prostate cancer overdiagnosis and mortality.

## BACKGROUND

Prostate cancer is the most common non-dermatologic male malignancy in the US and the second leading cause of cancer-related mortality in men. Despite the uncertain efficacy of PSA measurement as a tool for early detection of prostate cancer, its use as such has increased dramatically since 1988. By 1994 approximately half of men aged 65 or older in 1987 had had a PSA test<sup>1</sup>.

From 1992 to 2004, prostate cancer mortality in the US declined by 35% and the incidence of late-stage disease by 75%. However, while there is a general consensus that PSA screening explains much of the distant-stage decline, there is still considerable debate about its role in the observed mortality trends.

Many studies have explored the connection between PSA screening and prostate cancer mortality declines. Ecologic analyses have been widely used to compare prostate cancer mortality rates across geographic areas with different PSA utilization patterns. However, nearly all these efforts have yielded negative results. For example, prostate cancer mortality rates declined in both England and Wales, but PSA screening use is considerably lower in these countries than in the US. Another study found that prostate cancer death rates were virtually the same in Seattle and Connecticut even though PSA testing, biopsy, and treatment were much more common in Seattle. While concerns have been raised about the validity and interpretation of negative ecologic studies of PSA screening, there is no question that their persistently negative results have influenced both professional and public opinion about the value of the test.

Several investigators have suggested alternative explanations for declining rates of prostate cancer mortality. These include changes in treatment practices such as increases in curative therapy—surgery and radiation—for localized disease and hormone ablation therapy for localized disease or for early recurrence. In the US, the frequency of curative therapy has almost doubled since 1983, and studies have shown that the use of hormone therapy in conjunction with primary radiation therapy in the US increased substantially during the 1990s. Both of these treatment approaches have shown benefit in randomized studies. However, the role of treatment advances in explaining mortality declines also remains unclear.

The value of PSA screening is a pressing question because it carries high costs in terms of overdiagnosis and overtreatment. As results from two screening trials in the US and Europe are not expected for several years, important insights at present must rely on careful examination of the growing knowledge base concerning disease natural history, progression, and mortality. We use mathematical modeling to connect this information and quantify how much of the US prostate cancer mortality decline may plausibly be attributed to PSA screening.

## MODEL DESCRIPTION

Dr. Etzioni and colleagues previously developed a model of serial PSA screening<sup>2</sup>. The FHCRC prostate cancer model is an extension of this earlier work. The basic premise of the model is to distinguish cases from the total population simulated and to measure the benefit of stage shifting for the cases that are screen detected. A life history of a hypothetical case is presented in Figure 1.

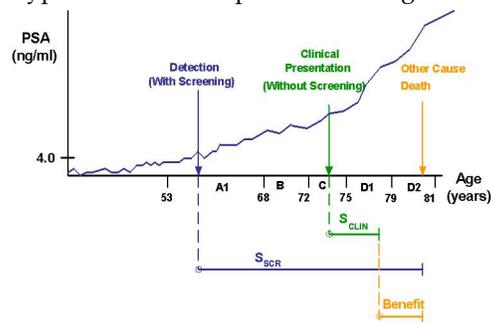


Figure 1. The life history of a hypothetical case, with disease transitions and major events in the absence and presence of screening marked. The difference in endpoints between survival from screen detection ( $S_{SCR}$ ) and survival from clinical diagnosis ( $S_{CLIN}$ ) produces the individual benefit due to screening.

The microsimulation generates clinical and disease histories for a hypothetical cohort of men beginning at age 30. The model comprises five basic modules.

### Natural history

The *natural history module* generates independent:

1. clinical histories (year of birth, age/stage at diagnosis, age of other cause death), and
2. disease histories (age of asymptomatic onset, stage lengths for disease progression as described by Cowen<sup>3</sup> and Whitmore<sup>5</sup>).

We combine data from the Surveillance, Epidemiology, and End Results (SEER) program, the US Census Bureau (USCB), and the National Center for Health Statistics (NCHS) to generate clinical histories. Disease histories are generated by combining data from Etzioni's asymptomatic onset study<sup>6</sup> with Cowen's disease progression rates<sup>3</sup>.

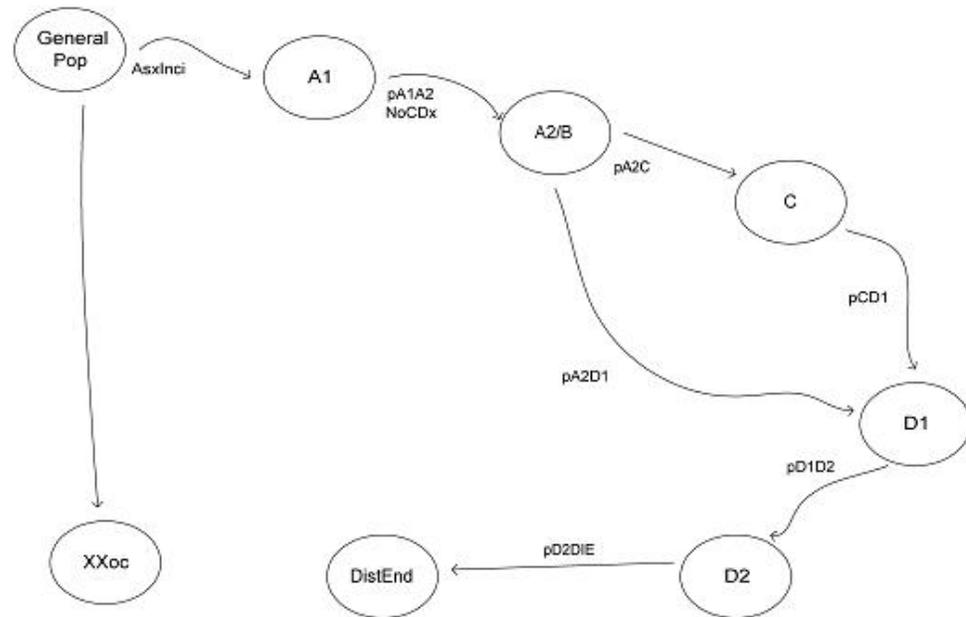


Figure 1. Markov Model diagram for Natural History of Prostate Cancer Onset and Progression. XXoc denotes death due to causes other than prostate cancer. AsxInci denotes asymptomatic incidence which occurs at the transition to stage A1. Transition probabilities (rate parameters for exponential dwelling-time distributions) between the American Urological Association (AUA) pathologic stages as defined by Cowen et al are prefixed with the letter p. NoCDx indicates that clinical diagnosis may be disallowed during the earliest part of stage A1.

**Clinical diagnosis** The *clinical diagnosis module* matches one disease history with each clinical history, thereby producing a complete disease profile for each hypothetical subject. We have explored several methods for matching disease and clinical histories and determined that uniform random matching, while slower, sidesteps artificial anomalies. The model projections of disease incidence prior to the PSA era (i.e., before 1988) are calibrated to match clinical incidence rates observed in the population.



**Serial PSA screening** The *screening module* assigns screening events to subjects. Subjects are eligible for a screen if they are alive and have not been previously diagnosed with prostate cancer. Screen dates are assigned based on Mariotto et al<sup>7</sup>. A positive test is defined as PSA > 4.0 ng/ml. We do not model digital rectal exam (DRE) testing.

**PSA growth** PSA trajectories have different growth rates, dependent on whether the subject is in a cancerous or non-cancerous state. Not all subjects experience disease onset in their lifetime. The PSA growth model is based on work by Inoue et al.<sup>8</sup>. Prior modeling work used the studies of Oesterling<sup>9</sup> and Carter<sup>10</sup>.

**Prostate cancer survival** The survival module generates age at prostate cancer death for each subject based on his complete disease profile under screening and non-screening scenarios. We use **SEER** survival data from 1980 to 1987 to determine each case's age at death following prostate cancer diagnosis. Years of survival after diagnosis depend on age and stage; years are added to the age at clinical diagnosis, which is termed "lead-time delay." Model projections of disease-specific mortality rates prior to the PSA era are calibrated to match those observed in the population.

Disease-specific survival is irrelevant for latent subjects since, by definition, all latents die from some other cause before prostate cancer affects their lifespan.

## CONTRIBUTORS

## REFERENCES:

- 1 Legler J, Feuer E, Potosky A, Merrill R, Kramer B "The role of Prostate-Specific Antigen testing patterns in the recent prostate cancer incidence decline in the USA" in *Cancer Causes Control* 1998; 9: 519-527
- 2 Etzioni, Ruth, Cha, Raymond, Cowen, Mark E. "Serial Prostate Specific Antigen Screening For Prostate Cancer: A Computer Model Evaluates Competing Strategies" in *J Urol* 1999; 162: : 741-748
- 3 Cowen ME, Chartrand M, Weitzel WF "A Markov Model of The Natural History of Prostate Cancer" in *J Clin Epidemiol* 1994; 47: 1: 3-21
- 4 Whitmore WF "Background for screening: natural history and treatment. **EORTC Genitourinary Group Monograph 5: Progress and Controversies in Oncological Urology II.**" 1988; : 123-130
- 5 Whitmore, WF "Natural History of low-stage prostatic cancer and the impact of early detection." in *Urol Clin N Am* 1990; 17: 689-697
- 6 Etzioni R, Cha R, Feuer EJ, Davidov O "Asymptomatic Incidence and Duration in Prostate Cancer." in *American Journal of Epidemiology* 1998; 148: 775-785
- 7 Mariotto A, Etzioni R, Krapcho M, Feuer EJ "Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey" in *Cancer* 2007; in press:
- 8 Inoue LY, , Etzioni R, , Slate EH, , Morrell C, , Penson DF. "Combining longitudinal studies of PSA. " in *Biostatistics* 2004; 5: 3: 483-500
- 9 Oesterling JE Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM "Serum prostate-specific antigen in a community-based population of healthy men." in *JAMA* 1993; 270: 860-864
- 10 Carter HB, Morrell CH, Pearson JD, Brant LJ, et al. "Estimation of prostatic growth using serial Prostate-Specific Antigen measurements in men with and without prostate disease." in *Cancer Research* 1992; 52: 3323-3328



# ASSUMPTION OVERVIEW

## SUMMARY

The assumptions inherent in the FHCRC modeling approach are described below. When possible, we discuss the potential impact of these assumptions on our results.

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)

## BACKGROUND

Our model combines information on both the observed and latent aspects of the disease. Most of the assumptions made pertain to the latent natural history, but some also relate to the interface between the observed data and the latent disease history.

Our natural history model (onset and progression through disease stages) is based on two published studies: the Markov model of Cowen et al<sup>1</sup> and the asymptomatic onset and duration study of Etzioni et al<sup>2</sup>. Our first main assumption is that these are accurate reflections of the frequency of disease onset and the rates of disease progression through the clinical stages of prostate cancer as defined by the American Urological Association (AUA, aka Whitmore-Jewitt) staging system.

Our second main assumption comes when we link natural histories with clinical diagnosis. We use a matching algorithm that randomly selects natural histories at the correct time so as to match observed age- and stage-specific clinical incidence. While the algorithm achieves the desired result, it also induces a structure on the natural histories that ultimately are selected to be clinically diagnosed; these end up having earlier ages at onset and shorter stage durations than those natural histories that do not have a corresponding date of clinical diagnosis (these “latent” histories are ultimately our candidates for overdiagnosis). See Figure 1. A further assumption concerning clinical incidence is that this would have remained constant at its pre-PSA level (the level observed in 1987) in the absence of screening.

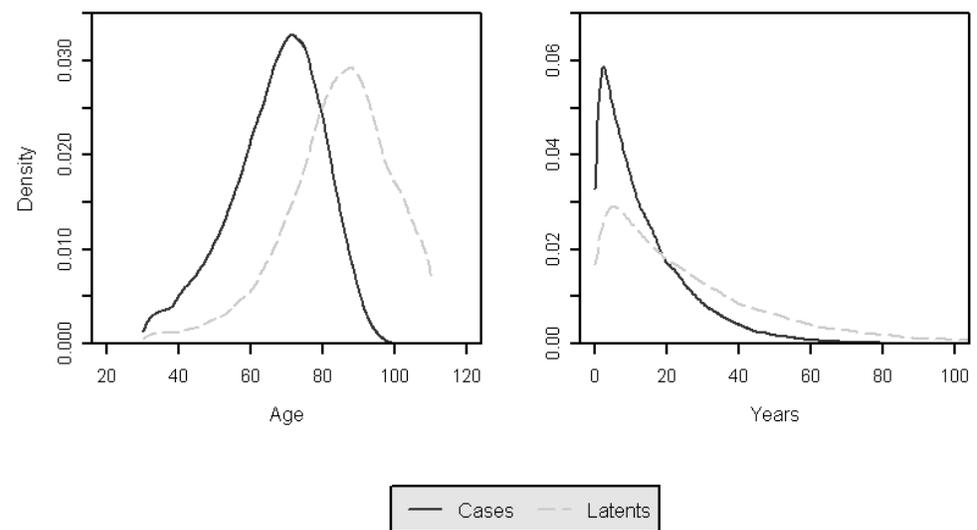


Figure 1. Age at disease onset (left) and stage A1 duration (right) distributions for cases and latents resulting from the matching algorithm.

One of the hidden assumptions that is implicit in our matching algorithm is that stage D2 disease is always symptomatic.

Each individual is assigned a PSA growth trajectory that is based on a meta-analysis of stored serum data, conducted by Inoue et al<sup>3</sup>. This dataset provides information on PSA growth for clinical cases by stage at clinical diagnosis. We assume that the PSA growth for latent cases is on average approximately half that of the PSA growth for the local-regional clinical cases. We link PSA growth for an individual with his natural history as follows: the quantile in the distribution of PSA slopes across individuals is set to be one minus the individual's quantile in the distribution of stage A1 durations. Thus, those individuals with the longest stage A1 durations receive the lowest annual PSA growth rates and vice versa.



Our next major assumption relates to screening and biopsy practices in the population. One of our observed inputs is a set of screening histories that we use to assign individuals to screening tests. These inputs have been rigorously estimated based on data from the 2000 NHIS and the linked [SEER](#) -Medicare databases (Mariotto et al<sup>4</sup>). We assume that a PSA level of 4.0 ng/ml is the trigger for biopsy, which may not be an accurate reflection of practice. Based on this assumption, we use biopsy frequencies from the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial (by age and PSA level) to assign men to receive a biopsy. We also assume that biopsy accuracy increases over time, in accordance with increases in the number of cores typically sampled at biopsy. Until the late 1980s, four-core biopsies were standard; by the mid-1990s six-core biopsies were standard, and by the early 2000s, 8-12 and extended-core biopsies were standard. We have conducted a literature review and assume that for cases with stage A1 disease, 6-core biopsy accuracy is 80%, 4-core biopsy accuracy is 2/3 of this amount, and extended-core biopsies are 100% accurate. For cases with more advanced disease, biopsy accuracy is assumed to be 100%.

Our final major assumption is one that underlies all of the screening models in CISNET, namely that stage shift implies survival shift. A case who would have been detected clinically in late stage but is shifted by PSA screening to detection in local-regional stage has his survival from clinical diagnosis re-set to reflect that of a local-regional stage case. We assume that if the distant-stage survival is relatively good (or poor), then this will be the case with the local-regional survival as well. To achieve this correspondence, the quantile of the shifted survival within the local-regional stage distribution is set to be equal to the quantile of the individual's original distant-stage survival in its distribution.

The validity of these assumptions is not tested directly. The model is validated by comparing a results with published studies (see [Validations Overview](#)) and the model-projected prostate cancer incidence and mortality trends are calibrated against those observed in [SEER](#) prior to the PSA era.

## ASSUMPTION LISTING

### **Mortality and clinical incidence:**

- Age-and stage-specific clinical incidence rates would have remained at 1987 levels in the absence of screening. Thus, this assumption does not explicitly take into account changes in the frequency of transurethral resections of the prostate (TURPs) during the PSA era. TURPs were closely linked with increases in prostate cancer incidence during the 1980s (Merrill et al<sup>5</sup>), but use of this procedure declined sharply in the 1990s following the dissemination of medical approaches to manage benign prostatic hyperplasia. Telesca et al<sup>6</sup> have recently estimated a background trend in incidence in the absence of PSA screening. This trend levels off after 1987 (i.e., it does not continue its historical increase), which is consistent with the constant secular trend in incidence assumed in the model.
- Age- and stage-specific incidence prior to 1973 is adequately approximated by the rates observed in 1973 to 1975.



- Stage D2 is symptomatic. Latents (individuals who are not clinically diagnosed in their lifetimes) must have an age at other-cause death that precedes their age at transition to AUA stage D2 (distant metastases).

#### **Asymptomatic onset:**

- Asymptomatic onset<sup>7</sup> used in the model is estimated from autopsy studies performed in the US in the 1950s. The model assumes that these adequately reflect the prevalence of latent disease. Based on these data, we have estimated that approximately 36% of men develop prostate cancer in their lifetimes. It is likely that this is an underestimate of the true amount of latent disease in the population (newer biopsy studies using more modern technology have yielded higher age-specific prevalences), but this assumption still yields sufficient latent cases for our modeling purposes.

#### **Disease progression and clinical presentation:**

- A Markov model is used to describe the progression of disease through AUA stages. Stage transition rates are based on work by Cowen<sup>8</sup>.
- Disease progression rates are independent of patient age, race, and date of disease onset. Stage durations are exponentially distributed and are not correlated with each other.

#### **PSA growth:**

- Pre-cancerous PSA growth is based on Oesterling et al<sup>9</sup>. PSA increases by approximately 3% annually.
- Cancerous PSA growth is derived from a study by Inoue et al<sup>10</sup>. This study analyzed data on mostly clinical cases. The mean annual growth rate for cases destined to be diagnosed in distant stage is 60%, and for cases destined to be diagnosed in local-regional stage it is 15%. For latents, we assume that the annual increase in PSA is half that estimated by Inoue et al<sup>10</sup> for local-regional cases.
- PSA growth accelerates at the time of entry into stage A1. It is also possible to specify a lag time (as a fraction of the stage A1 duration) until the start of PSA acceleration.
- PSA growth for an individual is inversely associated with the rate of disease progression from stage A1 to subsequent stages. An individual's quantile in the population distribution of PSA slopes is set to be one minus the individual's quantile in the population's distribution of stage A1 durations. Thus, those individuals with the longest stage A1 durations receive the lowest annual PSA growth rates and vice versa.



### PSA test schedule:

- The PSA dissemination schedule is based on the work of Mariotto et al<sup>4</sup>. A positive test is defined as PSA > 4.0 ng/ml.
- We do not model digital rectal exam (DRE) testing. We effectively assume that the frequency of DRE screening remains at pre-PSA-era levels. Thus, we do not capture any possible increase in the frequency of DRE as a consequence of the increase in PSA use. If use of DRE testing increases during the PSA era (e.g., DRE may be routinely conducted in conjunction with PSA screening), then this may lead to underascertainment of cases at screening tests because we will not be capturing any increase in detection due to DREs with positive results in the absence of positive PSA test results. However, we anticipate these to be relatively small in number.

### PSA test follow-up:

- Not all men with a positive PSA test will submit to a follow-up biopsy. The model assumes that the biopsy rate following a positive PSA test is similar to the one-year biopsy frequencies presented in Pinsky et al<sup>11</sup>.
- No men with a PSA test
- Biopsy accuracy parameters for stage A1 cases are based on our assessment of trends in number of cores based on an extensive literature review. We have determined that 4-core biopsies (assumed accuracy 53%) were standard at the start of the PSA era, 6-core biopsies (assumed accuracy 80%) were standard in the mid 1990s, and higher numbers of cores (assumed accuracy 100%) were standard by the early 2000s.
- We assume that biopsy is 100% accurate when disease has progressed beyond stage A1.

### Survival following diagnosis:

- The major survival benefit assumption for the model is that prostate cancer is a disease whose natural progression can be interrupted by intervention at an early stage; specifically, stage shift (from distant to local-regional) implies survival shift (from distant-stage survival to local-regional-stage survival).
- We do not model within-stage shifts, so a case shifted from regional to local or within local stage receives no survival benefit.
- We assume no improvements in survival during the PSA era due to treatment since we are trying to isolate the effect of the screening-induced stage shift on population mortality. Thus, in the absence of PSA testing, we assume that disease-specific survival observed for cases diagnosed from 1987 to 2000 would have been the same as the survival observed for cases diagnosed from 1980 to 1987.
- Among stage-shifted cases, the shifted survival begins declining only once the lead time has elapsed, i.e., at the time of clinical diagnosis. Thus, we explicitly disallow negative survival benefit under screening.



- The survival from clinical diagnosis without and survival with screening are correlated by quantile: the quantile of the shifted survival within the local-regional stage distribution is set to be equal to the quantile of the individual's original distant-stage survival in its distribution.

## REFERENCES:

- <sup>1</sup> Cowen ME, Chartrand M, Weitzel WF "A Markov Model of The Natural History of Prostate Cancer" in J Clin Epidemiol 1994; 47: 1: 3-21
- <sup>2</sup> Etzioni R, Cha R, Feuer EJ, Davidov O "Asymptomatic Incidence and Duration in Prostate Cancer." in American Journal of Epidemiology 1998; 148: 775-785
- <sup>3</sup> Inoue LY, , Etzioni R, , Slate EH, , Morrell C, , Penson DF. "Combining longitudinal studies of PSA. " in Biostatistics 2004; 5: 3: 483-500
- <sup>4</sup> Mariotto A, Etzioni R, Krapcho M, Feuer EJ "Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey" in Cancer 2007; in press:
- <sup>5</sup> Merrill RM, Feuer EJ, Warren JL, Schussler N, Stephenson RA, "Role of transurethral resection of the prostate in population-based prostate cancer incidence rates" in Am J Epidemiol 1999; 150: 8: 848-60
- <sup>6</sup> Telesca D, Etzioni R, Gulati R "Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends" in Biometrics 2007; in press:
- <sup>7</sup> Etzioni R, Cha R, Feuer EJ, Davidov O "Asymptomatic Incidence and Duration in Prostate Cancer." in American Journal of Epidemiology 1998; 148: 775-785
- <sup>8</sup> Cowen ME, Chartrand M, Weitzel WF "A Markov Model of The Natural History of Prostate Cancer" in J Clin Epidemiol 1994; 47: 1: 3-21
- <sup>9</sup> Oesterling JE Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM "Serum prostate-specific antigen in a community-based population of healthy men." in JAMA 1993; 270: 860-864
- <sup>10</sup> Inoue LY, , Etzioni R, , Slate EH, , Morrell C, , Penson DF. "Combining longitudinal studies of PSA. " in Biostatistics 2004; 5: 3: 483-500
- <sup>11</sup> Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK "Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial" in J Urol 2005; 173: 3: 746-50



# PARAMETER OVERVIEW

## SUMMARY

This page describes the model inputs with which we have developed the FHCRC prostate cancer microsimulation.

See the [Assumption Overview](#) for detailed assumptions associated with these model inputs.

## BACKGROUND

### PARAMETER LISTING OVERVIEW

The FHCRC microsimulation comprises five fundamental modules.

#### [Natural history and clinical presentation:](#)

- All-cause mortality data are based on Berkeley life tables containing annual mortality rates by birth cohort from birth year 1900 to 2000 by single year ages from 0 to 119. We use data provided by National Cancer Institute (NCI) to subtract out prostate cancer death rates from 1950 to 2000, yielding other cause (i.e., not due to prostate cancer) death rates. We use these to generate age at other-cause death.
- A cumulative distribution of age at asymptomatic onset is computed from the results of Etzioni et al<sup>1</sup> and is used to generate an age at disease onset for each individual. If the age at onset precedes other-cause death, the individual becomes asymptomatic during his lifetime.
- Stage transition rates from Cowen et al<sup>2</sup> are used to generate clinical stage durations from stage A1 through the end of stage D2.
- A year of birth distribution (uniform between 1895 and 1950) produces a multi-cohort population including men aged 50 to 84 for all years between 1980 and 2000. See Figure 1.
- The distribution of the lifetime probability of clinical incidence is created using [Dev Can](#) software provided by NCI. Inputs consist of [SEER](#) age-specific incidence rates from 1973 to 1987. We assume that incidence prior to 1973 is approximated by the rate observed in 1973 and that incidence after 1987 in the absence of PSA screening is approximated by that observed in 1987. We then use [Dev Can](#) to generate a cumulative distribution of age at clinical diagnosis in the absence of other-cause death for each birth cohort in the model. This is used to generate the clinical histories that correspond to the cases.
- The stage distribution at clinical presentation is based on [SEER](#) data. Prior to 1973 we assume the stage distribution to be approximated by that observed from 1973 to 1977. After 1987, we assume that the stage distribution in the absence of screening is approximated by the distribution observed from 1983 to 1987.

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)

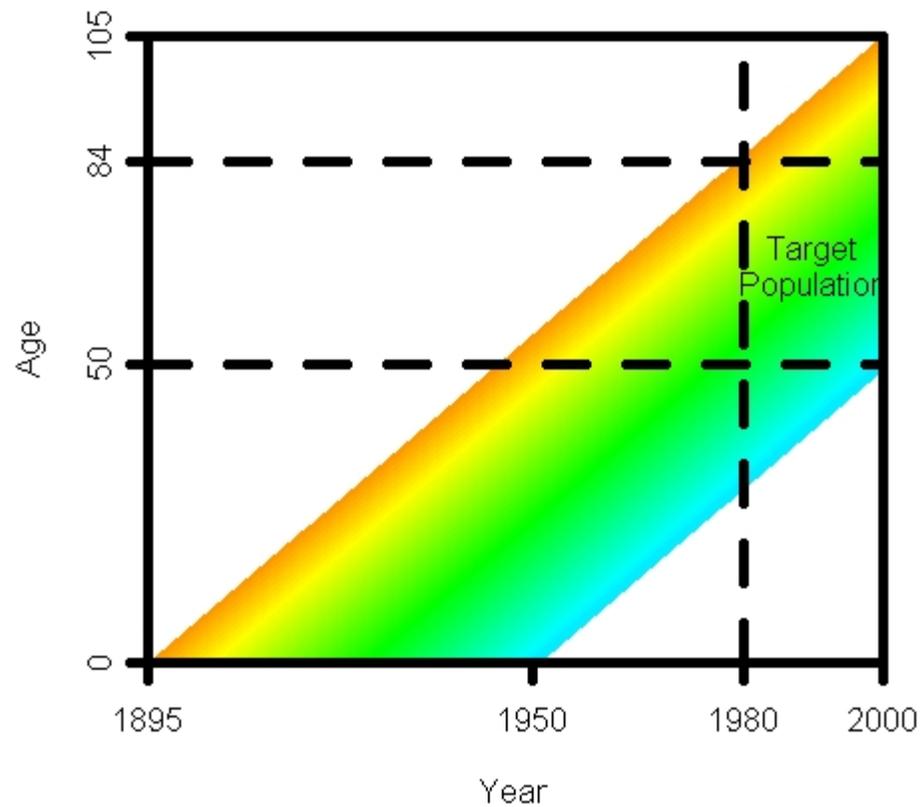


Figure 1. Illustration of birth years corresponding to target population for the ages (50-84) and years (1980-2000) of interest.

#### Screening: PSA testing and biopsy follow-up:

- A schedule for PSA testing is assigned to each subject based on the PSA dissemination model<sup>3</sup> developed by IMS and provided to CISNET modelers by our collaborators at NCI.
- The probability of follow-up biopsy after a positive PSA result is based on data from Pinsky et al<sup>4</sup>, who estimated the likelihood of a biopsy within one year of a PSA test by PSA level, age, and calendar year in the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial.
- Biopsy accuracy (ability to detect existing disease) for men with stage A1 disease is a function of the number of biopsy cores (4, 6, or more than 6). Based on an extensive review of the literature, we have estimated that prior to 1990, 4-core biopsies were standard, by 1995 6-core biopsies were standard, and by the early 2000s, 8- to 12-core biopsies were standard<sup>5</sup>. Following Presti et al<sup>6</sup>, we have utilized 80% as the sensitivity of 6-core biopsies, 100% as the sensitivity for extended-core biopsies, and 2/3 of 80% as the sensitivity for 4-core biopsies.

#### PSA growth:

- The distribution of PSA levels beginning at age 45 serves as an anchor point for the PSA growth curve. This is drawn from a lognormal distribution fit to the distribution of PSAs for 40 to 49 year olds in Oesterling et al<sup>7</sup>.
- Mean annual PSA growth rate for healthy subjects is 3% percent per year from Oesterling et al<sup>7</sup>.
- Within-person standard deviation of PSA level for healthy subjects is  $\exp(\sqrt{0.05})$ .
- Annual PSA growth rate after disease onset is modeled with an exponential growth<sup>8</sup>. Specifics of the model are:
  - Average annual percent change for distant-stage cases is 60%.
  - Average annual percent change for local/regional-stage cases is 15%.
  - Average annual percent change for latents is 6.5%.
  - After disease onset, between-individual standard deviation of annual percent change in PSA is 10% of mean growth rate.
  - Individual-specific annual percent change in PSA is determined by quantile  $q$  in the population distribution of PSA growth rates where  $1 - q$  is the individual's quantile in the initial stage distribution.

**Survival:** Survival inputs consist of relative survival curves from SEER, by age, stage, and calendar year of diagnosis. Data from cases diagnosed between 1973 and 1987 are used, i.e., we end at the start of the PSA era. We split years of diagnosis into three calendar periods: 1973-1977, 1987-1982, 1983-1987. For diagnoses prior to 1973 we apply the 1973-1977 results and for diagnoses after 1987 we apply the 1983-1987 results. Thus we assume no improvement in age- and stage-specific survival from clinical diagnosis during the PSA era, i.e., we do not model any increases in survival that might be due to treatment changes.

## REFERENCES:

- <sup>1</sup> Etzioni R, Cha R, Feuer EJ, Davidov O "Asymptomatic Incidence and Duration in Prostate Cancer." in American Journal of Epidemiology 1998; 148: 775-785
- <sup>2</sup> Cowen ME, Chartrand M, Weitzel WF "A Markov Model of The Natural History of Prostate Cancer" in J Clin Epidemiol 1994; 47: 1: 3-21
- <sup>3</sup> Mariotto A, Etzioni R, Krapcho M, Feuer EJ "Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey" in Cancer 2007; in press:
- <sup>4</sup> Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK "Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial" in J Urol 2005; 173: 3: 746-50
- <sup>5</sup> Amling C "Personal communication" 2006;
- <sup>6</sup> Presti JCJ, Chang JJ, Bhargava V, Shinohara K "The optimal systematic prostate biopsy scheme should include 8 rather than 6 biopsies: Results of a prospective clinical trial" in J Urol 2000; 163: 1: 163-6
- <sup>7</sup> Oesterling JE, Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM "Serum prostate-specific antigen in a community-based population of healthy men." in JAMA 1993; 270: 860-864
- <sup>8</sup> Inoue LY, , Etzioni R, , Slate EH, , Morrell C, , Penson DF. "Combining longitudinal studies of PSA. " in Biostatistics 2004; 5: 3: 483-500

# COMPONENT OVERVIEW

## SUMMARY

For each major module of the microsimulation, implementation details are discussed in this section.

## OVERVIEW

The FHCRC microsimulation comprises five fundamental modules; natural history, clinical diagnosis, PSA production, PSA screening, and survival. These modules are outlined in the figure below, and implementation details for each are discussed in broader detail in the Component Listing section. In addition, an output module collates the model results and creates summary output reports.

- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Validations Overview
- Key References

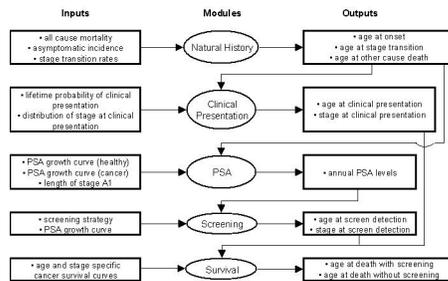


Figure 1. Overview of the model components

along with the inputs and outputs of each.

## COMPONENT LISTING

The FHCRC microsimulation comprises five fundamental modules. Implementation details for each module are discussed in this section.

**Natural history:** This module generates  $N$  disease histories and  $N$  clinical histories that are later used in the clinical diagnosis module.

### *Disease histories*

Generates age at asymptomatic disease onset and ages at stage transitions. We assume stage durations are distributed independently according to exponential distributions. Disease stage is converted from American Urological Association (AUA) staging to SEER historic stage using the mapping shown in the following table<sup>1</sup>.

| AUA Staging | SEER Historic Staging |
|-------------|-----------------------|
| A1, A2, B   | Local                 |
| C, D1       | Regional              |
| D2          | Distant               |



### *Clinical histories*

Conditional on age of birth generate age of diagnosis in the absence of other-cause death and (independently) generate age at other-cause death. Two types of clinical histories result: (1) histories that include an age at clinical diagnosis prior to the age at other-cause death and (2) histories that include only the age at other-cause death.

### **Clinical diagnosis:**

Histories of the first type are matched to appropriate natural histories; for example, a natural history that has a birth year of 1920, disease onset at age 45, and progression to distant stage disease at age 60 might be matched to a clinical history from the 1920-1925 birth cohort that specifies local-regional diagnosis at age 58 in 1978. The matched clinical histories are called “cases.”

Histories of the second type are paired with the remaining unmatched natural histories so that, within each pair, the age at other-cause death in the clinical history precedes the age at transition to distant stage disease in the natural history. This operation effectively assumes that advanced prostate cancer is generally symptomatic and would not remain undetected during the lifetime of the patient. Matched histories of this type are labeled “latents.” The latents include both latent cases (those who have disease onset but not clinical diagnosis within their lifetimes) and healthy men (those who never have disease onset within their lifetimes—these men account for approximately 60% of the total population, in agreement with the autopsy studies).

Clinical histories and disease histories are processed in batches to control memory usage. Each subset of disease histories is searched for matches with the clinical histories in the current batch. Disease histories that do not match any of the current clinical histories are retained for comparison with subsequent batches of clinical histories. For each clinical history, one matching disease history is selected and removed from further consideration. This process is repeated until all possible clinical histories have been matched. Unmatched clinical and natural histories generally constitute less than 1% of the total and are dropped from the population.

**PSA production:** Assigns PSA levels to each individual's PSA screening events. PSA



growth rates differ for cancerous and non-cancerous states, and by cancerous disease stage (local/regional or distant). Further details are on the [Parameter Overview](#) page.

**Serial PSA screening:** The screening module assigns screening schedules to subjects. Screening dissemination is based on the results of Mariotto et al<sup>2</sup>, who used retrospective data from the linked [SEER](#) -Medicare database and the National Health Interview Survey.

**Survival:** This module applies only to cases as latents do not benefit from screening. The module generates three ages at death: age at death due to prostate cancer without screening (XXCaClin), age at death due to prostate cancer with screening (XXCaScrn), and age at death due to causes other than prostate cancer (XXoc).

A subject's age at death in the absence of screening is the smaller of XXoc and XXCaClin; his age at death in the presence of screening is the smaller of XXoc and the larger of XXCaClin and XXCaScrn. Survival benefit is the difference between the age at death in the absence of screening and the age at death in the presence of screening.

Additional details are available in the [Screen Benefit Summary](#) page which can be accessed from the [Output Overview](#) page.

## REFERENCES:

- <sup>1</sup> Etzioni R, Legler JM, Feuer EJ, Merrill RM, Cronin KA, Hankey BF "Cancer Surveillance Series: Interpreting Trends in Prostate Cancer-Part III: Quantifying the Link Between Population Prostate-Specific Antigen Testing and Recent Declines in Prostate Cancer Mortality" in Journal of National Cancer Institute 1999; 91: 1033-1039
- <sup>2</sup> Mariotto A, Etzioni R, Krapcho M, Feuer EJ "Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey" in Cancer 2007; in press:



# OUTPUT OVERVIEW

## SUMMARY

This page describes the principal outputs of the FHCRC prostate cancer microsimulation, and their importance in understanding prostate cancer trends in the US population.

## OVERVIEW

The major outputs of the FHCRC model are as follows:

- Age-specific and (age-adjusted) stage-specific incidence of prostate cancer after 1987 in the absence and presence of PSA testing.
- Mean sojourn time (time from disease onset to clinical diagnosis). This can be computed for cases only, i.e., conditional on clinical diagnosis happening before other-cause death, in which case we refer to it as a “conditional sojourn time,” or it can be computed for all men with disease onset ignoring other-cause death, in which case we refer to it as an “unconditional sojourn time.” The sojourn time always starts at onset and ends at the date of clinical diagnosis.
- Mean lead time associated with PSA screening (time from screen to clinical detection). Like the sojourn time, this can be computed for cases only, i.e., conditional on clinical diagnosis happening before other-cause death, in which case we refer to it as a “conditional lead time,” or it can be computed for all screen-detected individuals, ignoring other-cause death, in which case we refer to it as an “unconditional lead time.” The lead time always starts at screen detection and ends at the date of clinical diagnosis.
- Age-specific and age-adjusted prostate cancer mortality rates after 1987 in the absence and presence of PSA screening. The difference between these two is our measure of screening benefit (see [Screen Benefit Computation](#)).

## OUTPUT LISTING



# RESULTS OVERVIEW

## SUMMARY

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)

## OVERVIEW

Selected numerical and graphical results from the microsimulation are explained below, including results for survival benefit, mortality and mortality reduction in the presence of screening, incidence in the presence of screening, and estimates for the mean lead-time.



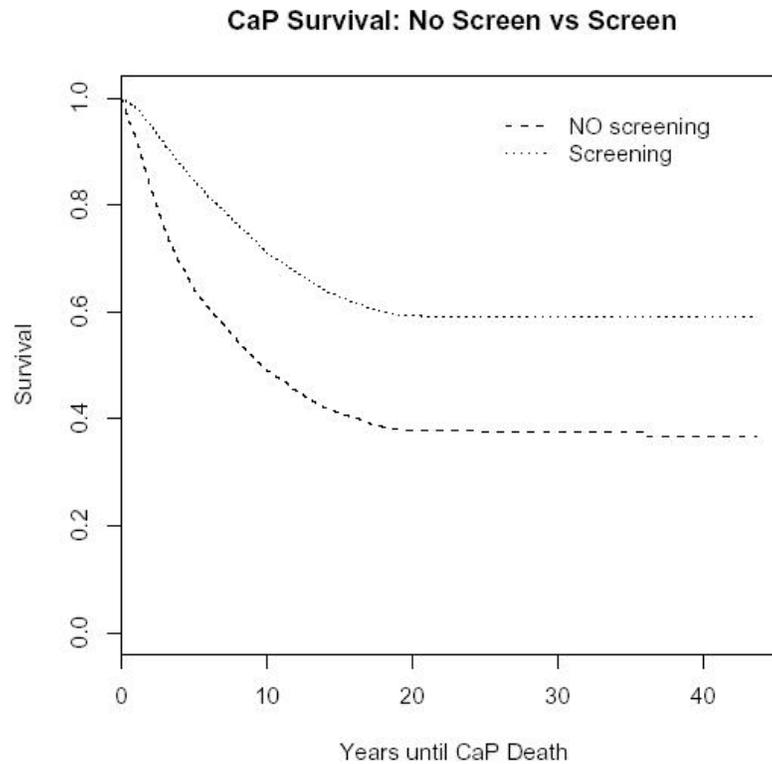
## RESULTS LIST

Fred Hutchinson CRC (PCSIM)  
Results Overview  
Results List



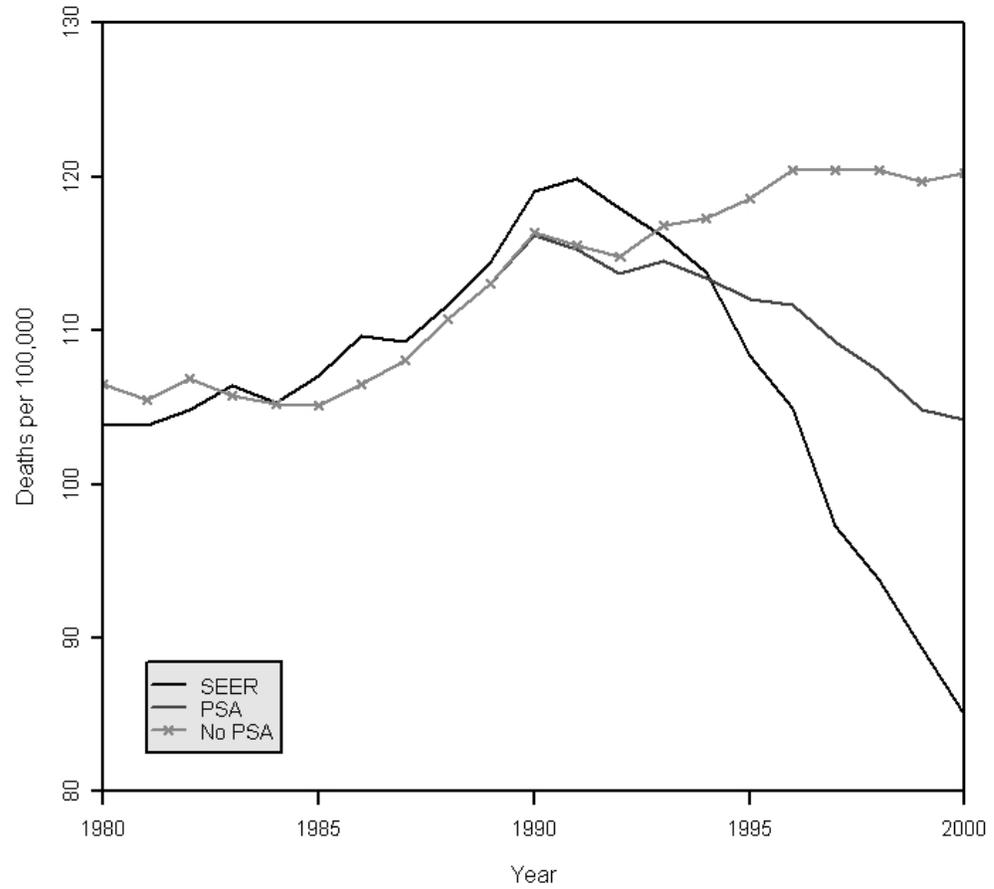
### Survival benefit:

The model predicts a survival benefit from PSA screening. Screening and the corresponding stage shift imply a relative risk of 0.48. The following figure shows the relative survival among modeled cases with and without screening.



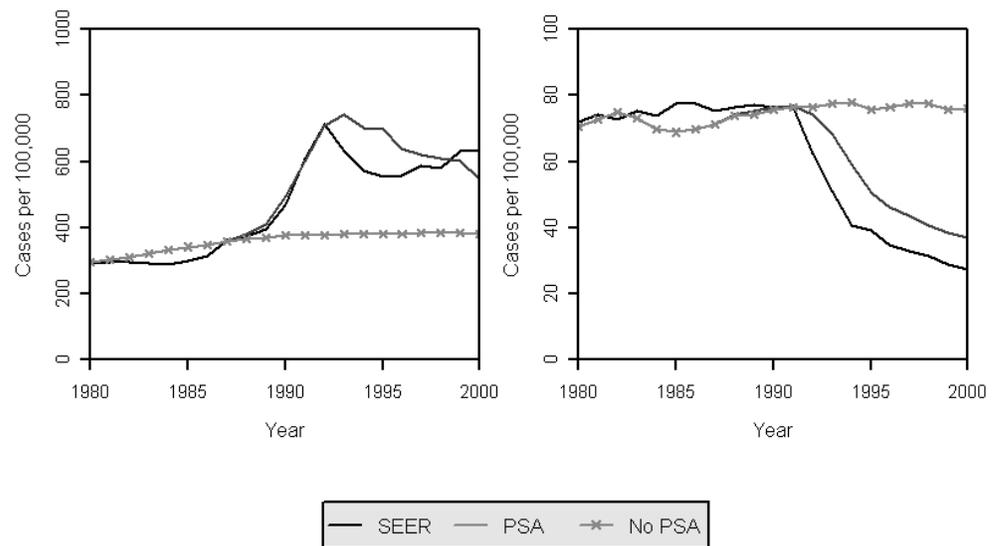
### Mortality reduction:

In the absence of PSA testing, the model predicts that mortality due to prostate cancer would have increased throughout the 1990s. Model results indicate that PSA testing may be responsible for about half of the reduction in mortality.



Note: For the years 1980 to 1987, the figure shows model validation; results from 1988 to 2000 are model results.

**Stage-specific incidence of prostate cancer:**





Note 1: For the years 1980 to 1987, the figure shows model validation; results from 1988 to 2000 are model results.

Note 2: Stage-specific incidence is compared to a projection of incidence from SEER that assumes that the stage distribution among unstaged cases is equivalent to stage distribution among staged cases.

**Sojourn and lead times associated with PSA testing:**

Sojourn time is the length of time from preclinical disease onset to clinical diagnosis. Lead time is the length of time by which diagnosis is advanced by screening, or the difference between the age at screen diagnosis and age at diagnosis in the absence of screening. Table 1 shows min, mean, and max sojourn and lead time estimates (in years) from the model based on random samples of 1000 cases across 10 simulations. Sojourn times are by age group at onset and lead times are by age group at screen detection. Min (Max) times are the minimum (maximum) of the mean times across the 10 simulations, and  $\bar{n}$  is the mean number of subjects in each age group entering into calculations

| Table 1. Sojourn and lead time estimates |               |      |      |           |            |      |      |           |
|--|---------------|------|------|-----------|------------|------|------|-----------|
| Age                                      | Sojourn times |      |      |           | Lead times |      |      |           |
|  | Min           | Mean | Max  | $\bar{n}$ | Min        | Mean | Max  | $\bar{n}$ |
| <b>50-59</b>                             | 12.9          | 13.7 | 14.5 | 164.8     | 8.8        | 11.3 | 13.8 | 11.5      |
| <b>60-69</b>                             | 8.7           | 9.1  | 9.6  | 287.7     | 5.5        | 7.2  | 8.6  | 24.3      |
| <b>70-79</b>                             | 6.3           | 6.6  | 6.9  | 324.4     | 4.4        | 5.3  | 6.0  | 26.1      |
| <b>80-84</b>                             | 4.4           | 5.1  | 5.4  | 102.2     | 2.6        | 4.3  | 5.4  | 9.3       |
| <b>Age-adjusted</b>                      | 8.3           | 8.6  | 8.8  | —         | 5.5        | 6.8  | 7.6  | —         |

# VALIDATIONS OVERVIEW

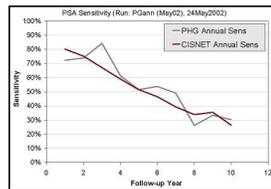
Model development requires many input parameters where reliable source data may not be available. This includes parameters describing disease natural history, the key points of which are generally not observable. Our model results are calibrated to prostate cancer incidence and mortality in the pre-PSA era and validated against data on PSA test characteristics, sojourn and lead times, and the cumulative probability of disease diagnosis in the presence of other-cause death.

- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Validations Overview
- Key References

## Calibrations and Validations

- **Validation of PSA sensitivity in a case-control study.**

PSA sensitivity was validated by comparing results from the model to a retrospective case-control study by Gann et al<sup>1</sup>. This study sampled men who had enrolled in the Physicians' Health Study and had provided a blood sample at the time of enrollment in 1980. Cases consisted of men diagnosed with prostate cancer within 10 years after enrollment; controls were age-matched to cases and had not been diagnosed with prostate cancer by the end of follow-up. The stored blood samples were retrospectively assayed for PSA and the sensitivity of PSA to detect disease diagnosed within  $k$  years (where  $k$  ranges from 1 to 10) was estimated. We simulated this study design and computed corresponding estimates of screen sensitivity by interval from test to clinical diagnosis.



- **Validation of model incidence to age 85 in presence of other-cause death**



The CISNET model takes as input estimates of the cumulative probability of clinical incidence in the absence of other-cause death. Disease cases consist of individuals whose clinical diagnosis precedes their other-cause death. Thus, as a validation exercise, we computed the model-generated probability of clinical diagnosis by age 85 by birth cohort and compared it with that produced by [Dev Can](#) software. This is shown below.

| Birth cohort | <a href="#">Dev Can</a> output | Model output |
|--------------|--------------------------------|--------------|
| 1900-1905    | 0.0905                         | 0.0874       |
| 1905-1910    | 0.1000                         | 0.0961       |
| 1910-1915    | 0.1114                         | 0.1055       |
| 1915-1920    | 0.1217                         | 0.1150       |
| 1920-1925    | 0.1284                         | 0.1210       |
| 1925-1930    | 0.1321                         | 0.1241       |
| 1930-1935    | 0.1346                         | 0.1254       |

- **Validation of sojourn and lead time estimates produced by the model.**

Sojourn and lead times by birth cohort and calendar year of diagnosis, aggregated across 50 million subjects, are summarized in the [Results Overview](#). Sojourn time is the time from disease onset to clinical diagnosis and is computed for cases by age group at onset. Since the distribution of age at onset is the same for all birth cohorts, any between-cohort differences in sojourn times result from differences in clinical diagnosis rates. The lead time is the time from screen detection to clinical diagnosis and is computed for screen-detected cases by age group at detection.

Model estimates of mean sojourn and lead times validate well with other studies. Our overall, model-projected sojourn time is close to the estimate of 10 to 12 years obtained by Etzioni et al and slightly lower than the estimate of 12.7 years obtained by Draisma et al<sup>2</sup>. The estimated mean lead time among clinical cases is between the 5 years obtained by Gann et al<sup>1</sup> and Telesca et al<sup>3</sup> and the 7 years implied by Tsodikov et al<sup>4</sup>.

- **Calibration of the model to prostate cancer incidence and mortality prior to the PSA era.**

Calibration involves informal optimization in a high-dimensional parameter space, which is aided by an internal linear interpolation algorithm that smooths inputs provided by age group and calendar interval. In calibrating the model so that it replicates prostate cancer incidence and mortality levels prior to the PSA era, we vary the mean stage A1 duration, the minimum local-regional stage duration, the precise stage distribution at clinical diagnosis, the PSA growth rates for latents and for cases clinically diagnosed in distant stage, and the case-latent ratio in the modeled population. No formal estimation procedure is conducted to identify the best-fitting input values for these parameters. Regarding the case-latent ratio, this is set originally in the clinical diagnosis module, which uses [Dev Can](#) to compute the cumulative probability of clinical diagnosis—individuals with clinical diagnosis in their lifetimes become cases and the rest of the population become latents. Only 1 out of 12 latents is preserved for computational efficiency, and the final model results upweight (i.e., inflate) any contributions from these latents by a factor of 12. However, for model calibration purposes we have found that a factor of 14 produces pre-PSA incidence and mortality rates that are considerably closer to those observed. Hence the incidence and mortality plots in the [Results Overview](#) use this as the latent inflation factor.

The calibrated model generates 5 million disease and screening histories and aggregates the resulting age- and stage-specific incidence rates over birth cohorts to produce results comparable to [SEER](#) rates, which are age-adjusted for the same age groups.

## REFERENCES:

- <sup>1</sup> Gann PH, Hennekens CH, Stampfer MJ “A prospective evaluation of plasma Prostate-Specific Antigen for detection of prostate cancer” in *JAMA* 1995; 273: 289-294
- <sup>2</sup> Draisma G, Boer R, Otto SJ, van der Crujisen IW, Damhuis RA, Schroder FH, de Koning HJ, “Lead times and overdetection due to prostate-specific antigen screening: Estimates from the European Randomized Study of Screening for Prostate Cancer” in *J Natl Cancer Inst* 2003; 95: 12: 868-78
- <sup>3</sup> Telesca D, Etzioni R, Gulati R “Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends” in *Biometrics* 2007; in press:
- <sup>4</sup> Tsodikov A, Szabo A, Wegelin J “A population model of prostate cancer incidence” in *Stat Med* 2006; 25: 16: 2846-66



Fred Hutchinson CRC (PCSIM)  
SEER

# SEER

The [Surveillance, Epidemiology, and End Results](#) program of the NCI :

[From the SEER website:](#)

"The [SEER Program](#) of the NCI is the most authoritative source of information on cancer incidence and survival in the United States. Information on more than 2.5 million cancer cases is included in the SEER database, and approximately 160,000 new cases are accessioned each year within the SEER catchment areas. SEER [data](#), [publications](#), and [resources](#) are available free of charge."

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)



Fred Hutchinson CRC (PCSIM)  
USCB

# USCB

The Bureau of the Census within the United States Department of Commerce

- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Validations Overview
- Key References



# Nc Hs

The [National Center for Health Statistics](#) is a division of the [CDC](#).

**From the NCHS website:**

"NCHS is the Federal Government's principal vital and health statistics agency. Since 1960, when the National Office of Vital Statistics and the National Health Survey merged to form NCHS, the agency has provided a wide variety of data with which to monitor the Nation's health. Since then, NCHS has received several **legislative mandates and authorities**.

"The NCHS is a part of the [CDC](#), [US Department of Health and Human Services](#). To meet priority data needs for public health, NCHS works closely with other Federal agencies as well as researchers and academic institutions.

"NCHS data systems include data on vital events as well as information on health status, lifestyle and exposure to unhealthy influences, the onset and diagnosis of illness and disability, and the use of health care. These data are used by policymakers in Congress and the Administration, by medical researchers, and by others in the health community."

- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Validations Overview
- Key References



# Nc I

The [National Cancer Institute](#), part of the [NIH](#).

[From the NCI website:](#)

"The [NCI](#) is a component of the [NIH](#), one of eight agencies that compose the Public Health Service (PHS) in the [US Department of Health and Human Services](#). The [NCI](#), established under the National Cancer Act of 1937, is the Federal Government's principal agency for cancer research and training. The National Cancer Act of 1971 broadened the scope and responsibilities of the [NCI](#) and created the National Cancer Program. Over the years, legislative amendments have maintained the NCI authorities and responsibilities and added new information dissemination mandates as well as a requirement to assess the incorporation of state-of-the-art cancer treatments into clinical practice.

"The National Cancer Institute coordinates the National Cancer Program, which conducts and supports research, training, health information dissemination, and other programs with respect to the cause, diagnosis, prevention, and treatment of cancer, rehabilitation from cancer, and the continuing care of cancer patients and the families of cancer patients."

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)



Fred Hutchinson CRC (PCSIM)  
Dev Can

# DEV CAN

[Dev Can](#) takes cross-sectional counts of incident cases from the standard areas of the Surveillance, Epidemiology, and End Results (SEER) Program conducted by the National Cancer Institute, and mortality counts for the same areas from data collected by the National Center for Health Statistics, and uses them to calculate incidence and mortality rates using population estimates from census data for these areas. These rates are converted to the probabilities of developing or dying from cancer for a hypothetical population.

[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Validations Overview](#)  
[Key References](#)

Software to perform the calculations is maintained and available free of charge from NCI: <http://srab.cancer.gov/devcan/devcan.html>



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)

# SCREEN BENEFIT SUMMARY

## SUMMARY

This section discusses how the FHCRC prostate cancer model computes survival benefit due to screening. The survival benefit computation compares a population strategy of PSA screening and diagnostic follow-up with a baseline strategy reflecting the level of diagnostic intervention in 1987, just prior to the start of the PSA era. This does not include PSA screening but may include other interventions that lead to detection of prostate cancer such as digital rectal exam (DRE) or trans-urethral resection of the prostate (TURP). We do not explicitly consider changes over time in these interventions.

## OVERVIEW

Of screened subjects, only Cases (those who would have been clinically diagnosed in the absence of PSA) receive screening benefits. The primary mechanism to achieve benefit is by a shift from distant stage to regional or local-regional stage.

The model does not link pre-diagnosis progression rates with post-diagnosis prognosis. Thus it does not reflect any length bias that may be present in population screening.

Lead time bias is not an issue in our model because we generate survival beginning at clinical diagnosis in both the absence and presence of screening. Thus, if a case is shifted by screening from a distant to a local-regional stage, then his new (local-regional) survival time begins from his date of original clinical diagnosis.

## BENEFIT: INPUT OR OUTPUT?

Individual screening benefit is an output calculated from three important parameters generated in the model:

- Age-at-death-due-to-clinically-diagnosed-disease (XXCaClin)
- Age-at-death-due-to-screen-diagnosed-disease (XXCaScrn)
- Death-due-to-other-causes (XXOc)

The screening benefit calculation looks like this:

If  $XXCaClin > XXCaScrn$  and  $XXOc < XXCaClin$ ; otherwise, if  $XXOc > XXCaClin$   
Population screening benefit is estimated from the disease-specific mortality curves generated by the model in the absence and presence of screening. For any given year, this is captured by the estimated percentage of the mortality decline attributable to PSA screening, given by  $100 \times (M_a - M_p) / (M_a - M_o)$  where  $M_a$  and  $M_p$  denote mortality in the absence and presence of PSA and  $M_o$  is observed mortality.



## BENEFIT: EXPLICIT OR IMPLICIT?

Benefit arises from the difference between explicitly modeled survival times for Cases with and without screening. Population screening benefit is estimated from the disease-specific mortality curves generated by the model in the absence and presence of screening. For any given year, this is captured by the estimated percentage of the mortality decline attributable to PSA screening, given by  $100 \times (M_a - M_p) / (M_a - M_o)$ , where  $M_a$  and  $M_p$  denote mortality in the absence and presence of PSA and  $M_o$  is observed mortality.

## ATTRIBUTES DRIVING SURVIVAL

We use survival curves from SEER for local, regional, and distant stage disease. The lookup parameters are year, age, and stage at diagnosis. The same tables are used for both clinical- and screened-detected subjects.

### Clinical survival:

- Stage at clinical diagnosis
- Age at clinical diagnosis
- Calendar year periods of clinical diagnosis: 1973-1977 (used for diagnoses prior to 1973), 1978-1982, and 1983-1987 (used for diagnoses after 1987)

### Screen survival:

- Stage at screen diagnosis
- Age at screen diagnosis

## ATTRIBUTES CHANGED BY SCREENING

Age and stage at diagnosis may be changed by screening and may be used to recalculate disease-specific survival.

## CORRELATION AND LINKING

The clinical and screen-diagnosed survivals for a given subject are computed at the same quantile of their respective survival distributions.

For the baseline model, benefit is not linked to any other attributes of the natural history model.

## ISSUES AND ARTIFACTS

Individuals may not die of prostate cancer during their lead time or sojourn time. This is enforced by beginning disease-specific survival times at the original date of clinical diagnosis, whether in the absence or presence of screening. We refer to this as “lead time delay”; the survival under screening is delayed until the date of clinical diagnosis.

Latents (individuals with no clinical detection during their lifetimes) do by definition die of other causes within their sojourn and/or lead time.



## OTHER ISSUES

### **What if XXCaScr is less than XXCaClin?**

It can happen that, by “luck of the draw”, a case is given an age-at-death-due-to-screened-PCa that is less than age-at-death-due-to-clinical-PCa. In this case, we move XXCaScr to be at the same date as XXCaClin. This is a relatively rare occurrence because of the lead time delay and the linkage by quantile of the screen- and clinically-diagnosed survival times.



Fred Hutchinson CRC (PCSIM)  
Cd C

## Cd C

The [Centers for Disease Control and Prevention](#), an agency of the US Department of Health and Human Services.

### [From the CDC website:](#)

"The Centers for Disease Control and Prevention (CDC) is recognized as the lead federal agency for protecting the health and safety of people—at home and abroad, providing credible information to enhance health decisions, and promoting health through strong partnerships. CDC serves as the national focus for developing and applying disease prevention and control, environmental health, and health promotion and education activities designed to improve the health of the people of the United States.

"CDC, located in Atlanta, Georgia, USA, is an agency of the [US Department of Health and Human Services](#). [Dr. Jeffrey P. Koplan](#) is the Director."

[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Validations Overview](#)  
[Key References](#)



# DHHS

The [United States Department of Health and Human Services](#) comprises the following agencies:

[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Validations Overview](#)  
[Key References](#)

- [Office of the Secretary of Health and Human Services \(OS\)](#)
- [Administration for Children and Families \(ACF\)](#)
- [Administration on Aging \(AOA\)](#)
- [Agency for Healthcare Research and Quality \(AHRQ\)](#)
- [Agency for Toxic Substances and Disease Registry \(ATSDR\)](#)
- [Centers for Disease Control and Prevention \(CDC\)](#)
- [Centers for Medicare & Medicaid Services \(CMS\)](#)
- [Food and Drug Administration \(FDA\)](#)
- [Health Resources and Services Administration \(HRSA\)](#)
- [Indian Health Service \(IHS\)](#)
- [National Institutes of Health \(NIH\)](#)
- [Program Support Center \(PSC\)](#)
- [Substance Abuse and Mental Health Services Administration \(SAMHSA\)](#)



# Ni H

The [National Institutes of Health](#) is an agency of the [US Department of Health and Human Services](#).

[From the NIH website:](#)

"Begun as a one-room Laboratory of Hygiene in 1887, the [NIH](#) today is one of the world's foremost medical research centers, and the Federal focal point for medical research in the U.S.

**"The [NIH](#) mission is to uncover new knowledge that will lead to better health for everyone. [NIH](#) works toward that mission by:**

- conducting research in its own laboratories;
- supporting the research of non-Federal scientists in universities, medical schools, hospitals, and research institutions throughout the country and abroad;
- helping in the training of research investigators; and
- fostering communication of medical information.

"The [NIH](#) is one of eight health agencies of the Public Health Services which, in turn, is part of the [US Department of Health and Human Services](#). Comprising 27 separate components, mainly Institutes and Centers, [NIH](#) has 75 buildings on more than 300 acres in Bethesda, MD. From a total of about \$300 in 1887, the [NIH](#) budget has grown to more than \$20.3 billion in 2001."

- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Validations Overview](#)
- [Key References](#)



# KEY REFERENCES

- (1995) GAUSS Mathematical and Statistical System Version 3.2.18.
- (1988) National Center for Health Statistics: Vital Statistics of the United States, 1986, Vol. II Mortality Part A. in *Public Health Service, Washington, U.S. Government Printing Office*
- Albertsen, PC** (1998) Editorial: Computer modeling-What should we look for? in *Journal of Urology* 159
- Amling C** (2006) Personal communication
- Babaian, R.J., Kojima, M., et al** (1996) Comparative Analysis of Prostate Specific Antigen and its Indexes in the Detection of Prostate Cancer. in *Journal of Urology* 156 : 2 , p 432-437
- Barry MJ , Fleming C , Coley CM , Wasson JH , Fahs MC, Oesterling JE** (1995) Should Medicare Provide Reimbursement for Prostate-Specific Antigen Testing for Early Detection of Prostate Cancer? Part IV: Estimating the Risks and Benefits of an Early Detection Program in *Urology* 4 : 46 , p 445-461
- Boring CC, Squires TS, Tong T, Montgomery S** (1994) Cancer Statistics, 1994 in *CA: A Cancer Journal for Clinicians* 44 , p 7-26
- Brown ML, Riley GF, Potosky AL, Etzioni RD** (1999) Obtaining long-term disease-specific costs of care: application to Medicare enrollees diagnosed with colorectal cancer. in *In press, Medical Care*
- Carter HB, Morrell CH, Pearson JD, Brant LJ, et al.** (1992) Estimation of prostatic growth using serial Prostate-Specific Antigen measurements in men with and without prostate disease. in *Cancer Research* 52 , p 3323-3328
- Carter HB, Piantadosi S, Isaacs JT** (1990) Clinical evidence for and implications of the multistep development of prostate cancer. in *J Urol* 143 , p 742-746
- Catalona WJ, Smith DS, Ratliff TL, Basler JW** (1993) Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. in *JAMA* , p 948-954
- Clark LC, Combs GF, Turnbull BW, Slate EH, Chalker DK, Chow J, Davis LS, Glover RA, Graham GF, Gross EG, Krongrad A, Leshner JL, Park HK, Sanders BB, Smith, CL, Taylor JR, and The Nutritional Prevention of Cancer Study Group** (1996) Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: A randomized clinical trial. in *JAMA* , p 1957-1963
- Coley, CM;, Barry, MJ;, Fleming, C;, Fahs, MC;, Mulley, AG** (1997) Early Detection of Prostate Cancer Part II: Estimating the Risks, Benefits, and Costs in *Ann Intern Med* 126 : 6 , p 468-479
- Cowen ME, Chartrand M, Weitzel WF** (1994) A Markov Model of The Natural History of Prostate Cancer in *J Clin Epidemiol* 47 : 1 , p 3-21
- Crawford ED, DeAntoni EP, Etzioni R, Schaefer VC, Olson RM, Ross CA , and the Prostate Cancer Education Council** (1996) Serum Prostate-Specific Antigen and Digital Rectal Examination for the early detection of prostate cancer in a national community-based program. in *Urology* 47 , p 863-869
- Cronin KA, Slate EH, Turnbull BW, Wells MT** (1994) Using the Gibbs sampler to detect changepoints: application to PSA as a longitudinal marker for prostate cancer. in *Computing Science and Statistics* 26 , p 314-318
- Draisma G, Boer R, Otto SJ, van der Crujisen IW, Damhuis RA, Schroder FH, de Koning HJ,** (2003) Lead times and overdetection due to prostate-specific antigen screening: Estimates from the European Randomized Study of Screening for Prostate Cancer in *J Natl Cancer Inst* 95 : 12 , p 868-78

- Dugan, JA, Bostwick, DG, Myers, RP, Qian, J, Bergstrahl, EJ, Oesterling, JE** (1996) The definition and preoperative prediction of clinically insignificant prostate cancer in *JAMA* 275 : 4 , p 288-294
- Etzioni R, Cha R, Feuer EJ, Davidov O** (1998) Asymptomatic Incidence and Duration in Prostate Cancer. in *American Journal of Epidemiology* 148 , p 775-785
- Etzioni R, Legler JM, Feuer EJ, Merrill RM, Cronin KA, Hankey BF** (1999) Cancer Surveillance Series: Interpreting Trends in Prostate Cancer-Part III: Quantifying the Link Between Population Prostate-Specific Antigen Testing and Recent Declines in Prostate Cancer Mortality in *Journal of National Cancer Institute* 91 , p 1033-1039
- Etzioni R, Pepe M, Longton G, Hu C, Goodman G** (1999) Incorporating the time dimension in receiver operating characteristic curves: A prostate cancer case study. in *Medical Decision Making* 19 , p 242-251
- Etzioni R, Ramsey SD, Berry K, Brown M** (1999) The impact of including future medical care costs when estimating the costs attributable to a disease: a colorectal cancer case study. in *In press, Health Economics*
- Etzioni R, Urban N, Baker M** (1996) Estimating the costs attributable to a disease with application to ovarian cancer in *Journal of Clinical Epidemiology* 49 , p 95-103
- Etzioni RD, Kadane JB** (1995) Bayesian statistical methods in public health and medicine. in *Annual Review of Public Health* 16 , p 23-41
- Etzioni, Ruth, Cha, Raymond, Cowen, Mark E.** (1999) Serial Prostate Specific Antigen Screening For Prostate Cancer: A Computer Model Evaluates Competing Strategies in *J Urol* 162 : , p 741-748
- Farkas A, Schneider D Perrotti M, Cummings KB, Ward WS** (1998) National Trends in the Epidemiology of Prostate Cancer, 1973 to 1994: Evidence for the Effectiveness of Prostate-Specific Antigen Screening. in *Journal of Urology* 52 , p 444-449
- Feuer EJ, Merrill RM, Hankey BF** (1999) Cancer Surveillance Series: Interpreting Trends in Prostate Cancer-Part II: Cause of Death Misclassification and the Recent Rise and Fall in Prostate Cancer Mortality. in *Journal of the National Cancer Institute* 2 : 91 , p 1025-1032
- Feuer EJ, Wun LM** (1992) How Much of the Recent Rise in Breast Cancer Incidence Can be Explained by Increases in Mammography Utilization: A Dynamic Population Approach. in *American Journal of Epidemiology* 136 , p 1423-1436
- Feuer, EJ, Wun, LM, Boring, CC** (1993) Probability of developing cancer. in *Cancer Statistics Review: 1973-1989, National Cancer Institute* , p 1-8
- Fleming, Craig, Wasson, JH, Albertsen, PC, Barry, MJ, Wennberg, JE** (1993) A Decision Analysis of Alternative Treatment Strategies for Clinically Localized Prostate Cancer in *JAMA* 269 : 20 , p 2650-2658
- Gann PH, Hennekens CH, Stampfer MJ** (1995) A prospective evaluation of plasma Prostate-Specific Antigen for detection of prostate cancer in *JAMA* 273 , p 289-294
- Gilliland FD, Welsh DJ, Hoffman RM, Key CR** (1995) Rapid rise and subsequent decline in prostate cancer incidence rates for New Mexico, 1989-1993 in *Cancer Epidemiology, Biomarkers and Prevention* 7 : 4 , p 797-800
- Gohagen JK, Prorok PC, Kramer BS, Cornett JE** (1994) Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial of the National Cancer Institute. in *Journal of Urology* 152 , p 1905-1909

- Hankey BF, Feuer EJ, Clegg LX, Hayes RB, Legler JM, Prorok PC, Ries LA** (1999) Cancer Surveillance Series: Interpreting Trends in Prostate Cancer-Part I: Evidence of the Effects of Screening in Recent Prostate Cancer Incidence, Mortality, and Survival Rates. in *Journal of the National Cancer Institute* 2 : 1 , p 1017-1024
- Heston JR, Kelly JB, Meigs JW, et al** (1986) Forty-five years of cancer incidence in Connecticut 1935-1980. in *Natl Cancer Institute Monogr*
- Inoue LY, , Etzioni R, , Slate EH, , Morrell C, , Penson DF.** (2004) Combining longitudinal studies of PSA. in *Biostatistics* 5 : 3 , p 483-500
- Jacobsen SJ, Bergstralh EJ, Guess HA, Katusic SK, Klee GG, Oesterling JE, Lieber MM** (1996) Predictive Properties of Serum Prostate-Specific Antigen Testing in a Community-Based Setting. in *Archives of Internal Medicine* 156 , p 2462-2468
- Labrie F, Candas B, Dupont A, Cusan L, Gomez JL, Suburu RE, Diamond P, L'vesque J, Belanger A** (1998) Screening Decreases Prostate Cancer Death: First Analysis of the 1988 Quebec Prospective Randomized Controlled Trial in *The Prostate* 38 , p 83-91
- Laird NM, Ware JH** (1982) Random-effects models for longitudinal data. in *Biometrics* 38 , p 963-974
- Legler J, Feuer E, Potosky A, Merrill R, Kramer B** (1998) The role of Prostate-Specific Antigen testing patterns in the recent prostate cancer incidence decline in the USA in *Cancer Causes Control* 9 , p 519-527
- Lin DY, Feuer EJ, Etzioni R, Wax Y** (1997) Estimating medical costs from incomplete follow-up data. in *Biometrics* 53 , p 419-434
- Mariotto A, Etzioni R, Krapcho M, Feuer EJ** (2007) Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey in *Cancer* in press :
- Mccarthy** (1998) PSA screening said to reduce prostate-cancer deaths, or does it? in *The Lancet*
- Merrill RM, Feuer EJ, Warren JL, Schussler N, Stephenson RA,** (1999) Role of transurethral resection of the prostate in population-based prostate cancer incidence rates in *Am J Epidemiol* 150 : 8 , p 848-60
- Mettlin C, Murphy G, Babaian R, et al** (1996) The results of a five-year early prostate cancer detection intervention. in *Cancer* 1 : 77 , p 150-159
- Mettlin CJ, Murphy GP** (1998) Why is the prostate cancer death rate declining in the United States? in *Cancer* 2 : 82 , p 249-51
- Morrell CH, Pearson JD, Carter HB, Brant LJ** (1997) Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. in *Journal of the American Statistical Association* 90 , p 45-53
- National Hospital Discharge Survey Web Page** Hospital Discharge and Ambulatory Surgery Data (Latest update - July 21, 1999). <http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm>.
- Newman AJ Jr, Graham MA, Carlton CE Jr Lieman S** (1982) Incidental carcinoma of the prostate at the time of transurethral resection: importance of examining every chip in *Journal of urology* 128 , p 948-950
- Oesterling JE Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM** (1993) Serum prostate-specific antigen in a community-based population of healthy men. in *JAMA* 270 , p 860-864

- Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH, Barnhart S, Hammar S** (1996) Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. in *New England Journal of Medicine* 18 : 334 , p 1150-5
- Pearson JD, Morrell CH, Landis PK, Carter HB, Brant LJ** (1994) Mixed-effects regression models for studying the natural history of prostate disease. in *Statistics in Medicine* 13 , p 587-601
- Pienta K, Goodson J A, Esper S** (1996) Epidemiology of prostate cancer: molecular and environmental clues in *Urology* 5 : 48 , p 676-683
- Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK** (2005) Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial in *J Urol* 173 : 3 , p 746-50
- Potosky AL, Kessler L, Gridley G, Brown CC, Horm JW** (1990) Rise in prostatic cancer incidence associated with increased use of trans-urethral resection. in *Journal of National Cancer Institute* 82 , p 1624-1628
- Potosky AL, Miller BS, Kramer BS, Albertsen PC** (1995) The role of increasing detection in the rising incidence of prostate cancer. in *JAMA* 273 , p 548-552
- Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG** (1993) Potential for cancer related health services research using a linked Medicare-tumor registry database in *Medical Care* 8 : 31 , p 732-748
- Presti JCJ, Chang JJ, Bhargava V, Shinohara K** (2000) The optimal systematic prostate biopsy scheme should include 8 rather than 6 biopsies: Results of a prospective clinical trial in *J Urol* 163 : 1 , p 163-6
- Richie JP, et al** (1993) Effect of patient age on early detection of prostate cancer with serum Prostate-Specific Antigen and Digital Rectal Examination. in *Urology* 42 , p 365-374
- Roberts RO, Bergstralh EJ, Katusic SK, Lieber MM, Jacobsen SJ** (1999) Decline in prostate cancer mortality from 1980 to 1997, and an update on incidence trends in Olmsted County, Minnesota. in *Journal of Urology* 161 : 2 , p 529-33
- Slate EH, Clark LC** (1999) Using PSA to detect prostate cancer onset: An application of Bayesian retrospective and prospective changepoint identification. In: *Case Studies in Bayesian Statistics IV.* , p 511-534
- Smith Ds, Catalona WJ** (1994) The nature of prostate cancer detected through prostate-specific antigen based screening. in *Journal of Urology* 152 , p 1732-1736
- Stanford JL, Wicklund KG, McKnight B, Daling JR, Brawer MK** (1999) Vasectomy and Risk of Prostate Cancer in *Cancer Epidemiology* 8 , p 881-886
- Stanford JL, Stephenson RA, Coyle LM, Cerhan J, Correa R, Eley JW, Gilliland F, Hankey B, Kolonel LN, Kosary C, Ross R, Severson R, West D** (1999) Prostate Cancer Trends 1973-1995, Seer Program. in *National Cancer Institute*
- Telesca D, Etzioni R, Gulati R** (2007) Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends in *Biometrics* in press :
- Tsodikov A, Szabo A, Wegelin J** (2006) A population model of prostate cancer incidence in *Stat Med* 25 : 16 , p 2846-66
- Urban N, Drescher C, Etzioni R, Colby C** (1997) Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening in *Controlled Clinical Trials* 18 , p 251-270
- Weiss NS, Rossing MA** (1996) Healthy screenee bias in epidemiologic studies of cancer incidence in *Epidemiology* 3 : 7 , p 319-22



- Whitmore WF** (1988) Background for screening: natural history and treatment. **EORTC Genitourinary Group Monograph 5: Progress and Controversies in Oncological Urology II.** , p 123-130
- Whitmore, WF** (1990) Natural History of low-stage prostatic cancer and the impact of early detection. in *Urol Clin N Am* 17 , p 689-697
- Whittemore AS, Lele C, Friedman GD, Stamey T, Vogelmann JH, Orentreich N** (1995) Prostate-specific antigen as predictor of prostate cancer in black men and white men. in *Journal of the National Cancer Institute* 87 , p 354-360
- Whittemore, AS, Keller, JB, Betensky, R** (1991) Low grade latent prostate cancer volume: predictor of clinical cancer incidence? in *Journal of the National Cancer Institute* 83 , p 1231-1235
- Wun LM, Merrill RM, Feuer EJ** (1998) Estimating lifetime and age-conditional probabilities of developing cancer in *Lifetime Data Analysis* 4 , p 169-186

---

[Core Docs](#)



# UNIVERSITY OF MICHIGAN

FLEXKB DOCUMENT  
Version: HI.001.12212009.86043  
Document generated: 12/21/2009

**Important note:** This document will be updated periodically. The most current version is available at <http://cisnet.cancer.gov/profiles>. Note that unlike most PDF documents, the CISNET model profiles are not suitable for printing as they are not typically written or read in sequential fashion.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

We recommend you let your interests guide you through this document, using the navigation tree as a general guide to the content available.

The intent of this document is to provide the interested reader with insight into ongoing research. Model parameters, structure, and results contained herein should be considered representative but preliminary in nature.

We encourage interested readers to contact the contributors for further information.

Go directly to the: [Reader's Guide](#).



# READERS 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.



Readers Guide  
Model Overview  
Assumption Overview  
Parameter Overview  
Component Overview  
Output Overview  
Results Overview  
Key References

### 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.

## Further Reading

---

These topics will provide an intermediate level view of the model. Consider these documents if you are interested in gaining a working knowledge of the model, its inputs and outputs.

## Advanced Reading

---

These topics denote more detailed documentation about specific and important aspects of the model structure



# MODEL PURPOSE

## SUMMARY

This page summarizes the purposes for which this model was developed.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

## PURPOSE

The model provides a quantitative link between dissemination of cancer control processes and their impact on population and public health measures of cancer incidence, survival and mortality. Its purpose is to unravel the myriad causes and relationships that underlie recent trends in prostate cancer incidence and mortality, to quantify the relationships in terms of model parameters, and to enable researchers to perform inference on these parameters by means of confidence intervals and hypothesis tests.

The model provides tools by which national population and cancer registry data may be analyzed, so that the population impact of cancer control processes may be understood and predicted. It exists that researchers might predict short- and long-term trends in national incidence and mortality under various scenarios; might analyze racial disparities as they pertain to factors associated with trends in treatment, survival, incidence and mortality; and might determine and evaluate optimal screening strategies.

A particular goal of this model is to enable researchers to determine in what way, if at all, PSA screening of asymptomatic men is linked to the recent decline in prostate cancer mortality. In fact, the model has already generated predictions for prostate cancer incidence and mortality under current PSA utilization patterns, and for the baseline case of no PSA screening. The latter prediction is counterfactual, in that it expresses what incidence and mortality would have been during the years 1970-2000 if there had been no PSA screening, other things being equal. Thus it yields an estimate of the differences in incidence and mortality that are purely associated with PSA utilization.



# MODEL OVERVIEW

## SUMMARY

This document provides an overview of the modeling effort, and describes the model itself in general terms.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

## PURPOSE

This is a model of prostate cancer incidence and mortality. It was developed to analyze national population and cancer registry data. It is used to understand, predict and optimize the population impact of cancer control processes in prostate cancer. See [Model Purpose](#) for more details.

## BACKGROUND

Excluding skin cancers, prostate cancer is the most common cancer in American men. It claims over 40,000 lives annually, ten percent of cancer deaths among men, and is second only to lung cancer as a cause of cancer deaths<sup>1</sup>. Progressive prostate cancer is a serious disease. Thousands of men suffer pain and complications and die prematurely from progressing tumors.

Management and control of prostate cancer is a significant public health problem. For more than a decade since the introduction of PSA testing in the late 80s, the incidence rates of newly diagnosed prostate cancers have seen a dramatic increase to over 190,000 cases in early 90s, followed by an equally dramatic decline<sup>2</sup> ([Figure\\_1](#)). At the same time, mortality slowly increased from the 70s to the early 90s, and has been declining since then<sup>3</sup> PSA screening has spread through the population because of the hope that it ultimately may reduce mortality. But the mere fact that screens can detect organ-confined prostate cancer does not in itself constitute a sufficient ground for their implementation. Screening cannot be justified unless patients who are screened actually have improved outcomes, and this has not yet been shown.

To make appropriate decisions regarding treatment and public health policy, we must understand the causes of these trends. To do so, models are needed that can unravel and disentangle all the factors behind the observed population trends, including length bias, overdiagnosis, early detection of cases that would become clinical, shifts in stage and grade of cancer associated with early detection, and other possible factors, such as a change in the survival curve within stage and grade following the advent of PSA screening.

The model reported here has several distinctive features that meet this challenge. These include:

- Estimation from population data instead of from screening trial data.
- A flexible regression framework, accomodating explicit adjustments for differences in screening and treatment utilization patterns.
- Analytic, rather than simulation-based, procedures for estimation and prediction.
- The identification of within-stage shift as a factor affecting survival and mortality.



- Confidence intervals and tests of statistical hypotheses for all model [parameters](#).
- A general structure, making the model applicable to cancer sites other than the prostate.

Before describing the "guts" of the model, we briefly discuss these features.

#### **Estimation from population data.**

The [parameters](#) in our model are estimated from population databases such as SEER, not from screening trial data. This mode of estimation is possible because we incorporate random PSA schedules into the estimation procedure. This approach is appropriate, since the focus of CISNET is on population trends in incidence and mortality. To measure such trends, a model must unravel and disentangle a set of competing risks and confounded effects, including length bias, overdiagnosis, advancement of diagnosis due to screening, stage shift due to early detection, and other possible effects, such as within-stage shift. No randomized trial for which data are currently available has been designed to measure all these effects. To model such effects, it is critical that methods be developed that can exploit population data for estimation. Our model does this.

Since our estimation procedure is based on population databases, we are able to exploit the wealth of information available from these databases. The consequences for precision and power are substantial, since population databases are typically much larger than the number of individuals participating in a screening trial. Increased power enables us to obtain reliable estimates on a potentially larger set of model [parameters](#) than would be possible were the estimation based on screening trials.

#### **A flexible regression framework.**

The model is constructed within a flexible framework that uses concepts of survival analysis, yet is not limited to the standard Cox proportional hazard model. This framework accommodates adjustments for variable screening and treatment utilization patterns through built-in lead-time, length-bias and stage- and within-stage shift, all of which may affect survival. This framework allows the researcher to derive realistic estimates of mortality through joint modeling of incidence and survival in a dynamic population environment. For example, clinical covariates being equal, the model generally would provide different survival estimates for subjects from low and high PSA utilization areas.

Since this regression framework is applied to population data, there is no need to perform additional calibration of model predictions after [parameter](#) estimation. All adjustments of the model are explicitly built into the fitting procedure. In case the fit is unsatisfactory, this framework naturally leads the researcher or policymaker to directly examine model assumptions and [parameter](#) values. Thus the framework is flexible, allowing reviewers, researchers outside the development team, and policymakers to analyze and evaluate the model, creating an environment conducive to further model improvement.

#### **Analytic procedures for estimation and prediction.**

A second distinctive feature of our model is that it is analytic in all its components. We do not use simulation or stochastic approximation either to fit the model or to predict



from it. This fact has two consequences, one pertaining to ease of model fitting and prediction, one pertaining to model interpretation.

A fit or prediction run for our model takes a few minutes on a standard PC, much less time than if we relied on simulation. This results in a quick feedback and consequently allows interactive dialog between the model and the user. This feature is particularly important for providing a policy maker with a tool to quickly evaluate a number of cancer control hypotheses in an interactive environment. We have already developed a working prototype software package (SCANS, Self-Consistency Analysis of Surveillance) for Windows.

In addition to the practical advantage of speed, the analytic nature of our model means that it is transparent. The [parameters](#) can be directly interpreted in terms of processes of interest, so that the model itself is in no way a "black box."

#### **Within-stage shift.**

It is customary to explain survival and mortality differences associated with PSA screening by shifts in stage and grade of the cancer associated with the lead time between screen diagnosis and the time at which an individual would have been diagnosed clinically. Under the current model, however, we find evidence of a survival shift within stage, associated with PSA dissemination. For more information, see [Within Stage Shift](#).

#### **Additional distinctive features of this model.**

We provide confidence intervals and tests of statistical hypotheses for all estimated model [parameters](#). Thus, the user has an idea as to the significance of model findings.

While the model is applied to prostate cancer, its structure is general and open to immediate application to other diseases.

## MODEL DESCRIPTION

The model provides a means by which [parameters](#) may be estimated that enable us to explain and predict trends in incidence, survival and mortality. As suggested in the Background section above, the model is probabilistic -- yielding p-values and confidence intervals -- and is accompanied by procedures that permit estimation from the same kind of data that we are seeking to predict and explain, namely, large population databases.

#### **Model Assumptions**

Please see [Assumption Overview](#) for the assumptions on which the model is based.

#### **Model Inputs**

Please see [Parameter Overview](#) for a list of model inputs.

#### **Model Outputs**

The model yields estimates of a set of [parameters](#) that together constitute a comprehensive model for prostate cancer incidence, survival, and mortality.



These [parameters](#) govern such basic characteristics of the model as age at tumor onset, sojourn time, lead time, overdiagnosis, delay time, sensitivity of the PSA test, and the correlation between the age at tumor onset and the sojourn time. From these basic characteristics follow estimates of incidence a function of calendar year, age, stage, and grade; survival as a function of stage, grade, and delay time; and mortality by calendar year and age. From these [parameters](#) we finally derive estimates of the effect of PSA screening on prostate cancer incidence and mortality.

For more information, see [Output Overview](#).

### **Model Limitations**

Since the data are observational, we do not have the benefit of complete elimination of confounders, as is possible in a well-conducted clinical trial. The urgent need to gain understanding of the processes, however, and the long follow-up time required by a screening trial, do not afford us the luxury of waiting for experimental results.

Because the model is based on past data, its generalizability to the future may be limited.

The current version of the model does not explicitly describe PSA growth.

## **CONTRIBUTORS**

We gratefully acknowledge the collaboration of the following individuals.

Dr. Ray Merrill of the Department of Health Sciences, Brigham Young University, helped in model building, interpretation and prediction of national prostate cancer trends.

Dr. Marco Zaider, Head of Brachytherapy at the Memorial Sloan-Kettering Cancer Center, brought his expertise in prostate cancer treatment to the project. He assisted us in analyzing and interpreting clinical data, providing a link between screening strategies and prostate cancer post-treatment survival.

Dr. Gilda Garibotti worked on computer implementation of the profile information matrix methodology in the survival analysis module. She provided advice on software development and implementation of survival analysis machinery in the population model.

Dr. Aniko Szabo provided help on software development and implementation of methods and population models. She especially provided advice on the integration of the model software into the population software shell, and helped in testing computer code that implements extended population models.

## **REFERENCES:**

- <sup>1</sup> Stewart, S.L., King, J.B., Thompson, T.D., Friedman, C., Wingo, P.A. "Cancer mortality surveillance--United States, 1990-2000" in *MMWR Surveillance Summary* 2004; 53: 3: 1-108
- <sup>2</sup> Potosky, A.L., Feuer, E.J., Levin, D.L. "Impact of screening on incidence and mortality of prostate cancer in the United States" in *Epidemiological Review* 2001; 23: 1: 181-186



University of Michigan  
Model Overview  
References:

<sup>3</sup> Hsing, A.W., Devesa, S.S. "Trends and patterns of prostate cancer: What do they suggest?" in *Epidemiologic Reviews* 2001; 23: : 3-13



# ASSUMPTION OVERVIEW

## SUMMARY

In this section we summarize the main assumptions on which the model is based.

## BACKGROUND

Researchers generally agree that prostate cancer is the result of an irreversible transition of the disease through three consecutive stages: the disease free stage, the pre-clinical stage and the clinical stage. This three-stage model entails the following potential time points in an individual's life: birth, [onset](#) of prostate cancer, time of clinical diagnosis, time of death due to prostate cancer, and alternatively the time of death due to a competing risk.

Although this model may be accurate as far as it goes, it does not capture the processes currently affecting incidence, survival and mortality. The dissemination of [TURP](#) and, more drastically, the dissemination of PSA testing have made the picture more complicated, because both [TURP](#) and the PSA test can advance the diagnosis of prostate cancer.

In addition, it has customarily been believed that stage shift is the only reasonable explanation for any benefit derived from early detection. We do not make this assumption, and in fact have found evidence to the contrary (see [Introduce Within Stage Shift](#)).

For both these reasons, a more complex set of assumptions must be spelled out.

## ASSUMPTION LISTING

**Tumor onset** (See [Age At Tumor Onset](#) for details.)

- The baseline hazard of tumor onset may depend on age.
- The hazard of tumor onset may depend on calendar year.
- The effect of calendar year on the hazard of tumor onset is multiplicative.

**Sojourn Time** (See [Sojourn Time Distribution](#) for details.)

- Baseline [Sojourn Time](#) may depend on age.
- The hazard function associated with baseline [Sojourn Time](#) may include a multiplicative trend in calendar time.
- [Sojourn Time](#) may depend on age at tumor onset.
- Given the time of tumor onset, [Sojourn Time](#) does not depend on the cancer screening process.

**Delay Time.** The distribution of the duration of the [latent\\_disease\\_stage](#) is an average over random patterns of all possible modes by which the disease can be detected. (See [Incidence Model](#) for details.) These include

- Clinical diagnosis through symptoms





- PSA screening
- Transurethral resection of the prostate ([TURP](#)).

**Screening tests** occur randomly in time, subject to the following assumptions (see [PSAscreening Model](#) for details):

- Age at first PSA test has a distribution (alternatively, hazard function) that depends both on age and calendar time.
- The times between consecutive PSA tests occur as a non-homogeneous Poisson process, with an intensity that depends on age and calendar year.
- The sensitivity of the screening test is an increasing function of time since tumor onset.

**Survival time after diagnosis** follows a semiparametric regression model. (See [Survival Component](#) for details.)

**Mortality** in the population can be adequately modeled by combining information from the incidence and survival models. (See [Mortality Model](#) for details.)

**Within-Stage Differences in Prognosis.** We allow the possibility that stage and grade at diagnosis are not the only variables associated with a patient's prognosis for survival. (See [Introduce Within Stage Shift](#).)



# PARAMETER OVERVIEW

## SUMMARY

This document lists and defines the inputs to the modeling algorithm.

## BACKGROUND

The parameters discussed on this page are parameters of a **model**, not of a population. Thus, in this context, the term "parameter" has an entirely different meaning from the classical statistical use of the term.

In classical statistics, a parameter is a number or set of numbers that characterize a population. Typically, we estimate parameters by drawing a sample from the population, measuring a variable or set of variables on each individual in that sample, and using these measurements (i.e., using a set of data) as input to an estimation algorithm. The estimation algorithm produces parameter estimates as output. These may be point estimates, interval estimates, p-values, or higher-dimensional objects such as densities or cumulative distribution functions.

In the current context, however, "parameter" does not refer to a characteristic of the population but rather to any input to the modeling algorithm. Thus, even a set of measurements made on each individual in a sample drawn from the population are considered "parameters" if they are used as input to the modeling algorithm.

## PARAMETER LISTING OVERVIEW

The inputs to the model (also called the "model parameters," as explained in the Background section of this page) consist of population data as well as "given" parameters and distributions provided by the National Cancer Institute (NCI). These include:

- The distribution of PSA utilization in the population. This distribution is based on an algorithm that can be used to simulate life histories of the times that individual men undergo PSA tests. For more information see [base Case PSA](#).
- Surveillance, Epidemiology and End Results (SEER) data on every individual diagnosed with prostate cancer in nine areas of the United States (San Francisco-Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta), more than 350,000 cases ([SEER\\_Medicare](#)). The data include tumor characteristics as well as standard follow-up and outcome variables. In particular, for each age  $a$  (over 50) and each year  $t$  the number of new prostate cancer cases  $C(a, t)$  is derived and "fed" to the model. For details see [Likelihood In The Incidence Model](#).
- Population count files belonging to the same areas from which the prostate cancer case data were obtained. From this source the number of people at risk for prostate cancer for each age  $a$  and each year  $t$  is derived,  $P(a, t)$ , and "fed" into the model. For details see [Likelihood In The Incidence Model](#).
- Age distribution in the U.S. population in the year 2000 for men over 50.
- Risk of death from other causes, derived from the [Human Mortality Database](#)<sup>1</sup>.





University of Michigan  
Parameter Overview  
References:

## REFERENCES:

- <sup>1</sup> Wilmoth, John R. (Director), Shkolnikov, Vladimir Shkolnikov (Co-Director),  
"Human Mortality Database (HMD)." 2003;



# COMPONENT OVERVIEW

## SUMMARY

This document outlines the analytic components of which the model is constructed.

## OVERVIEW

The model is composed of three distinct components by which predictions and estimates are made based on population data.

The [Incidence Component](#) takes population data as input and yields estimates and predictions of prostate cancer incidence by calendar year and age. In addition it yields predictions both in the presence and in the absence of PSA testing, thereby yielding an estimate of the difference in prostate cancer incidence that is due to the presence of PSA testing.

The [Survival Component](#) also takes population data as input. It yields a model for the relationship between a set of covariates (including age, year of diagnosis, cancer stage and tumor grade) and a man's survival prognosis.

The [Mortality Component](#) combines the [Incidence Component](#) and the [Survival Component](#). It yields estimates and predictions of prostate cancer mortality by calendar year, age, and presence or absence of PSA testing. Thus, similar to the [Incidence Component](#), the [Mortality Component](#) yields an estimate of the difference in prostate cancer mortality that is due to the presence of PSA testing.

## COMPONENT LISTING

[Incidence Component](#)

[Survival Component](#)

[Mortality Component](#)



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



# OUTPUT OVERVIEW

## SUMMARY

This page lists and describes the statistics computed by the model, the [parameters](#) for which these statistics serve as estimates, and the predictions available from the model.

## OVERVIEW

The model yields estimates of a set of [parameters](#) (some one-dimensional, some multi- or high-dimensional) that together constitute a comprehensive model for prostate cancer incidence, survival, and mortality.

## OUTPUT LISTING

- Prostate cancer incidence as a function of calendar year, age, and presence or absence of PSA testing. (See [Incidence Figure](#).)
- Survival as a function of calendar year, age, stage, grade, and screening schedule. (See [Survival Component](#).)
- Mortality by calendar year, age, and presence or absence of PSA testing. (See [Mortality Component](#).)
- Mean lead time as a function of birth cohort. (See [define Lead Time](#) and [Results Overdiagnosis Lead Time](#).)
- Overdiagnosis as a function of birth cohort. (See [define Overdiagnosis](#) and [Results Overdiagnosis Lead Time](#).)
- Delay time as a function of calendar year and age, with and without PSA testing. (The latter is a counterfactual scenario; see [define Delay Time](#) and [Mean Posterior Delay Time Marginal Incidence Model](#).)
- Relationship between delay time and survival. (See [Within Stage Shift By Delay Time](#).) Note that this estimate entails differences in survival (or prognosis) associated with differences in screening schedule, even after adjustment has been made for stage and grade of cancer. We call this phenomenon within-stage shift. To our knowledge this phenomenon has not been discovered or quantified by any other research group.
- Estimates of the differences in incidence and mortality that are due to PSA screening. These estimates are based on a scenario in which incidence and mortality are estimated in the counterfactual case of no PSA screening. (See [results Age Adjusted Incidence Mortality](#).)
- Parameters governing the distribution of baseline sojourn time. (See [Incidence Model, Analysis Of Population Data](#) and [Table\\_1](#).)
- Parameters governing age at tumor onset. (See [Incidence Model, Analysis Of Population Data](#), and [Table\\_1](#).)
- Sensitivity of the PSA test. This was estimated at 100% by the model, but can be set to zero to predict what incidence would be in the absence of PSA. (See [Modeling Cancer Detection Through Screening](#) and [Analysis Of Population Data](#).)
- Correlation between age at onset and sojourn time. (This was found to be negligible, and removed from the model; see [Analysis Of Population Data](#).)





# RESULTS OVERVIEW

## SUMMARY

This page lists discoveries that have been made through the current modeling effort.

## OVERVIEW

The model constitutes a framework for analysis of population databases. Within this framework, and by means of its estimation procedures, researchers obtain estimates of parameters that matter in the quest to understand the causes and processes of change in prostate cancer incidence, survival, and mortality. These parameters represent factors that either cause, or are associated with, differences in the outcomes that researchers and the public ultimately care about: survival prognosis for those diagnosed with prostate cancer, and mortality in the population due to prostate cancer.

## RESULTS LIST

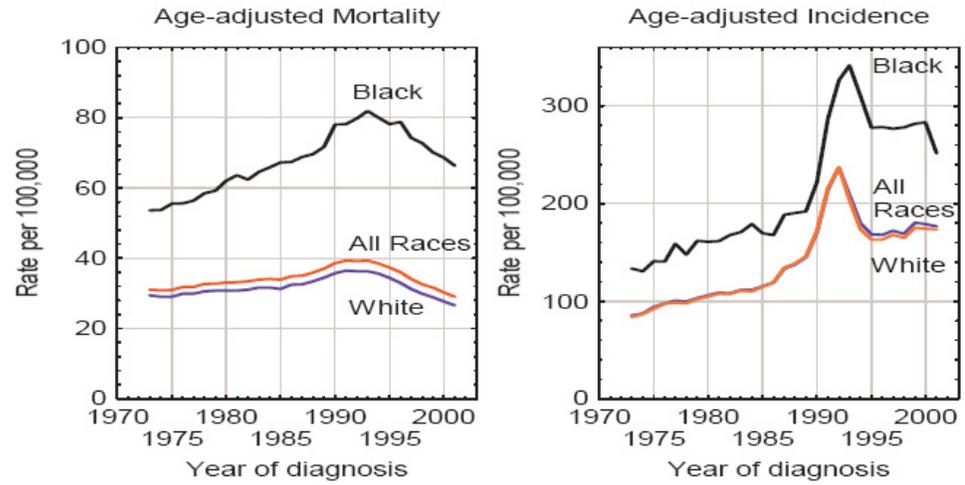
- [resultsAgeAdjustedIncidenceMortality](#), and consequently
- Estimation of the difference in incidence that is due to PSA testing.
- [resultsAgeAdjustedIncidenceMortality](#), and consequently
- Estimation of the difference in mortality that is due to PSA testing.  
For more information on results regarding incidence and mortality, see [results Age Adjusted Incidence Mortality](#).
- Estimation of [Delay Time](#) without using survival data. This represents an independent significant prognostic factor for post-treatment survival, particularly for cancers in the localized-regional stage. See [Mean Posterior Delay Time Marginal Incidence Model](#).
- Identification of differences in survival associated with differences in [Delay Time](#). The changes in survival associated with early detection have been customarily modeled by stage shift. We have found, however, that, even after stage has been accounted for, differences in delay time (the time from onset of cancer until detection) are associated with differences in survival. See [Within Stage Shift By Delay Time](#).
- [Publications](#) accepted or submitted (click on the link for a list).





- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

# FIGURE\_1



Figure\_1. Prostate cancer incidence and mortality rates by year of diagnosis age-adjusted to US population in year 2000. Data from the Surveillance, Epidemiology and End Results (SEER) database, National Cancer Institute.



University of Michigan  
parameter



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

# PARAMETER

In this model profile, the term "parameter" is used in the classical statistical sense. In this sense a parameter is a number or set of numbers that characterize a population. A primary object of classical statistics is to use data, also referred to as a sample, to estimate the parameters of a population. This sense of the term is entirely different from the sense referred to by the phrase [Parameter Overview](#). Please see the Background section of that document for an explanation of the difference between the two senses of the word.

For information on the particular parameters estimated by the model, see [Output Overview](#).



University of Michigan  
Within Stage Shift



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

## WITHIN STAGE SHIFT

The effects of [over diagnosis](#), [Length Bias](#) and [Lead Time](#) result in remarkable changes in the meaning of clinical covariates at diagnosis. With the introduction of screening, however, the prognostic value of such covariates is modified. The prognosis for cases diagnosed in the screening era is markedly different as compared to cancer cases from unscreened populations. This effect remains unexplained even when survival is adjusted for stage and grade ([Figure\\_2](#)). As PSA screening is intensified with the dissemination of the test in the U.S. population, survival in localized stage is improving while survival in distant stage is worsening as discussed above. A similar effect in the localized stage might be found before PSA was introduced, in association with early detection through TURPs. In contrast to the situation in a clinical trial, straightforward conditioning on clinical covariates in the analysis of population data may be misleading, and special care is needed to adjust for screening patterns in the population. For the same reason, the results of clinical trials are not immediately generalizable to the population setting.



University of Michigan  
define Onset

# DEFINE ONSET

**Onset** is defined as the beginning of prostate cancer in an individual. Note that the time of onset for an individual cannot be directly determined.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
define TURP

# DEFINE TURP

Transurethral resection of the prostate (TURP) is a surgical procedure performed to treat benign prostate hyperplasia (BPH) and urinary obstruction symptoms.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
Introduce Within Stage Shift

# INTRODUCE WITHIN STAGE SHIFT

See [Delay\\_Time\\_Approach](#) for modeling of within-stage shift.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



# AGE AT TUMOR ONSET

## Age at tumor onset

We use a Weibull distribution for the baseline age at tumor onset. Its baseline hazard function is given by

$$h_O(y) = s_O \left( \frac{\Gamma(1 + 1/s_O)}{\mu_O} \right)^{s_O} y^{s_O-1},$$

where  $y$  is the age past 50. In the above expression Weibull distribution is parameterized through the mean  $\mu_O$  and the shape parameter  $s_O$  related to the coefficient of variation

$$\sqrt{\frac{\Gamma(1 + \frac{2}{s_O})}{\Gamma^2(1 + \frac{1}{s_O})} - 1}.$$

Included in the model is a trend function  $T_O(t)$  that depends on calendar time. This function exerts a multiplicative effect on the baseline hazard so that the hazard of tumor onset depends on age and birth cohort

$$\lambda_O(y|x) = h_O(y)T_O(x + y).$$

The trend is used to model possible changes in the pattern of the disease onset with calendar time due to unspecified factors such as changes in diet, environment and biology of the disease. Note that it is hardly possible to give a biological definition for the tumor onset. From the modelling prospective, tumor onset represents the earliest point in time where cancer could be detected by screening. For this reason changes in detection technology, practice of biopsies for the disease following a positive screens and other diagnostics management issues may also affect the definition. Changes in such practices that are not modelled in a mechanistic fashion are thought of as part of the trend function. We used truncated linear trend functions in data analysis.

(This is extracted from an early draft of<sup>1</sup>.)

## REFERENCES:

- <sup>1</sup> Tsodikov, A., Szabo, A., Wegelin, J. "A population model of prostate cancer incidence" in *Statistics in Medicine* 2006; 25: 16: 2846-2866



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References



University of Michigan  
define Sojourn Time

# DEFINE SOJOURN TIME

**Sojourn time** is defined as the potential (other risks removed) time from tumor [onset](#) to its clinical diagnosis. Thus it is the duration of the preclinical stage in the absence of screening. We speak of the sojourn time distribution even for individuals who receive screenings. In this way we model the competing risks of clinical and screening diagnosis.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



# SOJOURN TIME DISTRIBUTION

## The sojourn time distribution.

Sojourn time is defined as the potential (other risks removed) time from tumor onset to its clinical diagnosis. A Weibull distribution with mean  $\mu_{CDx}$  and shape parameter  $s_{CDx}$  is used to model the baseline sojourn time hazard. Two effects can be imposed on the baseline sojourn time distribution:

- Age. Sojourn time may be affected by age for various reasons. Tumor growth biology may depend on the age of the person. Also, tumors developing at a younger age may represent a special subtype that can have different progression characteristics. To model age dependency, the mean sojourn time is regressed on the age at tumor onset  $y$  as  $\mu_{CDx} \exp(-\beta_{CDx}y)$ , where the parameter  $\beta_{CDx}$  models the correlation between the sojourn time and the onset time.
- Secular trend. Sojourn time may be affected by changes in the practice of cancer detection other than the studied modality of screening. Most notably, before PSA was introduced, prostate cancer was often detected as a result of surgery (Transurethral Resection of the Prostate, **TURP**) for benign prostate disorders<sup>1</sup>. Other changes in prostate cancer awareness in the population and detection practices may have contributed to a trend of increasing incidence observed before PSA was introduced. These trends in calendar time are modelled using a multiplicative trend function  $T_{CDx}(t)$  acting on the baseline sojourn time hazard.

We have the sojourn time hazard in the form

$$\lambda_{CDx}(\xi|x, y) = h_{CDx}(\xi|y)T_{CDx}(x + y + \xi),$$

where  $x$  is the birth year,  $y$  is age (past 50) at tumor onset,  $\xi$  is time since tumor onset, and  $h_{CDx}(\xi|y)$  is Weibull hazard with shape parameter  $s_{CDx}$  and mean  $\mu_{CDx} \exp(-\beta_{CDx}y)$ .

## REFERENCES:

- <sup>1</sup> Merrill, R.M., Feuer, E. J., Warren, J. L. , Schussler, N. , Stephenson, R. A. "Role of transurethral resection of the prostate in population-based prostate cancer incidence rates" in American Journal of Epidemiology 1999; 150: 8: 848-860



University of Michigan  
define Delay Time

# DEFINE DELAY TIME

**Delay time** is defined as the duration of the [latent\\_disease\\_stage](#), i.e., the time from onset until detection of cancer by any means, including PSA screening or clinical detection.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
define Latent Disease Stage

# DEFINE LATENT DISEASE STAGE

The **latent disease stage** is defined as the time when an individual has cancer but the cancer has not yet been detected by any means.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



# INCIDENCE MODEL

## INCIDENCE MODEL

We use the classical three-stage model of the natural history of a chronic disease<sup>1</sup>. Prostate cancer is a result of an irreversible transition of the disease through three consecutive stages: disease free stage, pre-clinical stage and clinical stage. The time spent in disease-free stage is characterized by the age  $Y$  (a random variable) at onset of the disease. In the pre-clinical stage disease is asymptomatic and can be detected by a screening test. The duration of the preclinical stage in the absence of screening (a random variable) is termed the sojourn time. If undetected by screening, the disease can either reach the clinical stage or, alternatively, the event of clinical diagnosis is precluded by a competing risk other than the disease of interest.

The distribution of any random duration can be specified by one of the following functions: a hazard function (h.f.,  $\lambda$  or  $h$ ), a survival function (s.f.,  $G$ ), a distribution function (d.f.,  $\overline{G}$ ), or a probability density function (p.d.f.,  $f$ ). Dependent on the situation, we will use the most convenient representation. Denote age by  $a$ , calendar year by  $t$ , year of birth by  $x$ , and time since tumor onset by  $\xi$ . We will follow the above notation unless noted otherwise. Prostate cancer incidence  $\lambda_I(a, t)$  by age and year can be written as  $\lambda_I(a|t - a)$  where  $\lambda_I(a|x)$  is the h.f. for cancer diagnosis for the  $x$ -birth cohort. Clearly,

$$\lambda_I(a|x) = \frac{f_I(a|x)}{G_I(a|x)}.$$

The functions  $f_I$  and  $G_I$  are in fact represented by a fairly complex mixture model. It is clear that cancer incidence is a convolution of two generally dependent survival times: age at tumor onset  $Y$  and duration of the latent disease stage  $T$ .

$$f_I(a|x) = \int_0^a f_I(a - y|x, y) f_O(y|x) dy,$$

where  $f_I(\xi|x, y)$  is a conditional p.d.f. of  $T$ , and  $f_O$  is the p.d.f. of  $Y$ . Generally,  $f_I(\xi|x, y)$  is an average over random patterns of screening operating in the population. It is clear that  $T$  is a result of two dependent competing risks: the one associated with natural clinical diagnosis through symptoms and the one associated with detection through screening. Dependency between the two risks is a consequence of natural detection and screen-based detection risks sharing the same disease development process in the subject. This dependency is modelled through the concept of shared mixed effect (frailty) Hougaard<sup>2</sup>, represented by  $Y$ . Conditional independence of potential risks of natural and screen-based detection, given  $Y$  gives

$$G_I(\xi|x, y) = G_{CDx}(\xi|x, y) G_{SDx}(\xi|x, y),$$



where  $G_{CDx}$  is the s.f. of time to clinical diagnosis (CDx) in the absence of screening (the sojourn time), and  $G_{SDx}$  is the s.f. of the potential time to screen-based diagnosis (SDx). Note, that  $G_{SDx}$  in our model corresponds to a continuous distribution as it is represented as a continuous mixture over random screening schedules in the population. Since incidence of prostate cancer before the age of 50 is negligible, we will associate the birth year  $x$  with the year in which the man turns 50. Weibull distribution with mean  $\mu_0$  and the shape parameter  $s_0$  is used for the baseline age at tumor onset. Weibull distribution with mean  $\mu_{CDx}$  and shape parameter  $s_{CDx}$  is used to model the baseline sojourn time hazard  $h_{CDx}$ . Two effects are imposed on the baseline sojourn time distribution, age dependence and a secular trend. To model age dependency, the mean sojourn time is regressed on the age at tumor onset  $y$  as  $\mu_{CDx} \exp(-\beta_{CDx}y)$ , where the parameter  $\beta_{CDx}$  models correlation between the sojourn time and the onset time. Secular trend models alterations in the practice of cancer detection other than the studied modality of screening. Most notably, in the pre-PSA era, many prostate cancers were incidentally detected through TURP

<sup>3</sup>. Secular trend is introduced as a multiplicative effect

## REFERENCES:

- <sup>1</sup> Zelen, M., Feinleib, M. "On the theory of screening for chronic diseases" in *Biometrika* 1969; 56: : 601-614
- <sup>2</sup> Hougaard, P. "Frailty models for survival data" in *Lifetime Data Analysis* 1996; 1: : 255-274
- <sup>3</sup> Merrill, R.M., Feuer, E. J., Warren, J. L. , Schussler, N. , Stephenson, R. A. "Role of transurethral resection of the prostate in population-based prostate cancer incidence rates" in *American Journal of Epidemiology* 1999; 150: 8: 848-860



# PSASCREENING MODEL

## The PSA screening model

The National Cancer Institute's Statistical Research and Applications Branch has developed a simulator for PSA schedules for arbitrary birth cohorts in the 1916--2000 box. This simulator uses data from the National Health Interview Survey (NHIS)<sup>1</sup> and Surveillance, Epidemiology and End Results (SEER) -- Medicare linked database<sup>2</sup>. To extrapolate the data beyond the original age--year box, generalized additive models (R procedure `\em gam`) were used to smooth the data. A logistic regression model was used for smoothing with the additive main effects of age  $a$  and calendar year  $t$  represented by thin plate regression splines<sup>3</sup>. No interaction smooth terms were specified. Shown in [Figure\\_3b](#), below, is an estimate for the risks of first  $\lambda_{1S}(a, t)$  and secondary  $\lambda_{2S}(a, t)$  PSA tests.

It is clear from the figure that the risk of secondary PSA test is several times higher the one for the first test. This observation prompted the development of the two-stage model for screening based detection described in [Modeling Cancer Detection Through Screening-B](#) and<sup>4</sup>. Frequency of PSA testing by age increases initially as the man enters the risk zone for prostate cancer. However for the older ages a decreasing pattern is observed perhaps because of limited residual life expectancy and associated diminishing relevance of detection of prostate cancer. Dissemination by calendar year is different for the first and secondary tests. In men who have been screened at least once the frequency increases as PSA is introduced into practice and the surface settles at stable values in the nineties. The risk of getting the first test by calendar year shows a spike in early nineties and settles at a lower level later showing a decreasing pattern in the late nineties. This phenomenon deserves further study. The effect could be a consequence of heterogeneity in people's acceptance of PSA testing. The group of men showing compliance for PSA testing is dissipating with time as such men get tested and leave the set of men "at risk" for the first test. Another explanation might be that the recent decline in the frequency of new PSA tests is associated with a dissemination of knowledge of various controversial issues surrounding screening for and treatment of prostate cancer.

## REFERENCES:

- 1 National Center for Health Statistics "National Health Interview Survey (NHIS)." 2004;
- 2 National Cancer Institute "Surveillance Epidemiology and End Results (SEER) - Medicare linked database" 2002;
- 3 Wood, S. "Thin plate regression splines" in Journal of the Royal Statistical Society Series B 2003; 65: : 95-114
- 4 Tsodikov, A., Szabo, A., Wegelin, J. "A population model of prostate cancer incidence" in Statistics in Medicine 2006; 25: 16: 2846-2866



# SURVIVAL COMPONENT

## SUMMARY

This document describes the survival component of the model.

## OVERVIEW

The survival component is an analytic model that describes the relationship between a set of covariates and a survival curve (alternatively a hazard function, a distribution of time to failure, or a density).

## DETAIL

The model yields differing survival curves depending on the following covariates.

- stage of cancer
- tumor grade
- calendar year of diagnosis
- age of patient
- therapy (the integration of this covariate into the model remains as future work.)

Here we present the main results. These results permit us to use composition to build flexible semiparametric survival models (nonlinear transformation models) and use them for estimation and hypothesis testing.

### Nonlinear transformation models

Let  $\gamma(x | \beta, z)$  be a parametrically specified distribution function with the  $x$ -domain of  $[0, 1]$ . Let  $F(t)$  be a nonparametrically specified baseline survival function. A semiparametric regression survival model is called a Nonlinear Transformation Model if its survival function can be represented as a composition

$$G(t | \beta, z) = \gamma \{F(t) | \beta, z\} = (\gamma \circ F)(t | \beta, z).$$

The NTM class and associated estimation procedures were developed by Tsodikov<sup>1</sup>. The key requirement that ensures monotonicity and convergence of the estimation algorithms (see Estimation Algorithm) is that of nondecreasing  $\Theta(x | \cdot)$ , where  $\gamma^{(c)}(x | \cdot) = \partial^c \gamma(x | \cdot) / \partial x^c$ ,  $c = 0, 1, \dots$ ,  $\gamma^{(0)}(x | \cdot) = \gamma(x | \cdot)$ . Using frailty models analogy,  $\Theta(F | \cdot, c)$  can be interpreted as a surrogate of the posterior risk for a subject observed with an event at time  $t$ , where  $c=0$  if right censored,  $c=1$  if failed.



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

### Model building by composition

If  $\gamma_\theta$  and  $\gamma_\eta$  are two different NT models with predictors  $\theta$ , and  $\eta$ , respectively, then

is a new semiparametric model with two predictors  $\theta$  and  $\eta$ . The fact that NTM--generating functions  $\gamma(x|\cdot)$  are all defined on  $x \in [0, 1]$  and have the range in the same interval allows us to compose as complex a hierarchical model as needed. Moreover, we proved that operation of composition preserves the key property of nondecreasing  $\Theta$  observed in frailty models \citep{tsomodelbuilding}. We also derived a chain rule that allows us to specify  $\Theta$  for the compound model based on  $\Theta$ -functions of the submodels

As we will see in the next section, knowledge of  $\Theta$  is all that is needed to specify an estimation procedure.

### Estimation algorithm

Let  $t_i, i = 1, \dots, n$  be a set of times, arranged in increasing order,  $t_{n+1} := \infty$ . Associated with each  $t_i$  is a set  $\mathcal{R}_i$  of subjects  $j \in \mathcal{R}_i$  at risk, with covariates  $z_{ij}$ . For any function  $A(t)$ , let  $A_i = A(t_i)$ ,  $\Delta A_i = |A(t_i) - A(t_i - 0)|$ . The following method (QEM) is used to obtain the profile likelihood.

where  $\{F^{(k)}\}$  and  $\{H^{(k)}\}$  are sequences of functions generated by the self-consistency equation (Equation QEM),  $D_m$  is the number of failures at  $t_m$  and  $\beta$  is a vector of regression coefficients.

It can be shown that if  $\Theta$  is nondecreasing, each update of  $H$  using the self-consistency equation (Equation QEM) strictly improves the likelihood, given  $\beta$ . This guarantees convergence of the sequence of likelihood values  $\ell\{\beta, H^{(k)}\}$  to the profile likelihood of  $\beta$ , and of the sequence  $\{H^{(k)}\}$  to  $H^*$ , the fixed point of (\ref{qem}), under fairly general conditions.

Under a frailty model, the procedure (Equation QEM) is an EM algorithm based on imputation of a missing predictor by its conditional expectation, given observed data, represented by  $\Theta(F|\beta, z, c)$ . Under an NT model, the procedure works as a Quasi-EM algorithm without the missing-data interpretation.

### Profile information matrix

To obtain confidence intervals and tests of statistical hypotheses for regression coefficients, we developed a solution for the exact observed profile information matrix of  $\beta^2$ . As the number of parameters of a semiparametric model is potentially unlimited, obtaining the inverse of the full information matrix becomes computationally prohibitive, and a profile information matrix would be very useful. The profile information matrix can be expressed as

where  $H = (\Delta H_1, \dots, \Delta H_n)^T$  and



$$I_{ab} = - \left. \frac{\partial^2 \ell(\beta, H)}{\partial a \partial b} \right|_{(\beta, H^*(\beta))}$$

for any two vectors  $a$  and  $b$ , where  $\ell$  is a log-likelihood and  $H^*$  is the fixed point of the self-consistency equation. Notice that  $I_{\beta\beta}^r$  has dimension  $d \times d$  with  $d = \dim(\beta)$ , therefore only a small matrix needs to be inverted in order to get an estimator of the covariance matrix of regression coefficients. The downside of (1) is that since  $H^*(\beta)$  is defined implicitly, so is the potentially large Jacobian matrix  $\partial H^*/\partial\beta$ . Therefore, the Jacobian is generally unavailable in a closed form. In the NTM case the problem reduces to solving a system of linear equations  $(D + R)x = b$ , where  $x$  represents a column-vector of the Jacobian,  $D$  is an  $n \times n$  diagonal matrix with diagonal elements  $d_i \neq 0, i = 1, \dots, n$ ,  $R = (R_{kl})$  is a  $n \times n$  matrix,  $R_{kl} = \sum_{i=\max\{k,l\}}^n a_i a_i$ ,  $i = 1, \dots, n$  are real numbers, and  $b$  be an  $n$ -dimensional vector. The main result used to obtain  $\partial H^*/\partial\beta$  is as follows. Let the functions  $\varphi_k : \mathbb{R} \rightarrow \mathbb{R}, k = 1, \dots, n$  be defined recursively as  $\varphi_n(y) = b_n/d_n - a_n/d_n y$ ,  $\varphi_k(y) = \left( b_k - \sum_{i=k}^n a_i y + \sum_{l=k+1}^n \sum_{i=k}^{l-1} a_i \varphi_l(y) \right) / d_k$ ,  $k = n-1, \dots, 1$ . Now, let  $\tilde{\varphi}$  be the function given by  $\tilde{\varphi}(y) = \sum_{k=1}^n \varphi_k(y)$ . The solution to the system of equations  $(D + R)x = b$  is the  $n$ -dimensional vector  $x = (\varphi_1(\tilde{y}), \dots, \varphi_n(\tilde{y}))$ , where  $\tilde{y} = \tilde{\varphi}(0)/(1 + \tilde{\varphi}(0) - \tilde{\varphi}(1))$ .

## RELEVANT ASSUMPTIONS

See [Assumption Overview](#).

## RELEVANT PARAMETERS

Recall that the term "parameter," in the language of this model profile environment, actually refers to a model input.

The inputs to the survival model consist of data from the SEER (Surveillance, Epidemiology and End Results) database, which includes approximately 350,000 men diagnosed with cancer<sup>3</sup>. More information on these data may be found by clicking the Details link in the footnote.

Each man's covariates (age, stage, grade, etc.) enter individually into the survival model. In this, this survival model differs from the incidence model (see [Incidence Component](#)).

## RELEVANT COMPONENTS

The components of the survival model are results that allow us to build flexible semiparametric survival models and use them for estimation and hypothesis testing. Many of these results are new discoveries, developed under this project. Details may be seen above, on this page.

## DEPENDENT OUTPUTS

[Mortality Component](#)



## RELEVANT RESULTS

**Methods.** Advances were made in statistical methods in the development of this component of the model. For more information, see the Relevant Components section of this document.

**Parameter estimates and prediction.** The output of the Survival Component is combined with the output of [Incidence Component](#) and used as input to the [Mortality Component](#), leading to predicted mortality and an estimate of the difference in mortality that is due to the introduction of PSA testing. See [results Age Adjusted Incidence Mortality and Analysis Of Population Data](#).

## REFERENCES:

- <sup>1</sup> Tsodikov, Alex "Semiparametric models: a generalized self-consistency approach" in *Journal of the Royal Statistical Society, Series B* 2003; 65: : 759-774
- <sup>2</sup> Tsodikov, A., Garibotti, G. "Profile information matrix for nonlinear transformation models" in *Lifetime Data Analysis* 2007; 13: 1: 139-159
- <sup>3</sup> National Cancer Institute "Surveillance Epidemiology and End Results (SEER) - Medicare linked database" 2002;



# MORTALITY MODEL

## THE MORTALITY MODEL

Let  $z$  be the vector of clinical covariates observed at diagnosis. The results of [Survival Component](#) allow us to study different survival models  $G_M(\tau|z, a, t)$  (s.f.), where  $\tau$  is the survival time post-diagnosis,  $a$  is age at diagnosis,  $t$  is calendar year of diagnosis, and  $z$  represents clinical covariates. We found that the proportional hazards model does not provide a good fit for the data by stage and grade. The list of adequate survival models for prostate cancer includes the PHPH cure model<sup>1,2</sup> and the proportional odds (PO) model<sup>4</sup>. We prefer the PO model with one linear predictor over the PHPH cure model with two predictors by the AIC model selection criterion. A test for the PO assumption vs. the PH assumption using the Gamma frailty model with covariates in shape and scale parameters of the frailty distribution can be found in Tsodikov2004. Confidence intervals for odds ratios of stage can be found in<sup>5</sup>.

In [Incidence Model](#) we presented a marginal model for cancer incidence by age  $a$  and year of diagnosis  $t$ . The marginal p.d.f.  $f_I(a|x)$  for the  $x$ -cohort,  $x = t - a$  resulting from this model can be partitioned into  $z$ -specific fractions using a regression of  $z$  on age and year,  $f_z(z|a, t)$ , so that  $f_I(a, z|t - a) = f_I(a|t - a)f_z(z|a, t)$ .

In practice, to specify  $f_z$  we use a categorical effect of  $t$  and define  $z$  as a categorical prognostic variable based on stage and grade. Cutpoints on PSA value and its velocity at diagnosis can be used to extend  $z$ . The period categorical variable associated with  $t$  serves as a surrogate of PSA utilization affecting  $z$ -shift and within--stage shift of survival with the introduction of PSA. We consider this approach preliminary in that it only pertains to the actually observed utilization pattern and does not easily generalize to hypothetical PSA impact scenarios (unlike the incidence model). As part of model improvement in this competing continuation application we plan to provide a more sophisticated link between the incidence and survival models that is necessary to address the specific aims of this project.

Now, mortality in year  $t_M$  at the age of  $a_M$ ,  $\lambda_M(a_M, t_M) = \lambda_M(\tau|x)$ , where  $\tau = t_M - t = a_M - a$  and  $x = t - a = t_M - a_M$ , is derived from the prostate cancer specific survival function counted out from birth, represented by the following convolution:

## REFERENCES:

- <sup>1</sup> Tsodikov, A. "Semiparametric models of long- and short-term survival: An application to the analysis of breast cancer survival in utah by age and stage" in *Statistics in Medicine* 2002; 21: : 895-920
- <sup>2</sup> Tsodikov, A., Ibrahim, J.G., Yakovlev, A.Y. "Estimating cure rates from survival data: An alternative to two-component mixture models" in *Journal of the American Statistical Association* 2003; 98: : 1063-1078
- <sup>3</sup> Tsodikov, Alex "Semiparametric models: a generalized self-consistency approach" in *Journal of the Royal Statistical Society, Series B* 2003; 65: : 759-774
- <sup>4</sup> Tsodikov, A. "Using composition to build semiparametric survival models" in *Statistical Modelling* 2006; Submitted:
- <sup>5</sup> Tsodikov, A., Garibotti, G. "Profile information matrix for nonlinear transformation models" in *Lifetime Data Analysis* 2007; 13: 1: 139-159



University of Michigan  
Mortality Model  
References:



University of Michigan  
base Case PSA



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

# BASE CASE PSA

The base case PSA simulation program can be used to simulate life histories of the times that individual men undergo PSA tests. The simulator is based on data from the National Health Interview Survey<sup>1</sup> and the Surveillance, Epidemiology and End Results (SEER) -linked database<sup>2</sup>. The simulation is based on two submodels for the "risk" of PSA test, both of which depend on age and calendar year. The first submodel is a survival model for the time to first PSA test among men who have not yet had a test. The second submodel is a non-homogeneous Poisson process model for the schedule of subsequent PSA tests in men who have already had at least one PSA test. For further information, see [figure First PSAtest Secondary PSAtest](#).

## REFERENCES:

- <sup>1</sup> National Center for Health Statistics "National Health Interview Survey (NHIS)." 2004;
- <sup>2</sup> National Cancer Institute "Surveillance Epidemiology and End Results (SEER) - Medicare linked database" 2002;



University of Michigan  
Seer\_Medicare

# SEER\_MEDICARE

## REFERENCE

National Cancer Institute (2002), "Surveillance Epidemiology and End Results (SEER) - Medicare linked database"

## URL:

<http://healthservices.cancer.gov/seermedicare/>



Readers Guide  
Model Overview  
Assumption Overview  
Parameter Overview  
Component Overview  
Output Overview  
Results Overview  
Key References

## NOTES AND DISCUSSION

The SEER data used in the current study consist of two parts:

- Population data: an age-by-year table with a count in each cell of the number of men at risk for prostate cancer
- Survival data: one row for each patient, with several variables such as age at diagnosis, stage, grade, and outcome

The data are available only from the following locations: San Francisco-Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta. They include all prostate cancer cases diagnosed in these regions, approximately 350,000 men.

CategoryReferences



# LIKELIHOOD IN THE INCIDENCE MODEL

## LIKELIHOOD IN THE INCIDENCE MODEL

Observed data for the incidence model include a count  $P(a, t)$  of people at risk of cancer and a count of cancer cases  $C(a, t)$  by age and year. The conditional likelihood of the data is built as a product of conditional probabilities of cancer detection given that the subject is in the risk set for each  $a, t$  combination from the observed box. Except for terms that do not depend on the model parameters, the likelihood takes the form

$$\ell = \sum_{a,t} C(a, t) \log \lambda_I(a, t) - P(a, t) \lambda_I(a, t).$$

Note that the same likelihood would result if we assumed that  $C$  is Poisson distributed with expectation  $P\lambda_I$  and that  $C(a, t)$  represent independent random variables for different  $(a, t)$  pairs (which is not the case in Equation Likelihood). Maximum likelihood inference is used to obtain point estimates and confidence intervals for the model parameters entering  $\lambda_I$ . Maximization of the likelihood can be regarded as minimizing a certain distance between the empirical incidence  $C/P$  and its model-based counterpart  $\lambda_I$ .



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
Human Mortality Database

# HUMAN MORTALITY DATABASE

## REFERENCE

Wilmoth, John R. (Director), Shkolnikov, Vladimir Shkolnikov (Co-Director), (2003), "Human Mortality Database (HMD)."

## URL:

<http://www.mortality.org/>



Readers Guide  
Model Overview  
Assumption Overview  
Parameter Overview  
Component Overview  
Output Overview  
Results Overview  
Key References

## NOTES AND DISCUSSION

CategoryReferences



# INCIDENCE COMPONENT

## SUMMARY

This document summarizes the incidence component of the model.

## OVERVIEW

The incidence component is an analytic model for the time until prostate cancer diagnosis of a randomly-selected man from the population at risk. This model is expressed in terms familiar from survival analysis, and thus may be expressed as a hazard function, a density, a cumulative distribution function, or a survival function. A likelihood-based estimation procedure is part of this component.

## DETAIL

The incidence model uses population data to estimate [parameters](#) that characterize prostate cancer incidence as a function of age and calendar year.

## RELEVANT ASSUMPTIONS

Please see [Assumption Overview](#).

## RELEVANT PARAMETERS

Recall that the term "parameter," in the language of this model profile environment, actually refers to a model input. The inputs to the incidence model consist of data and of [parameters](#) (in the statistical sense of the word) belonging to a model that has been specified and estimated independently.

The incidence model currently uses data from nine areas of the United States: San Francisco-Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta. Before being used in the estimation procedure, the data are summarized by age ( $a$ ) and calendar year ( $t$ ).

The particular variables summarized by age and calendar year, and used as input to the model, are:

- The number of men at risk for prostate cancer,  $P(a, t)$  for each age  $a$  and calendar year  $t$ .
- The number of new prostate cancer cases,  $C(a, t)$  by age  $a$  and calendar year  $t$ , obtained by computing summaries from the SEER (Surveillance, Epidemiology and End Results) database<sup>1</sup>.

In addition, the estimation procedure uses a prior estimate of the distribution of PSA utilization in the population. This distribution is based on an algorithm that can be used to simulate life histories of the times that individual men undergo PSA tests. For more information see [base Case PSA](#).

## RELEVANT COMPONENTS

- Distribution of time until tumor [onset](#) (see [Incidence Model](#))
- Duration of the latent disease stage (also called delay time), which is further broken down into the following competing risks:



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



- Risk of clinical diagnosis (see [Incidence Model](#))
- Risk of screen-based diagnosis (see [Modeling Cancer Detection Through Screening](#))

(Other means of detection, such as [TURP](#), are not currently included in the model.)

- Likelihood and estimation algorithm (see [Likelihood In The Incidence Model](#))

## DEPENDENT OUTPUTS

### [Mortality Component](#)

## RELEVANT RESULTS

- Predictions of incidence with and without PSA testing (see [results Age Adjusted Incidence Mortality](#)), and consequently
- Estimation of the difference in incidence that is due to PSA testing.

## REFERENCES:

- <sup>1</sup> National Cancer Institute "Surveillance Epidemiology and End Results (SEER) - Medicare linked database" 2002;



# MORTALITY COMPONENT

## SUMMARY

This document describes the mortality component of the model.

## OVERVIEW

The [Mortality Component](#) yields estimates of mortality by calendar year, age, and presence or absence of PSA testing.

## DETAIL

See [Mortality Model](#).

## RELEVANT ASSUMPTIONS

See [Assumption Overview](#).

## RELEVANT PARAMETERS

Recall that the term "parameter," in the language of this model profile environment, actually refers to a model input, not a [parameter](#) in the classical statistical sense.

The inputs to this component are the outputs of the [Incidence Component](#) and [Survival Component](#).

This component does not use population mortality data as input.

## RELEVANT COMPONENTS

## DEPENDENT OUTPUTS

## RELEVANT RESULTS

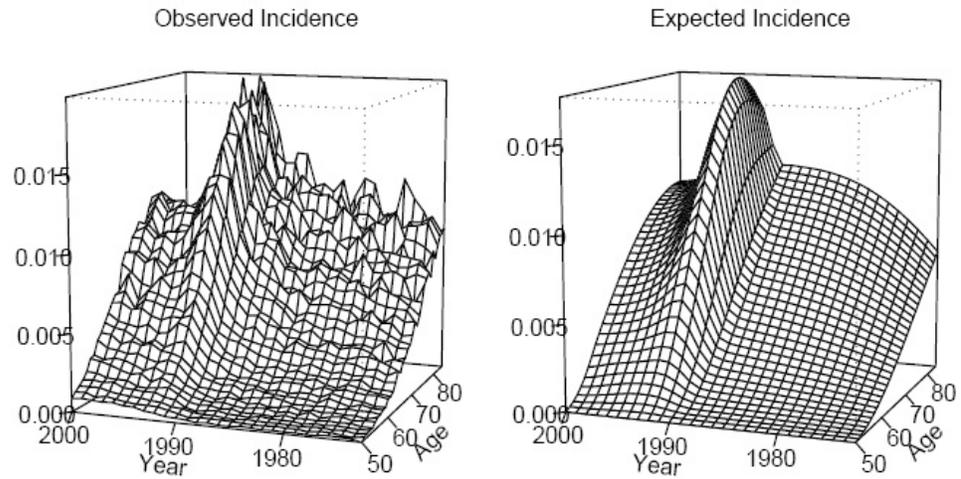
- Predictions of prostate cancer-specific mortality with and without PSA testing (see [results Age Adjusted Incidence Mortality](#)) and consequently
- Estimation of the difference in mortality that is due to PSA testing.





- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

# INCIDENCE FIGURE



**Incidence Figure.** Prostate cancer incidence (rate per person). Observed incidence of prostate cancer is displayed on the left. This is a histogram empirical estimate obtained by dividing incident cancer cases by the population at risk, for each age and each calendar year. Expected incidence is displayed on the right, also by age and calendar year, as predicted by the model. The model captures the basic pattern of prostate cancer incidence. The spike occurring with the introduction of PSA testing gets more pronounced with increasing age, except in very old men. The decrease in older men is a consequence of the fact that latent prevalence accumulates with age.



University of Michigan  
define Lead Time

# DEFINE LEAD TIME

**Lead time** refers to the amount by which detection of prostate cancer is advanced due to PSA screening. It adds to the observed survival time even if early detection and treatment were of no benefit. The lead-time effect targets patients who would still be detected later without screening. This effect could result in apparently improved short-term survival even if there were no mortality benefit.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

# RESULTS OVERDIAGNOSIS LEAD TIME

Estimates of lead time and overdiagnosis reveal the potential natural history of the disease and of population screening exposure over the lifespan of an individual. These parameters are estimated by birth cohort, and presented in [Figure\\_5](#). Overdiagnosis can be variously defined as a fraction of all detected cancers (the solid curve in the left panel of [Figure\\_5](#)), or as a fraction merely of screen-detected cancers (the dashed curve).

Recall that the horizontal axis represents the year of a man's fiftieth birthday, so that older men are represented toward the left of each panel in [Figure\\_5](#) and younger men toward the right. In younger cohorts, more of the cohort life span falls in the PSA era. This leads to a pattern of increasing lead time and increasing overdiagnosis among all detected cancer patients as we move toward the right in each panel (solid curves). For men entering the age risk zone for prostate cancer at the present time, the model predicts about a six-year mean lead time and 25% overdiagnosis among all detected patients.

Overdiagnosis in screen-detected cases is represented by the dashed curve in the left panel. This must always be higher than the solid curve, because screen-detected cases are a proper subset of all cancer cases and thus there is a smaller denominator in computing the fraction. But in addition to being higher than the solid curve, the dashed curve reveals a trend in the opposite direction. This can be understood as follows. Men whose fiftieth birthday occurs in the 1950s were already very old in the PSA era. A prostate cancer detected by screening in a man of this age has a high probability of being overdiagnosed, because of his very small expected residual lifetime. In men who are younger during the PSA era, on the other hand, the pool of screen-detected cases include includes many cancers that would have surfaced clinically in the man's residual lifetime if the man had not received a PSA test. These relevant cancers reduce the proportion of overdiagnosed cancers in younger men. Overdiagnosis in screen-detected cases settles at about 30% for men who reach their fiftieth birthday during the PSA era.



University of Michigan  
define Overdiagnosis



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

## DEFINE OVERDIAGNOSIS

**Overdiagnosis.** A large proportion of prostate cancers identified through screening would never be detected in the absence of screening. This phenomenon is called overdiagnosis. Screening brings such cancers to the surface predominantly in the localized stage of the disease, leading to an apparent "favorable" stage shift. Overdiagnosis has multiple consequences. It leads to over-treatment of men who would never be detected without screening. Also, it modifies apparent estimates of post-treatment survival as over-diagnosed cases appear to be "cured." Injection of overdiagnosed cases into the pool of all prostate cancer presentations at diagnosis changes the distribution and the meaning of clinical covariates in men diagnosed with prostate cancer in the PSA era. Overdiagnosis could lead to apparently improved long-term survival of patients with localized stage of the disease even if there were no mortality benefit.



# MEAN POSTERIOR DELAY TIME MARGINAL INCIDENCE MODEL

---

[Figure\\_8](#) shows an estimate of delay time computed from SEER data ([Analysis Of Population Data](#)). Introduction of PSA testing is associated with earlier detection, and the older the man the more so. The slight decrease in delay time in the no-PSA prediction is a transient process resulting from freezing the pre-PSA trend estimates in the year 1988.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

For more information on the model from which this estimate was obtained, see [Incidence Model](#).

For more information on delay time and its integration into the model, see [Delay\\_Time\\_Approach](#).



# WITHIN STAGE SHIFT BY DELAY TIME

[Figure\\_9](#) shows that the delay-time approach, developed in the current project, captures within-stage-and-grade shift. It should be stressed that delay time was computed without using survival data, and represents an independent significant prognostic factor for post-treatment survival, particularly in the localized-regional. These results can be compared with [Figure\\_2](#), where a similar effect is expressed by year of diagnosis. For estimates of delay time as a function of age and calendar year, see [Mean Posterior Delay Time Marginal Incidence Model](#). For more information on the model from which these estimates were obtained, see [Delay\\_Time\\_Approach](#).



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)



# RESULTS AGE ADJUSTED INCIDENCE MORTALITY

In [Figure\\_6](#), model predictions are displayed for age-adjusted incidence and mortality, along with their empirical estimates. To generate predictions in the counterfactual case of no PSA testing, all trend functions were frozen at a constant in the year 1988, and PSA sensitivity was set at zero. The mortality figure (right) indicates that the introduction of the PSA test has led to a decline in mortality.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

For a description of the data analysis that yielded these results, see [Analysis Of Population Data](#).

For a deeper understanding of the model components on which the results are based, see [Incidence Model](#) and [Mortality Model](#).



# ANALYSIS OF POPULATION DATA

## ANALYSIS OF POPULATION DATA

The SEER database was used to obtain data on more than 350,000 cases of prostate cancer diagnosed in nine areas of the United States (San Francisco-Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta) as well as population count files corresponding to those cases. We use the modeling box corresponding to age interval [50,85] and calendar year interval [1973-2000]. Age distribution in the U.S. population in year 2000 for men over 50 is used as a standard when age-adjusted characteristics are reported. Risk of death from other causes (used in estimates of lead-time and overdiagnosis) was derived from the [Human Mortality Database](#)<sup>1</sup>.

As shown in [Figure\\_1](#), incidence of prostate cancer before the introduction of PSA showed an increasing trend in calendar time reportedly related to TURPs<sup>2</sup>. In order to model this effect, a linear trend was specified for the sojourn time model ([Equation LCDX](#)) for the period 1973-1987, saturating in 1988. The parameter  $c$  specifies the slope of the trend during 1973-1987. We did not have a compelling evidence for changes in the onset time distribution over time, and  $T_O(t)$  was removed from the model. Also, we did not find any improvement in the fit from introducing a correlation between age at onset and the sojourn time, and this term was removed from the model. PSA sensitivity was specified as an increasing function of the time since tumor onset. When fitting the model, the estimate settled at 100% sensitivity. Likelihood was maximized by the Powell's method (Himmelblau1972) of conjugate directions. Confidence intervals for the model parameters are based on Likelihood Ratio and inverting the profile likelihood surface for each parameter. Estimates of key model parameters and the corresponding confidence intervals are shown in [Table\\_1](#).

Note that the estimated mean age at tumor onset goes well beyond the normal human lifetime. This is a consequence of the fact that only a proportion of men would ever develop prostate cancer in their life span. Shown in [Figure\\_4](#) is a histogram empirical estimate of prostate cancer incidence  $C(a, t)/P(a, t)$  and its model--predicted counterpart  $\lambda_I(a, t)$  by age and calendar year.

The model captures the basic pattern of prostate cancer incidence. The spike effect in the incidence occurring with the introduction of PSA gets more pronounced with age except for very old people. This is a consequence of latent prevalence of the disease accumulating with age. Shown in [Figure\\_5](#) is an estimate of lead time and overdiagnosis. Both notions relate to the potential natural history of the disease and population screening exposure over the life span of an individual. Therefore we represent them by birth cohort. Overdiagnosis can be measured as a fraction relative to all detected cancers or to screen-detected cancers only. As we move the year of birth to the right, more and more of the cohort life span falls on the PSA era. This leads to an increasing pattern of lead time and overdiagnosis among all detected cancer patients (solid curves). For men entering the age risk zone for prostate cancer at the present



time, the model predicts about 6-year mean lead time and 25% overdiagnosis among all detected patients. Interestingly, overdiagnosis in screen-detected cases is a decreasing function of the birth year and settles at about 30% for the present era. Initially for a person born in the fifties only older ages are affected by PSA utilization. If detected at such an age, the case is very likely to be overdiagnosed. Indeed, if screening were ignored the disease would have little chance to surface because of the very small expected residual lifetime in older people. This is why the dashed curve in [Figure\\_5](#) (left) starts high. As we move the potential life history more and more under the PSA exposure, the pool of screen-detected cases gets enriched with relevant cancers that have advanced diagnosis due to PSA yet would surface clinically in their potential residual lifetime if PSA were not applied. Since screen-detected cases represent a subset of all cancer cases, overdiagnosis relative to screen-detected cases (the dashed curve) is always higher than the one relative to all cancer cases (the solid curve). Shown in [Figure\\_6](#) are model predictions for age-adjusted incidence and mortality and their empirical estimates. To generate predictions without PSA, all trend functions were frozen at a constant in year 1988, and PSA sensitivity was set at zero. The mortality figure (right) indicates that introduction of PSA test has led to mortality decline. Explaining this effect and its partitioning into fractions attributable to early detection and treatment is one of the emphases of this project.

The model was implemented in a prototype software package for Windows that brings incidence, survival, mortality and other model blocks into a common GUI shell that uses unified data input, output, menu and graphics structure. Shown in [Figure\\_7](#) are screen shots of the software.

## REFERENCES:

- <sup>1</sup> Wilmoth, John R. (Director), Shkolnikov, Vladimir Shkolnikov (Co-Director), "Human Mortality Database (HMD)." 2003;
- <sup>2</sup> Merrill, R.M., Feuer, E. J., Warren, J. L. , Schussler, N. , Stephenson, R. A. "Role of transurethral resection of the prostate in population-based prostate cancer incidence rates" in American Journal of Epidemiology 1999; 150: 8: 848-860



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

# TABLE\_1

| Parameter   | Legend                              | Point Estimate | 95% CI             |
|-------------|-------------------------------------|----------------|--------------------|
| $\mu_{CDx}$ | Mean baseline sojourn time          | 18.558         | (18.345, 18.775)   |
| $s_{CDx}$   | Shape sojourn time                  | 1.541          | (1.5191, 1.5644)   |
| $c$         | Slope of trend for sojourn time     | 0.09354        | (0.09068, 0.09641) |
| $\mu_O$     | Mean age past 50 at tumor onset     | 72.732         | (72.498, 72.965)   |
| $s_O$       | Shape of age past 50 at tumor onset | 1.6153         | (1.6067, 1.6239)   |

[Table\\_1](#). Estimates of model parameters and confidence intervals. Time and age are measured in years.



# MODELING CANCER DETECTION THROUGH SCREENING

## MODELING CANCER DETECTION THROUGH SCREENING

Let  $G_{SDx}(\xi|x, y)$  be the survival function of the potential time to screen-based diagnosis (SDx). This section is devoted to modeling this distribution.



Readers Guide  
Model Overview  
Assumption Overview  
Parameter Overview  
Component Overview  
Output Overview  
Results Overview  
Key References

For an arbitrary individual from the target population, consider the "risk" of getting his first screen. Age at first screen may be regarded as a survival time with the instantaneous risk represented by the hazard function  $\lambda_{1S}(a, t)$  that depends on age and calendar year. An empirical histogram estimate for  $\lambda_{1S}(a, t)$  can be obtained by dividing the number of subjects at the age of  $a$  receiving their first screen in year  $t$  by the total number of person-years with no evidence of the disease in the  $(a, t)$  cell. Of course, this estimate is inconsistent unless the data are grouped<sup>1</sup>.

The probability of no screens by the age of  $a$ ,  $G_{1S}$ , is a survival function obtained by integrating the hazard  $\lambda_{1S}$  over a life line on the so-called Lexis diagram<sup>2</sup>:

Denote by  $\lambda_{2S}(a, t)$  the intensity of screening in subjects who have already had their first screen. The fact that the subject has had his first PSA test may identify him as a member of the group that enjoys a higher screening utilization for various reasons. Therefore,  $\lambda_{2S}$  is larger than  $\lambda_{1S}$ , as we see in [Figure\\_3](#).

Consider the unconditional probability  $G_{2SDx}(\tau|x, a, y)$  that a subject is not diagnosed by screening in the age interval  $[a, a + \tau]$ ,  $a \geq y$ . Under the assumptions stated in our Assumptions section, we have

$$G_{2SDx}(\tau|x, ay) = \exp \left\{ - \int_{\max(y-a, 0)}^{\tau} \lambda_{2S}(a + \zeta, x + a + \zeta) \alpha(\zeta + a - y) d\zeta \right\},$$

where  $\int_a^b = 0$  for any  $b \leq a$ . The conditional probability of no screening diagnosis by the age of  $y + \xi$ ,  $G_{SDx}(\xi|x, y)$ , takes the form

$$G_{SDx}(\xi|x, y) = G_{1S}(y+\xi|x) + \overline{G}_{1S}(y|x) G_{2SDx}(\xi|x, y, y) + \int_0^\xi \overline{\alpha}(\nu) f_{1S}(y + \nu|x) G_{2SDx}(\xi - \nu|x, y + \nu, y) d\nu.$$

The first term in this equation addresses the possibility of no screens by the age of  $y + \xi$ . The second term addresses the situation when the first screen occurs before onset of the disease at the age of  $y$  and no diagnosis is achieved through secondary screens that might happen in the age interval  $(y, y + \xi)$ . The third term accumulates the probability that cancer is missed at the first and secondary screens occurring after disease onset.



University of Michigan  
Modeling Cancer Detection Through  
Screening  
References:

## REFERENCES:

- 1 Burr, D. "On inconsistency of Breslow's estimator as an estimator of the hazard rate in the cox model" in *Biometrics* 1994; 50: 4: 1142-1145
- 2 Breslow, N. E., Lubin, J. H. , Marek, P., Langholz, B. "Multiplicative models and cohort analysis" in *Journal of the American Statistical Association* 1983; 78: 381: 1-12



University of Michigan  
Publications



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

## PUBLICATIONS

- Broet, P., , Tsodikov, A., , De Rycke, Y., , Moreau, T.** (2004) Two-sample statistics for testing the equality of survival functions against improper semi-parametric accelerated failure time alternatives: An application to the analysis of a clinical trial in *Lifetime Data Analysis* 10 : , p 103-120
- Tsodikov, A., Garibotti, G.** (2007) Profile information matrix for nonlinear transformation models in *Lifetime Data Analysis* 13 : 1 , p 139-159
- Tsodikov, A., Ibrahim, J.G., Yakovlev, A.Y.** (2003) Estimating cure rates from survival data: An alternative to two-component mixture models in *Journal of the American Statistical Association* 98 : , p 1063–1078
- Tsodikov, A.** (2006) Using composition to build semiparametric survival models in *Statistical Modelling* Submitted :
- Tsodikov, A., Szabo, A., Wegelin, J.** (2006) A population model of prostate cancer incidence in *Statistics in Medicine* 25 : 16 , p 2846-2866
- Tsodikov, Alex** (2003) Semiparametric models: a generalized self-consistency approach in *Journal of the Royal Statistical Society, Series B* 65 : , p 759–774



University of Michigan  
Over Diagnosis



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

# OVER DIAGNOSIS

A large proportion of prostate cancers identified through screening would never be detected in the absence of screening. This phenomenon is called overdiagnosis. Screening brings such cancers to the surface predominantly in the localized stage of the disease, leading to an apparent "favorable" stage shift. Overdiagnosis has multiple consequences. It leads to over-treatment of men who would never be detected without screening. Also, it modifies apparent estimates of post-treatment survival as over-diagnosed cases appear to be "cured." Injection of overdiagnosed cases into the pool of all prostate cancer presentations at diagnosis changes the distribution and the meaning of clinical covariates in men diagnosed with prostate cancer in the PSA era. Overdiagnosis would lead to apparently improved *long-term* survival of patients with localized stage of the disease even if there were no mortality benefit.



University of Michigan  
Length Bias



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

# LENGTH BIAS

It has long been recognized that screening preferentially detects slower growing tumors<sup>1</sup>. Slower growing tumors are likely to be associated with better prognosis. Among other consequences, length-bias effect would lead to apparently worsened survival of patients in distant stage under screening as compared to the unscreened population. Indeed, the pool of advanced tumors detected in the unscreened population is heterogeneous in terms of growth rates. With the introduction of screening some of the would-be distant cases will be detected earlier in a localized stage. These are likely to be the "best" slower growing fraction of the would-be distant cases. As a result, cases missed by screening that are still detected with distant disease under screening, show worse prognosis as compared to the unscreened population.

## REFERENCES:

- <sup>1</sup> Zelen, M., Feinleib, M. "On the theory of screening for chronic diseases" in *Biometrika* 1969; 56: : 601-614



University of Michigan  
Lead Time

# LEAD TIME

Lead-time measures an advance in the diagnosis of prostate cancer due to screening. It adds to the observed survival time even if early detection and treatment were of no benefit. The lead-time effect pertains to patients who would still be detected later without screening. Lead-time would lead to apparently improved *short-term* survival even if there were no mortality benefit.

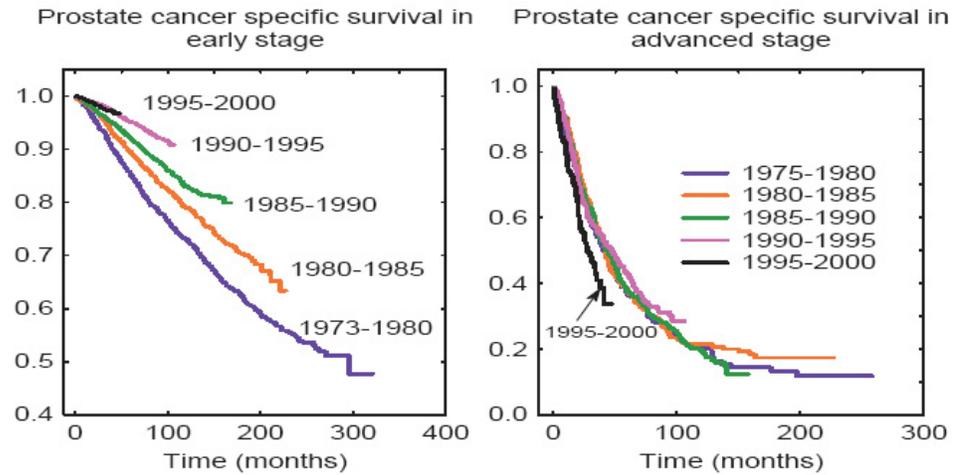


[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

## FIGURE\_2



**Figure\_2.** Within-stage shift. Prostate cancer specific survival by year of diagnosis and stage. The trend reflects improvement of prognosis in localized disease and worsening of prognosis in distant disease with dissemination of screening. Lead-time, length-bias and overdiagnosis provide part of the explanation for the within-stage shift. Data from the Surveillance, Epidemiology and End Results (SEER) database, National Cancer Institute.



# DELAY\_TIME\_APPROACH

## LINKING INCIDENCE AND SURVIVAL: THE DELAY TIME APPROACH

on the level of the subject, age at cancer diagnosis and survival post-diagnosis as well as cancer-specific lifetime are confounded by screening schedules. this confounding is expressed through lead-time, length bias and stage- and within-stage shift (see [Modeling Cancer Detection Through Screening](#)). As a consequence, as we discussed earlier, subjects with different screening schedules will have different distributions of age at diagnosis, and, clinical covariates being equal, they will still show different survival (within-stage shift). We have shown preliminary evidence that within-stage shift is a very significant effect in prostate cancer (Figure\_2). The within-stage shift effect is a consequence of heretogeneity in the latent natural history of the disease and its strong effect on cancer detection processes. As a result, conditioning on different screening histories, other things equal, selects different subsets of natural histories of the disease. It then comes at no surprise that different natural histories are associated with different prognosis. In population data, person-level screening schedules are typically unavailable. Latent heterogeneity in the population data is much higher due to the contribution of uncertainty in screening schedules. In our population model, discussed in the Model Description above, the effect of screening on survival was modeled through the observed stage-shift and the within-stage shift adjusted for empirically through the categorized year of diagnosis variable. This allowed us to make mortality prediction within the observed period of 1973-2000, including a no-PSA predictive run performed by freezing all trend functions in 1988 and removing PSA from the incidence model. In this project we plan to improve the predictive potential of the model by linking population characteristics of interventions such as utilization of PSA, TURPs, Treatment, etc., to survival through natural history surrogates without using empirical variables such as year of diagnosis. This would make predictions a function of utilization characteristics and enable long-term predictions, optimization of interventions, unbiased assessment of treatment effects from population data, and many other model applications discussed below.

Our approach to linking incidence and survival using population data will be based on the concept of frailty. Conditional on the information available at diagnosis

1. Age at diagnosis,  $a$ , year of diagnosis,  $t$ ,
2. Clinical covariates observed at diagnosis,  $z$ ,
3. Dissemination of interventions over calendar time (rates of PSA testing, TURPs),  $d(t)$ ,

we will derive the posterior distribution of the age at tumor onset  $Y$ . With the marginal incidence model discussed in [Incidence Model](#) (we continue using the notation introduced in this section), we have

$$Y | \{a, t, d(t)\} \propto f_O(y|a, t, d(t)) = \frac{f_I(a|x, y) f_O(y|x)}{f_I(a|x)},$$

where  $x = t - a$  is the year of birth.

Delay time  $A$  is the time interval from tumor onset to diagnosis,  $A = a - Y$ . We will use

$A$  as a surrogate of the natural history of the disease as far as its effect on survival is concerned. A frailty model will be formulated for survival with  $A$  as a frailty variable. The effect of early detection due to surveillance is expressed as decreasing  $A$ . As a result, the survival function post-diagnosis will be a functional of the distribution of  $A$ .

In our first approach we will summarize the effect of  $A$  on survival by using the posterior mean delay time as a covariate for survival. Using a specific complete data model for survival, given  $A$ , say, a proportional hazards frailty model, may lead to model misspecification. However, if we keep the form of the incomplete data semiparametric survival model flexible, a model building procedure based on the data will absorb such misspecification. This approach is more attractive than traditional frailty modeling, since the choice of a complete data model is difficult to justify anyway, because complete data are not available. Survival methodology developed in our previous project ([Survival Component](#)) specifically addresses flexible model building procedures and guarantees that inference procedures will be available for all models constructed by these procedures.

[Figure\\_8](#) shows mean posterior delay time (DT) ([Equation Post Y](#)) computed using the marginal incidence model fit to SEER data ([Analysis Of Population Data](#)).

From the [Figure\\_8](#) it is clear that introduction of PSA is associated with earlier detection, and the older the person the more so. The slight decrease of DT in the no-PSA prediction is a transient process resulting from freezing the pre-PSA trend estimates in year 1988. [Figure\\_9](#) shows that the DT approach captures the within-stage--and--grade shift. It should be stressed that DT was computed without using survival data and represents an independent significant prognostic factor for post-treatment survival, particularly in localized/regional stage (compare with [Figure\\_2](#) where a similar effect is expressed by year of diagnosis).

We will proceed as follows.

1. Identify a regression model  $f(z|x, y, \xi)$  of stage and grade ( $z$ ) at diagnosis conditional on independent variables represented by year of birth  $x$ , age at tumor onset  $y$ , and delay time  $\xi = a - y$ . This is a regression model for ordered categorical response. We will consider Proportional Odds and Continuation Ratio models. If necessary, custom models will be developed specifically to address this problem if the modelled effect proves to be non-standard. The PI has experience developing models for ordered categorical data and an efficient algorithm for statistical inference with general ordinal models<sup>1</sup>.
2. Using the marginal incidence model  $f_I(a|x, y)$  [Incidence Model](#), obtain the joint distribution of age and clinical covariates at diagnosis  $f(a, z|x, y) = f(z|x, y, a)f_I(a|x, y)$ . This will contribute to the refined stage- and grade--specific incidence model block that does not use year of diagnosis as a surrogate variable to model stage and grade shift. This block for  $f(a, z|x)$  is obtained by integrating out  $y$ .



3. Using the joint distribution  $f(a, z|x)$ , develop a model block for predicting posterior mean age at tumor onset by extending (Equation Post Y) to include information on stage and grade at diagnosis. This will improve prediction of DT and the proportion of explained variation in post-treatment survival attributable to the within stage shift.
4. Develop the procedure that will adjust survival time for known screening utilization patterns.

Given a population sample of prostate cancer survival, the procedure will be organized as follows.

1. With the stage- and grade--specific prostate cancer incidence model, obtain the mean posterior DT for each subject in the survival sample.
2. Determine an adequate semiparametric model for survival data with covariates represented by  $a, t = x + a$  and mean posterior DT. This would require a trial and error loop through model building using composition techniques (Model Building By Composition), fitting using the Quasi-EM algorithm (Estimation Algorithm) and hypotheses testing using the profile information matrix (Profile Information Matrix).

The significance of our DT approach is that it adjusts survival model for a complex "early detection" confounder. In clinical trials, ignoring significant confounders leads to underestimated treatment effects. With population data, straightforward estimation of the treatment effect is biased since screening utilization is uncontrolled for, and study design is retrospective without randomization. In this project the DT approach serves two main purposes

1. An unbiased assessment of treatment effects with population (and generally retrospective, nonrandomized) data ([Aim Develop Unbiased Assessment Of Treatment Effects](#)).
2. Enables a model that can predict cancer mortality under arbitrary scenarios of utilization of screening and treatment. This paves the way to partitioning mortality into attributable fractions ([Aim Study The Joint Effect Of Progress](#)) short- and long-term predictions of mortality trends ([Aim Make Short And Long Term Predictions](#)), predicting the effects of cancer control strategies that have never been used before, optimization of screening schedules ([Aim Determine Evaluate Optimal Screening Strategies](#)), and addressing other specific aims of this project.



We recognize that mean DT may not provide all of the necessary reduction in the unexplained variation of survival. This will be evaluated by preserving year of diagnosis variable in the model jointly with DT and assessing whether it is still significant. If it turns out that the use of year of diagnosis for explanation of the within stage shift is still necessary, we will extend the DT approach to include the variance of the posterior delay time in addition to the mean and will develop an adjustment of survival using both parameters (mean and variance). Also, as a better but more computer intensive alternative, we will consider using a frailty model approach where posterior distribution for the delay time is used for the frailty variable. The frailty would then represent the uncertainty in the tumor onset given information available at diagnosis in a functional way rather than by one or two surrogate parameters. If necessary, year of diagnosis trend in addition to the DT-adjustment will be preserved to cover yet unexplained variation of survival.

#### REFERENCES:

- <sup>1</sup> Tsodikov, Alexander "A Proportional Hazards Model Taking Account of Long-Term Survivors" in *Biometrics* 1998; 54: 4: 1508-1516

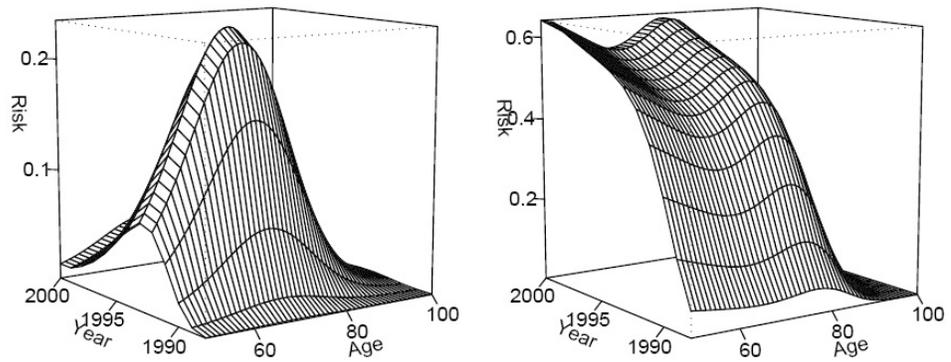
# FIGURE\_3B



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

First PSA Test

Secondary PSA Test



Figure\_3b. Risks of first  $\lambda_{1S}$  and secondary  $\lambda_{2S}$  PSA tests as estimated from the simulation model by age and calendar year. Left: Proportion of never screened men at risk getting their first PSA test. Right: Proportion of men screened at least once getting a secondary PSA test.

## REFERENCES:

- <sup>1</sup> Tsodikov, A., Szabo, A., Wegelin, J. "A population model of prostate cancer incidence" in *Statistics in Medicine* 2006; 25: 16: 2846-2866

# MODELING CANCER DETECTION THROUGH SCREENING-B

## Modelling cancer detection through screening



Readers Guide  
Model Overview  
Assumption Overview  
Parameter Overview  
Component Overview  
Output Overview  
Results Overview  
Key References

This section (from<sup>1</sup>) is devoted to modeling the distribution of potential time to screen-based detection  $G_{SDx}(\xi|x, y)$  conditional on the year of birth  $x$  and age at tumor onset  $y$ . It is a somewhat more detailed version of [Modeling Cancer Detection Through Screening](#).

For an arbitrary individual from the target population, consider the "risk" of getting the first screen in his life. Age at first screen may be regarded as a survival time with the instantaneous risk represented by the hazard function  $\lambda_{1S}(a, t)$ . Naturally,  $\lambda_{1S}$  depends on age  $a$  of the person and the current calendar year  $t$ . Generally, it is expected that  $\lambda_{1S}(a, t)$  increases in  $t$  starting with the year of PSA introduction. As a function of  $a$ , it is reasonable to expect that  $\lambda_{1S}(a, t)$  is increasing initially while the residual life expectancy is still substantial and then decreasing for very old people. An empirical histogram estimate for  $\lambda_{1S}(a, t)$  can be obtained by dividing the number of subjects at the age of  $a$  receiving their first screen in year  $t$  by the total number of subjects with no evidence of the disease in the  $(a, t)$  cell. More precisely, we should count tests in the interval  $(t, t + dt)$  and divide by  $dt$ , which results in the same estimate for the grouping interval  $dt = 1$  year. Note that this estimate is inconsistent unless the data are grouped<sup>2</sup>.

The evolution of an  $x$ -birth cohort up to the age of  $a$  can be represented as a line connecting points  $(\tau, x + \tau)$ , where  $\tau \in [0, a]$  on the age by year plane called the Lexis diagram<sup>3</sup>. The probability of no screens by the age of  $a$ ,  $G_{1S}$ , is a survival function obtained by intergrating (accumulating) the hazard  $\lambda_{1S}$  over the line

$$G_{1S}(a|x) = \left\{ - \int_0^a \lambda_{1S}(\tau, x + \tau) d\tau \right\}.$$

Denote by  $\lambda_{2S}(a, t)$  the intensity of screening in subjects who already had their first screen. Generally, we expect  $\lambda_{2S}$  to be larger than  $\lambda_{1S}$ . Indeed, the fact that the subject has had his first PSA test may identify him as a member of the group that is screened more frequently for reasons such as easier access to secondary testing having done this once already, favorable attitude towards screening in those who choose to have their first test, doctor's recommendations for serial secondary screens following the first one, etc.

The model for risk of diagnosis by cancer screening is based on the following assumptions.

- The probability that a subject born in year  $x$  who has never been screened by the age of  $a$  receives his first screen in the age interval  $(a, a + da)$  is  $\lambda_{1S}(a, x + a)da + o(da)$ .
- The probability that a subject born in year  $x$  who has been screened at least once by the age of  $a$  receives a screen in the age interval  $(a, a + da)$  is  $\lambda_{2S}(a, x + a)da + o(da)$ . This assumption defines secondary screens as following a non-homogeneous Poisson process in age with intensity  $\lambda_{2S}(a, x + a)$ .

- The probability that a subject born in year  $x$ , with the disease onset at the age of  $y$ , screened at the age of  $a$  is detected with cancer is

$$\begin{cases} 0, & y > a, \\ \alpha(a - y), & \text{otherwise,} \end{cases},$$

where  $\alpha(\xi)$  is the sensitivity of screening, and  $\xi$  is the age of tumor at the time of testing. It is natural to specify  $\alpha(\xi)$  as an increasing function.

It should be noted that the fact that  $\lambda_{1S} \neq \lambda_{2S}$  violates the notion that the entire screening schedule for a subject could be a realization of a non-homogeneous Poisson process.

Consider the probability  $G_{2SDx}(\tau|x, a, y)$  that a subject born in year  $x$ , with onset of the disease at the age of  $y$  who has had his first screen by the age of  $a$  is not diagnosed by screening in the age interval  $[a, a + \tau]$ ,  $a \geq y$ . Note that this is a probability of no event in the interval  $[0, \tau]$  for a non-homogeneous Poisson process in  $\zeta \in [0, \tau]$  with intensity  $\lambda_{2S}(a + \zeta, x + a + \zeta)$  thinned with probability  $\bar{\alpha}(\zeta + a - y) = 1 - \alpha(\zeta + a - y)$ . (We use the notation  $\bar{A} = 1 - A$  for any  $A$ .) The intensity of a Poisson process with intensity  $\lambda$  thinned with probability  $\bar{\alpha}$  is given by the product  $\lambda\bar{\alpha}$ , so that with  $a \geq y$ ,

$$G_{2Dx}(\tau|x, a, y) = \exp \left\{ - \int_0^\tau \lambda_{2S}(a + \zeta, x + a + \zeta) \alpha(\zeta + a - y) d\zeta \right\}.$$

If the interval in question is before onset,  $a + \tau \leq y$ , then there is no diagnosis and  $G_{2SDx}(\tau|x, a, y) = 1$ . If  $a < y$  and  $a + \tau > y$ , the time interval in  $\zeta$  where diagnosis is possible starts at  $y - a$ , so that  $G_{2SDx}(\tau|x, a, y)$  is given by an expression similar to (\ref{g2dxc}) with the lower limit in the integral set at  $y - a$ . Summarizing, we have

$$G_{2SDx}(\tau|x, a, y) = \exp \left\{ - \int_{\max(y-a, 0)}^\tau \lambda_{2S}(a + \zeta, x + a + \zeta) \alpha(\zeta + a - y) d\zeta \right\},$$

where  $\int_a^b = 0$  for any  $b \leq a$ .

We are now equipped to derive the probability of no screening diagnosis by the age of  $y + \xi$ , conditional on year of birth  $x$  and age at disease onset  $y$ , where  $\xi$  is time since onset. We have

$$G_{SDx}(\xi|x, y) = G_{1S}(y + \xi|x) + \bar{G}_{1S}(y|x) G_{2SDx}(\xi|x, y, y) + \int_0^\xi \bar{\alpha}(\nu) f_{1S}(y + \nu|x) G_{2SDx}(\xi - \nu|x, y + \nu, y) d\nu.$$

The first term in the above equation addresses the possibility of no screens by the age of  $y + \xi$ . The second term addresses the situation when the 1st screen occurs before onset of the disease at the age of  $y$  and no diagnosis is achieved through secondary screens that might happen in the age interval  $(y, y + \xi)$ . The third term accumulates the probability that cancer is missed at the first and secondary screens occurring after disease onset.



## REFERENCES:

- 1 Tsodikov, A., Szabo, A., Wegelin, J. "A population model of prostate cancer incidence" in *Statistics in Medicine* 2006; 25: 16: 2846-2866
- 2 Burr, D. "On inconsistency of Breslow's estimator as an estimator of the hazard rate in the cox model" in *Biometrics* 1994; 50: 4: 1142-1145
- 3 Breslow, N. E., Lubin, J. H. , Marek, P., Langholz, B. "Multiplicative models and cohort analysis" in *Journal of the American Statistical Association* 1983; 78: 381: 1-12



University of Michigan  
Equation QEM

# EQUATION QEM

$$\Delta H_m^{(k+1)} = \frac{D_m}{\sum_{ij \in \mathcal{R}_m} \Theta(F_i^{(k)} | \beta_{ij}, z_{ij}, c_{ij})}$$



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

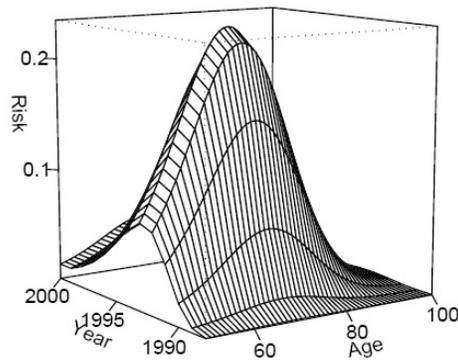


# FIGURE FIRST PSATEST SECONDARY PSATEST

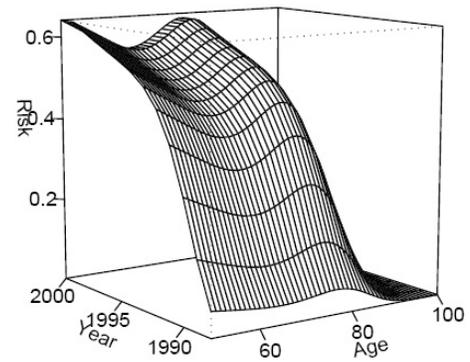


- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

First PSA Test



Secondary PSA Test

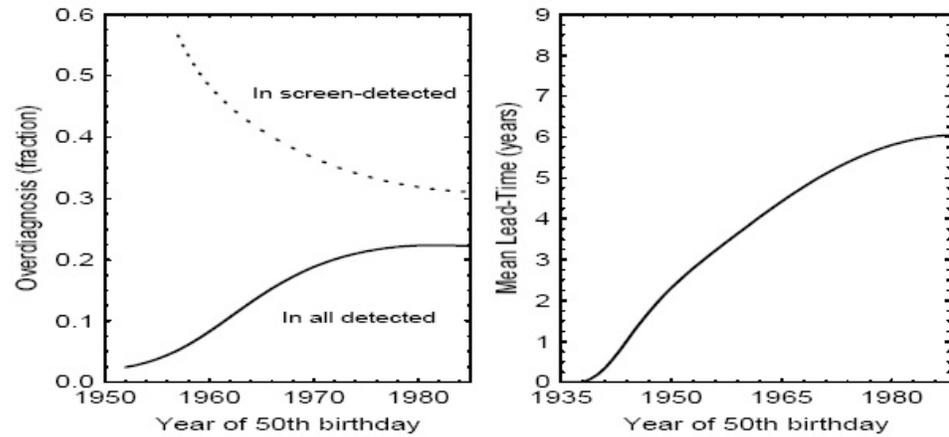


Risks (rate per person) of first  $\lambda_{1,S}$  and secondary  $\lambda_{2,S}$  PSA tests as estimated from the NIH simulation model by age and calendar year. This simulator is based on data from the National Health Interview Survey (NHIS) and Surveillance, Epidemiology and End Results (SEER) - Medicare linked database ([Seer\\_Medicare](#)). Left: Proportion of never screened men at risk of getting their first PSA test. Right: Proportion of men screened at least once getting a secondary PSA test. Original simulated data were smoothed by a generalized additive model with a logit link.

For more information see [PSAscreening Model](#).



# FIGURE\_5



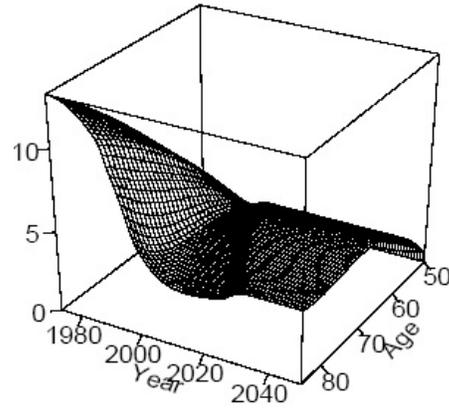
Figure\_5. Overdiagnosis (left) and lead-time (right) by birth cohort. Dashed line is the fraction of overdiagnosis in screen-detected patients. Solid line (left) is the fraction of overdiagnosis in all cancer patients.



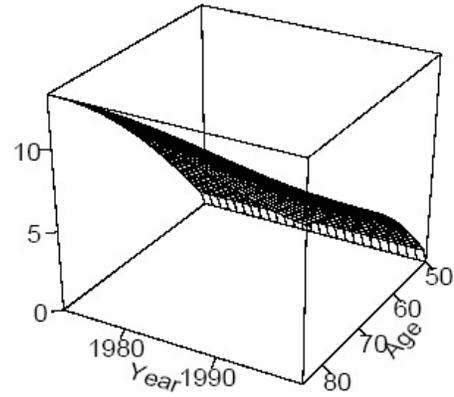
- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

# FIGURE\_8

(A) PSA, Long-term



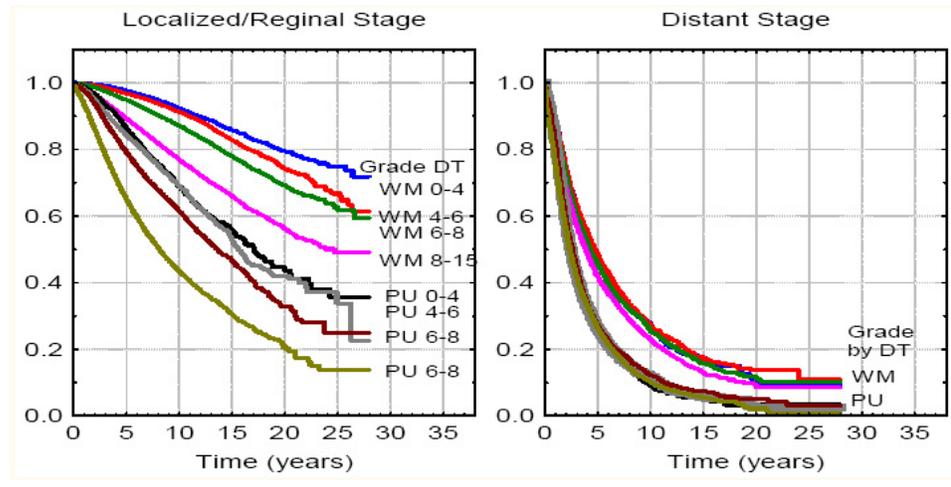
(B) No PSA



Figure\_8. Mean posterior delay time by age and year of diagnosis with (A) and without (B) PSA screening. Estimates from SEER data.



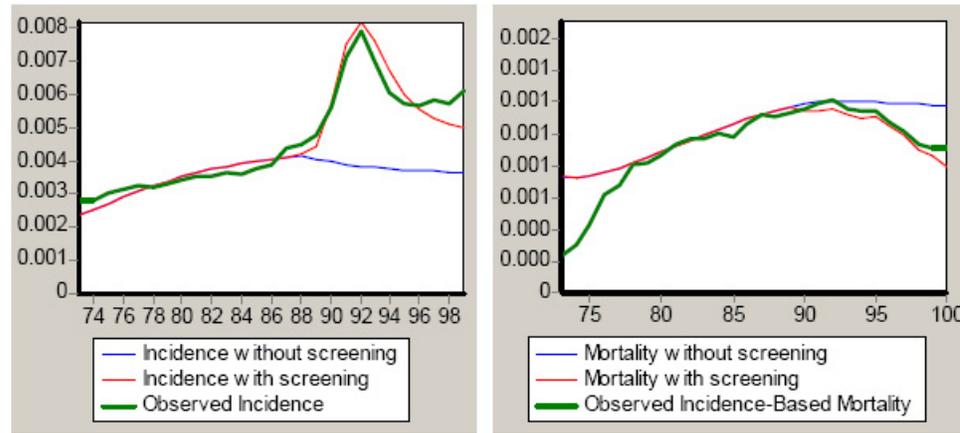
# FIGURE\_9



Figure\_9. Survival by mean posterior delay time (DT), stage (Localized/Regional, Distant) and Grade (Well or Moderately (WM), Poorly or Undifferentiated (PU)). Estimates were obtained from SEER data.



# FIGURE\_6



Figure\_6. Age-adjusted estimates of incidence (left) and mortality (right) in the presence of PSA (red) and prediction of the no-PSA case (blue). Rates are given per person. Thick green curves correspond to empirical estimates. Model-based predictions show overall mortality while empirical estimate is for incidence-based mortality only for cases diagnosed between 1973 and 2000. The discrepancy in the mortality figure for years close to 1973 shows the effect of prostate cancer cases prevalent in 1973.



University of Michigan  
Equation LCDX

# EQUATION LCDX

$$\lambda_{CDx}(\xi|x, y) = h_{CDx}(\xi|y)T_{CDx}(x + y + \xi).$$

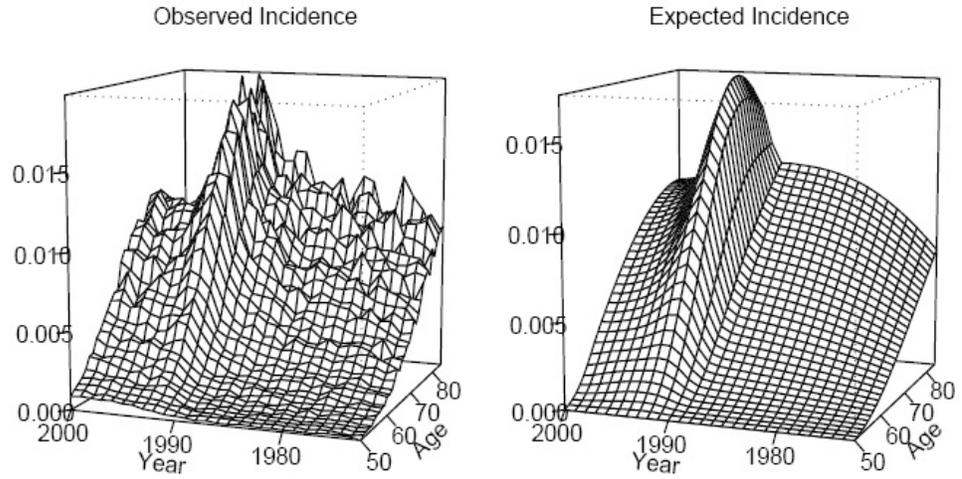


[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

# FIGURE\_4



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

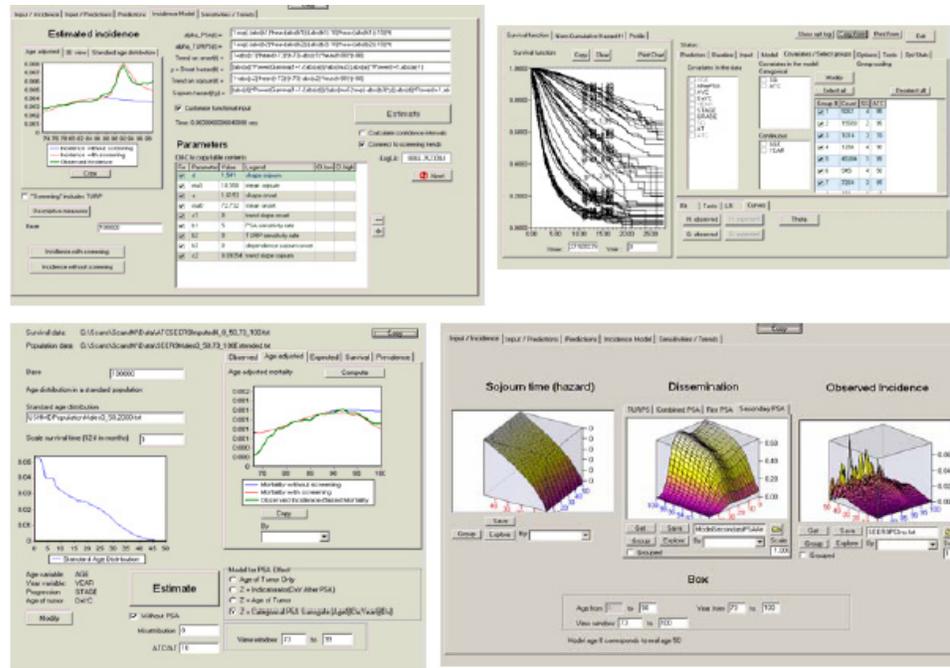


**Figure\_4.** Prostate cancer incidence (rate per person). Observed (left): Empirical estimate of prostate cancer incidence computed by dividing incident cancer cases from the SEER database by the population, for each age and calendar year. Expected (right): Model-predicted prostate cancer incidence by age and calendar year.



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

# FIGURE\_7



Figure\_7. Screen shots of the prototype software package implementing the model.



University of Michigan  
Equation Post Y

# EQUATION POST Y

$$Y | \{a, t, d(t)\} \propto f_O(y|a, t, d(t)) = \frac{f_I(a|x, y)f_O(y|x)}{f_I(a|x)},$$

where  $x = t - a$  is the year of birth.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
Aim Develop Unbiased Assessment Of  
Treatment Effects

# AIM DEVELOP UNBIASED ASSESSMENT OF TREATMENT EFFECTS



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)

To develop unbiased assessment of treatment effects from population data.



University of Michigan  
Aim Study The Joint Effect Of Progress

# AIM STUDY THE JOINT EFFECT OF PROGRESS

To study the joint effect of progress in treatment of prostate cancer and PSA utilization on observed national incidence and mortality trends.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
Aim Make Short And Long Term  
Predictions

# AIM MAKE SHORT AND LONG TERM PREDICTIONS

To make short- and long-term predictions of the trends in national incidence and mortality under various scenarios of projected behavior of key determinants of population processes.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
Aim Determine Evaluate Optimal  
Screening Strategies

# AIM DETERMINE EVALUATE OPTIMAL SCREENING STRATEGIES

To determine and evaluate optimal screening strategies and predict their effect on future national trends in prostate cancer incidence and mortality.



[Readers Guide](#)  
[Model Overview](#)  
[Assumption Overview](#)  
[Parameter Overview](#)  
[Component Overview](#)  
[Output Overview](#)  
[Results Overview](#)  
[Key References](#)



University of Michigan  
NHIS

# NHIS

## REFERENCE

National Center for Health Statistics (2004), "National Health Interview Survey (NHIS)."



Readers Guide  
Model Overview  
Assumption Overview  
Parameter Overview  
Component Overview  
Output Overview  
Results Overview  
Key References

## URL:

<http://www.cdc.gov/nchs/nhis.htm>

## NOTES AND DISCUSSION

CategoryReferences



# KEY REFERENCES

- Bolla, M., Gonzalez, D., Warde, P.** (1997) Improved survival in patients with locally advanced prostate cancer treated with radiotherapy and goserelin in *The New England Journal of Medicine* 337 : , p 295–300
- Breslow, N. E., Lubin, J. H. , Marek, P., Langholz, B.** (1983) Multiplicative models and cohort analysis in *Journal of the American Statistical Association* 78 : 381 , p 1-12
- Broet, P., , Tsodikov, A., , De Rycke, Y., , Moreau, T.** (2004) Two-sample statistics for testing the equality of survival functions against improper semi-parametric accelerated failure time alternatives: An application to the analysis of a clinical trial in *Lifetime Data Analysis* 10 : , p 103-120
- Burr, D.** (1994) On inconsistency of Breslow's estimator as an estimator of the hazard rate in the cox model in *Biometrics* 50 : 4 , p 1142-1145
- De Vita, V.T. Jr. , Hellman, S., Rosenberg, S.A.** (1997) Cancer: Principles and Practice of Oncology
- Draisma, G., Boer, R., Otto, S.J., van der Crujisen, I.W., Damhuis, R.A.M., Schröder, F.H., de Koning, H.J.** (2003) Lead times and overdetection due to prostate-specific antigen screening: Estimates from the European randomized study of screening for prostate cancer in *Journal of the National Cancer Institute* 95 : 12
- Etzioni, R., Cha, R., Feuer, E. J., Davidov, O.,** (1998) Asymptomatic incidence and duration of prostate cancer in *American Journal of Epidemiology* 148 : , p 775-785
- Etzioni, R., Legler, J.M., Feuer, E.J., Merrill, R.M., Chronin, K.A., Hankey, B.F.** (1999) Cancer surveillance series: Interpreting trends in prostate cancer - Part III: Quantifying the link between population prostate-specific antigen testing and recent declines in prostate cancer mortality in *Journal of the National Cancer Institute* 91 : , p 1033–1039
- Feuer, E.J., Merrill, R.M., Hankey, B.F.** (1999) Cancer surveillance series: Interpreting trends in prostate cancer, part ii: Cause of death misclassification and the recent rise and fall in prostate cancer mortality in *Journal of the National Cancer Institute* 91 : , p 1025–1032
- Gamel , J.W., Vogel, R.L.** (2001) Non-parametric comparison of relative versus cause-specific survival in Surveillance, Epidemiology and End Results (SEER) programme breast cancer patients in *Statistical Methods in Medical Research* 10 : , p 339-352
- Hanks, G.E., Pajak, T.F., Porter, A., Grignon, D., Brereton, H., Venkatesan, V., Horwitz, C., Lawton, E.M., Rosenthal, S.A., Sandler, H.M., Shipley, W.U. ,** (2003) Phase iii trial of long-term adjuvant androgen deprivation after neoadjuvant hormonal cytorreduction and radiotherapy in locally advanced carcinoma of the prostate in *Journal of Clinical Oncology*, 21 : , p 3972–3978
- Himmelblau, D.M.** (1972) Applied Nonlinear Programming
- Hoffman, R.M., Stone, S.N., Hunt, W.C., Key, C.R., Gilliland, F.D.** (2003) Effects of misattribution in assigning cause of death on prostate cancer mortality rates in *Annals of Epidemiology* 13 : , p 450-454
- Holmberg, L.H., Bill-Axelson, A., Helgesen, F., Salo, J.O., Folmerz, P., Haggman, M., Andersson, S.-O., Spangberg, A., Busch, C., Nordling, S., Palmgren, J., Adami, H.-O., Johansson, J-E., Norlen, B.J.** (2002) A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer in *The New England Journal of Medicine* 347 : , p 741-789
- Hougaard, P.** (1996) Frailty models for survival data in *Lifetime Data Analysis* 1 : , p 255–274

- Hsing, A.W., Devesa, S.S.** (2001) Trends and patterns of prostate cancer: What do they suggest? in *Epidemiologic Reviews* 23 : , p 3-13
- Hu, P., Zelen, M.** (1997) Planning clinical trials to evaluate early detection programmes in *Biometrika* 84 : , p 817-829
- Lawless, J.F.** (1987) Regression methods for Poisson process data in *Journal of the American Statistical Association* 82 : , p 808-815
- Lawton, C.A., Winter, K., Murray, K., Machtay, M., Mesic, J.B., Hanks, G.E., Coughlin, C.T., Pilepich, M.V.** (1998) Updated results of the phase iii radiation therapy oncology group (rtog) trial 85- 31 evaluating the potential benefit of androgen suppression following standard radiation therapy for unfavorable prog in *International Journal of Radiation Oncology, Biology and Physics* 93 : , p 937-46
- Lee, S.J., Zelen, M.** (2002) Statistical models for screening: planning public health programs in *Cancer Treat Res* 113 : , p 19-36
- Lee, S.J., Zelen, M.** (1998) Scheduling periodic examinations for the early detection of disease: applications to breast cancer in *Journal of the American Statistical Association* 93 : , p 1271-1281
- Merrill, R.M., Feuer, E. J., Warren, J. L. , Schussler, N. , Stephenson, R. A.** (1999) Role of transurethral resection of the prostate in population-based prostate cancer incidence rates in *American Journal of Epidemiology* 150 : 8 , p 848-860
- Merrill, R.M., Wiggins, C. L.,** (2002) Incidental detection of population-based prostate cancer incidence rates through transurethral resection of the prostate in *Urologic Oncology* 7 : , p 213-219
- Messing, E.M., Manola, J., Sarosdy, M., Wilding, G., Crawford, E.D., Trump, D.,** (1999) Immediate hormonal therapy compared with observation after radical prostatectomy and pelvic lymphadenectomy in men with node-positive prostate cancer in *The New England Journal of Medicine* 341 : , p 1781-1788
- National Cancer Institute** (2002) Surveillance Epidemiology and End Results (SEER) - Medicare linked database
- National Center for Health Statistics** (2004) National Health Interview Survey (NHIS).
- Optenberg, S. A., Thompson, I. M., Friedrichs, P., Wojcik, B., Stein, C. R., Kramer, B.** (1995) Race, treatment, and long-term survival from prostate cancer in an equal-access medical care delivery system in *Journal of the American Medical Association* 274 : , p 1599-1605
- Potosky, A.L., Feuer, E.J. , Levin, D.L.** (2001) Impact of screening on incidence and mortality of prostate cancer in the United States in *Epidemiological Review* 23 : 1 , p 181-186
- Roach III, M., Lu, J., Pilepich, M. V., Asbell, S. O., Mohiuddin, M., Grignon, D.** (2003) Race and survival of men treated for prostate cancer on radiation therapy oncology group phase iii randomized trials in *Journal of Urology* 169 : , p 245-250
- Roach, M., III, Krall, J., Keller, J. W., Perez, C. A., Sause, W. T., Doggett, R. L., Rotman, M., Russ, H., Pilepich, M. V., Asbell, S. O.** (1992) The prognostic significance of race and survival from prostate cancer based on patients irradiated on Radiation Therapy Oncology Group protocols (1976-1985) in *International Journal of Radiation Oncology Biology and Physics* 24 : , p 441-449
- Robbins, A. S., Whittemore, A. S., Van Den Eeden, S. K.** (1998) Race, prostate cancer survival, and membership in a large health maintenance organization in *Journal of the National Cancer Institute* 90 : , p 986-990
- Sasieni, P.D.** (1996) Proportional excess hazard in *Biometrika* 83 : , p 127-141

- Scardino P.T.** (1989) Early detection of prostate cancer in *Urol Clin North Am* 16 : , p 635-655
- Shen, Y., Zelen. M.** (1999) Parametric estimation procedures for screening programmes: Stable and nonstable disease models for multimodality case finding in *Biometrika* 86 : , p 503–515
- Stewart, S.L., King, J.B., Thompson, T.D., Friedman, C., Wingo, P.A.** (2004) Cancer mortality surveillance--United States, 1990-2000 in *MMWR Surveillance Summary* 53 : 3 , p 1-108
- Thompson, I., Tangen, C., Tolcher, A., Crawford, E., Eisenberger, M., Moinpour, C.** (2001) Association of african-american ethnic background with survival in men with metastatic prostate cancer in *Journal of the National Cancer Institute* 93 : , p 219–225
- Thompson, I.M., Goodman, P.J., Tangen, C.M., Lucia, M.S., Miller, G.J., Ford, L.G., Lieber, M.M., Cespedes, R.D., Atkins, J.M., Lippman, S.M., Carlin, S.M., Ryan, A., Szczepanek, C.M., Crowley, J.J., Coltman, C.A.** (2003) The influence of finasteride on the development of prostate cancer in *The New England Journal of Medicine* 349 : , p 215–224
- Tsodikov, A., Garibotti, G.** (2007) Profile information matrix for nonlinear transformation models in *Lifetime Data Analysis* 13 : 1 , p 139-159
- Tsodikov, A., Ibrahim, J.G., Yakovlev, A.Y.** (2003) Estimating cure rates from survival data: An alternative to two-component mixture models in *Journal of the American Statistical Association* 98 : , p 1063–1078
- Tsodikov, A.** (2002) Semiparametric models of long- and short-term survival: An application to the analysis of breast cancer survival in utah by age and stage in *Statistics in Medicine* 21 : , p 895–920
- Tsodikov, A.** (2006) Using composition to build semiparametric survival models in *Statistical Modelling* Submitted :
- Tsodikov, A., Szabo, A., Wegelin, J.** (2006) A population model of prostate cancer incidence in *Statistics in Medicine* 25 : 16 , p 2846-2866
- Tsodikov, Alex** (2003) Semiparametric models: a generalized self-consistency approach in *Journal of the Royal Statistical Society, Series B* 65 : , p 759–774
- Tsodikov, Alexander** (1998) A Proportional Hazards Model Taking Account of Long-Term Survivors in *Biometrics* 54 : 4 , p 1508-1516
- Weller, E.A., Feuer, E.J., Frey, C.M., Wesley, M.N.** (48) Parametric relative survival regression using generalized linear models with application to Hodgkin's lymphoma in *Applied Statistics* 1999 : , p 79–89
- Wilmoth, John R. (Director), Shkolnikov, Vladimir Shkolnikov (Co-Director),** (2003) Human Mortality Database (HMD).
- Wood, S.** (2003) Thin plate regression splines in *Journal of the Royal Statistical Society Series B* 65 : , p 95–114
- Wolf, S.H.** (1995) Screening for prostate cancer with prostate-specific antigen. in *he New England Journal of Medicine* T333 : , p 1401–1405
- Zelen, M., Feinleib, M.** (1969) On the theory of screening for chronic diseases in *Biometrika* 56 : , p 601-614

---

[Core Docs](#)