200 Years 1810 – 2010

Karolinska Institutet

ecpc
e-Science for Cancer Prevention and Control

Prediction and simulation

Mark Clements

NIASC Investigator

SeRC
Swedish e-Science Research Centre

Nordic Center of Excellence in Health-Related e-Sciences

Cancerfonden

Outline

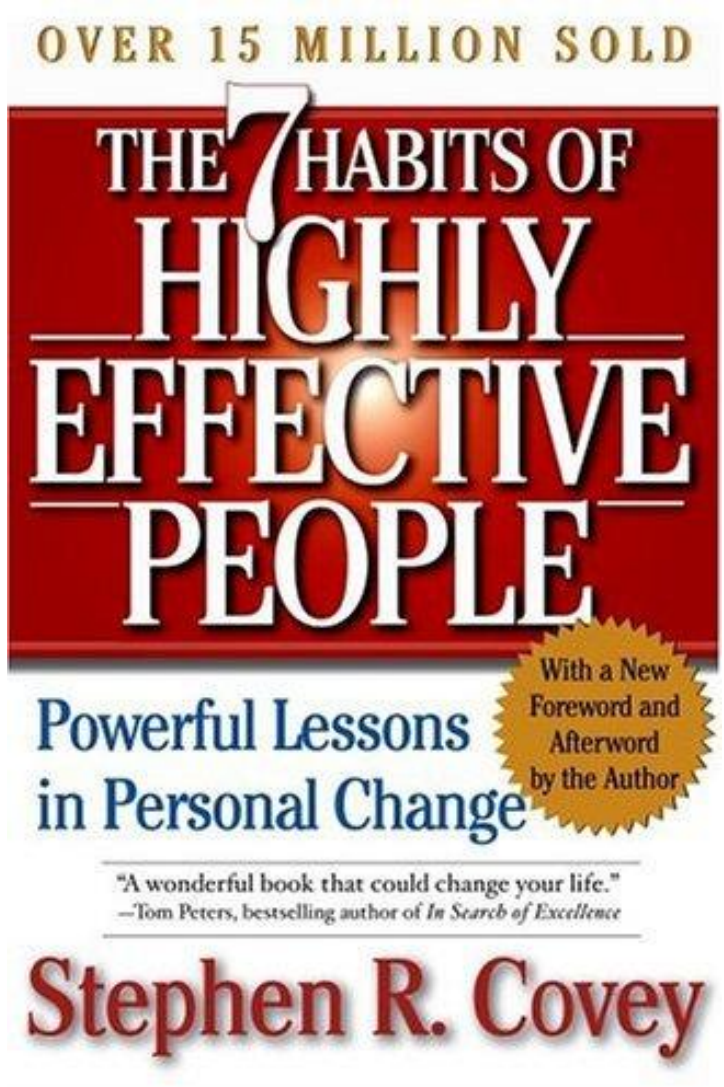
- Context: Prostate cancer testing
- Prediction
  - Review
  - Issues related to biomarker development for cancer screening
- Simulation
  - Review
  - Applications related to cancer testing/screening
- Discussion

Context: Who am I?

I am a biostatistician with an interest in modelling the cost-effectiveness of prostate cancer testing/screening. I am not:

- An expert in machine learning
- A health economist
- A *bona fide* epidemiologist
- A *bona fide* mathematical modeller

Covey's Second Habit: "Begin with the end in mind"



Context: Prostate cancer testing

- From a population health perspective, arguably, prostate cancer testing using the prostate-specific antigen (PSA) test is doing more harm than good and we could discourage its use
- From a health services perspective, prostate cancer testing has led to ballooning costs
- From a clinical perspective, a clinician wants to help their patient avoid metastatic prostate cancer
- From a health industry perspective, prostate cancer testing is good business

Context: Prostate cancer testing

- From a population health perspective, arguably, prostate cancer testing using the prostate-specific antigen (PSA) test is doing more harm than good and we could discourage its use
- From a health services perspective, prostate cancer testing has led to ballooning costs
- From a clinical perspective, a clinician wants to help their patient avoid metastatic prostate cancer
- From a health industry perspective, prostate cancer testing is good business
- Score: two for and two against

Context: We need to screen better for cancer

- New biomarkers are required to screen better for cancer
- How do we evaluate whether a new biomarker is cost-effective for cancer screening?
- Specifically, how can we evaluate whether a biomarker panel (with five blood-based biomarkers, a genetic risk score from 150 SNPs and self-reported family history) can lead to improved prostate cancer screening?

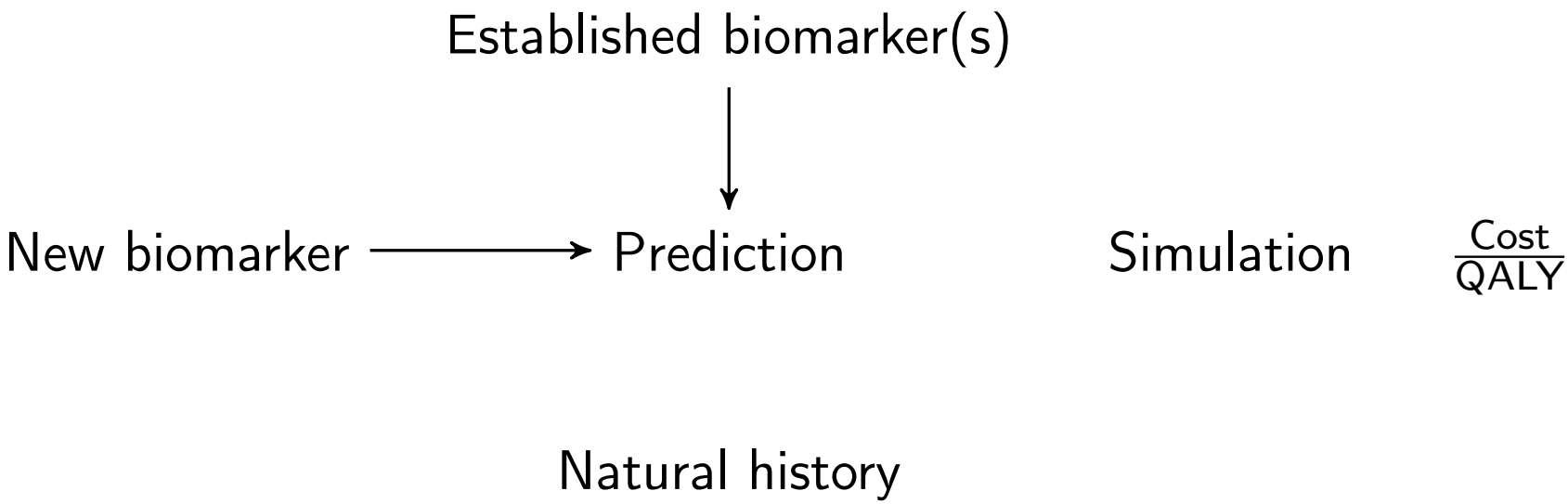
Conceptual overview

Established biomarker(s)

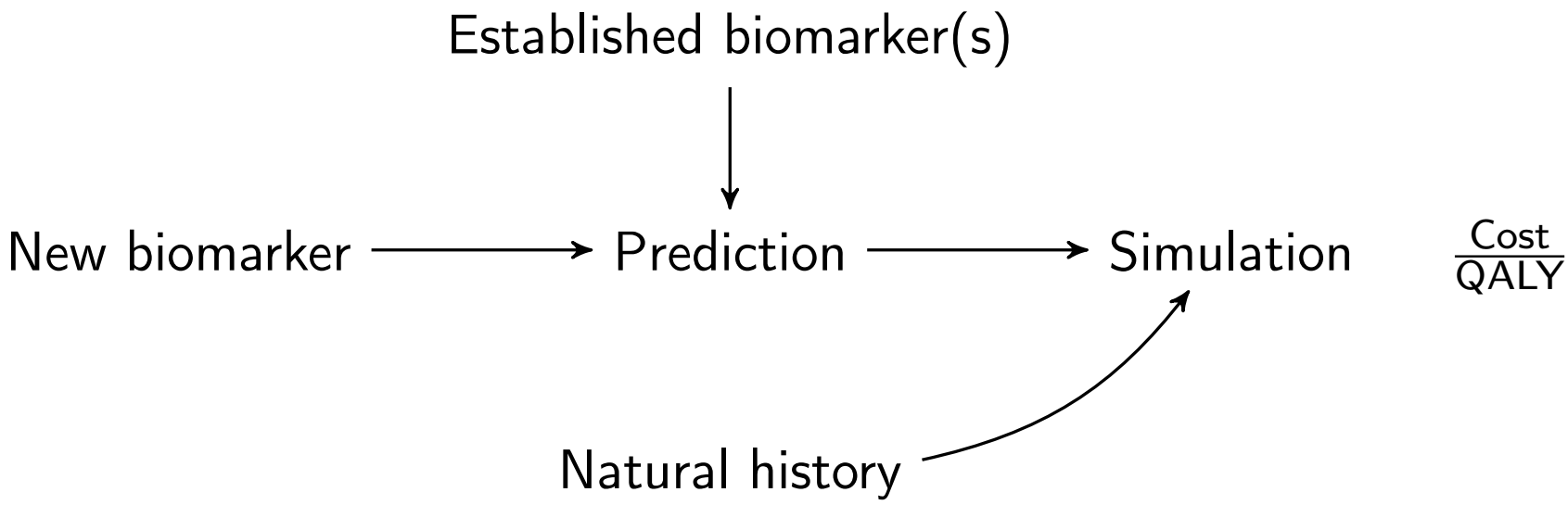
New biomarker | Prediction | Simulation | $\frac{\text{Cost}}{\text{QALY}}$

Natural history

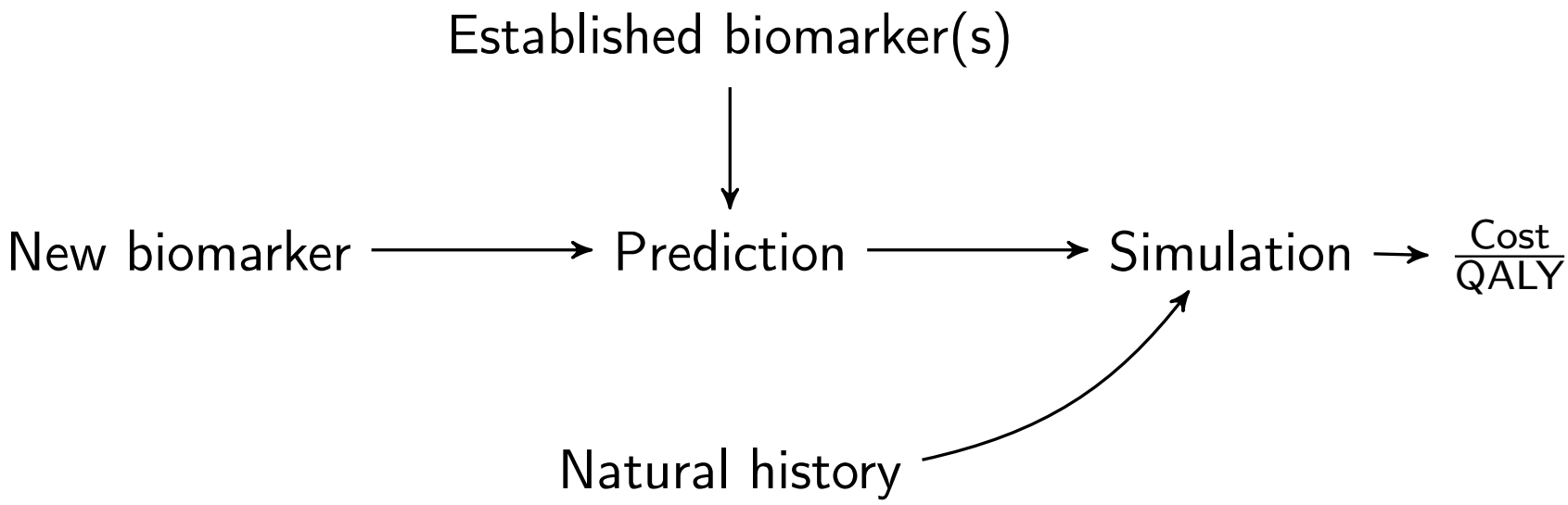
Conceptual overview



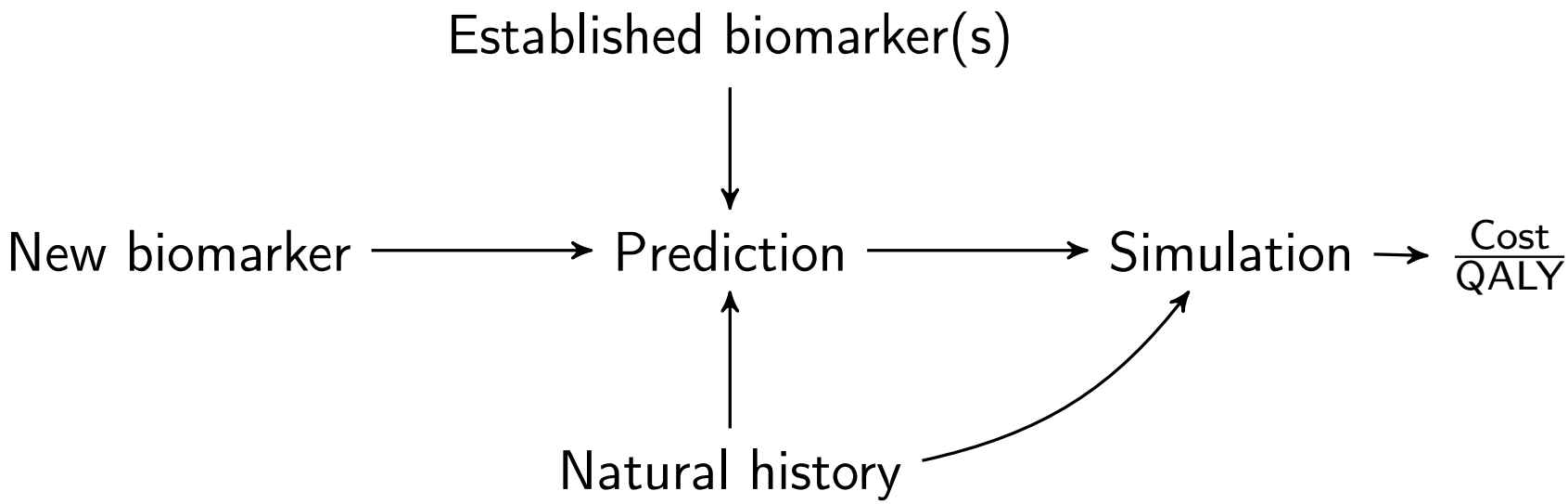
Conceptual overview



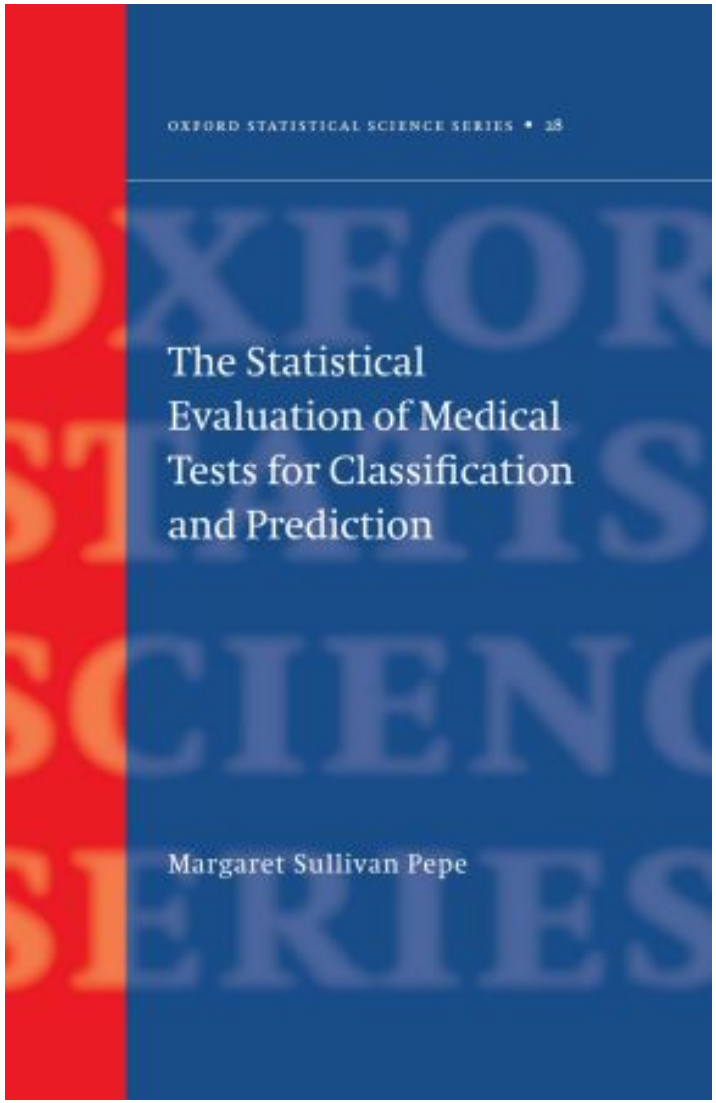
Conceptual overview



Conceptual overview



Prediction: Pepe (2003)



Prediction using new biomarkers: PROBE designs

COMMENTARY

Pivotal Evaluation of the Accuracy of a Biomarker Used for Classification or Prediction: Standards for Study Design

Margaret S. Pepe, Ziding Feng, Holly Janes, Patrick M. Bossuyt, John D. Potter

Research methods for biomarker evaluation lag behind those for evaluating therapeutic treatments. Although a phased approach to development of biomarkers exists and guidelines are available for reporting study results, a coherent and comprehensive set of guidelines for study design has not been delineated. We describe a nested case–control study design that involves prospective collection of specimens before outcome ascertainment from a study cohort that is relevant to the clinical application. The biomarker is assayed in a blinded fashion on specimens from randomly selected case patients and control subjects in the study cohort. We separately describe aspects of the design that relate to the clinical context, biomarker performance criteria, the biomarker test, and study size. The design can be applied to studies of biomarkers intended for use in disease diagnosis, screening, or prognosis. Common biases that pervade the biomarker research literature would be eliminated if these rigorous standards were followed.

J Natl Cancer Inst 2008;100:1432–1438

Prediction: Primer on test characteristics

"Most of epidemiology reduces to 2x2 tables"

		Disease status	
		+	−
Test status	+	a	b
	−	c	d

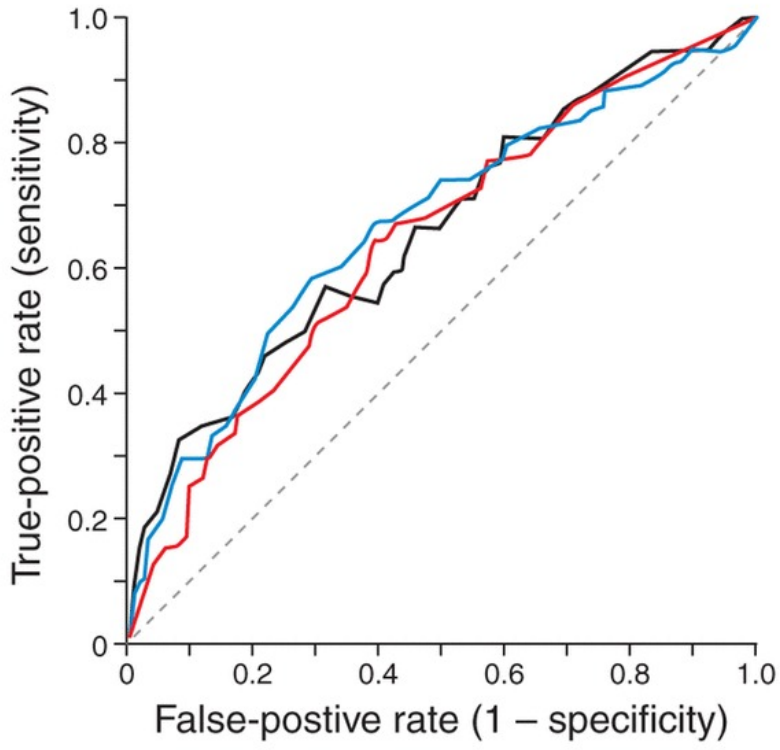
$$\text{Sensitivity} = \Pr(\text{Test } + \mid \text{Disease } +) = \frac{a}{a+c}$$

$$\text{Specificity} = \Pr(\text{Test } - \mid \text{Disease } -) = \frac{d}{b+d}$$

Continuous biomarkers: trade-off between Se and Sp

For a continuous biomarker, we can consider a threshold τ with sensitivity $Se(\tau) = \Pr(\text{Biomarker} > \tau | \text{Disease} +)$ and specificity $Sp(\tau) = \Pr(\text{Biomarker} < \tau | \text{Disease} -)$.

We can then look at the trade-off between sensitivity and specificity using the receiver operator characteristic (ROC) curve defined by $(1 - Sp(\tau), Se(\tau))$.



Prediction: Discrimination and calibration

- For discrimination, we are interested in whether a given predictor is good at distinguishing between individuals with and without the disease
- In the past, this has most often been measured using the area under the curve (AUC) for the ROC curve.

Prediction: Discrimination and calibration

- For discrimination, we are interested in whether a given predictor is good at distinguishing between individuals with and without the disease
- In the past, this has most often been measured using the area under the curve (AUC) for the ROC curve.
- Other measures have recently received support (e.g. NRI)

Prediction: Discrimination and calibration

- For discrimination, we are interested in whether a given predictor is good at distinguishing between individuals with and without the disease
- In the past, this has most often been measured using the area under the curve (AUC) for the ROC curve.
- Other measures have recently received support (e.g. NRI)
- For calibration, we are interested in whether a given risk prediction is unbiased: if the 10-year predicted risk of the disease is x%, then is that similar to the risk observed in out-of-sample data?

Prediction: Technical issues for cancer screening

- Specificity of a cancer screening test must be high. We are most interested in the LHS of the ROC curve (partial AUC)

Prediction: Technical issues for cancer screening

- Specificity of a cancer screening test must be high. We are most interested in the LHS of the ROC curve (partial AUC)
- We can combine information on different predictors using regression models, such as logistic regression and a range of machine learning algorithms

Prediction: Technical issues for cancer screening

- Specificity of a cancer screening test must be high. We are most interested in the LHS of the ROC curve (partial AUC)
- We can combine information on different predictors using regression models, such as logistic regression and a range of machine learning algorithms
- If we over-fit our models, then we can get overly optimistic AUCs. Use cross-validation, the bootstrap or external validation

Prediction: Technical issues for cancer screening

- Specificity of a cancer screening test must be high. We are most interested in the LHS of the ROC curve (partial AUC)
- We can combine information on different predictors using regression models, such as logistic regression and a range of machine learning algorithms
- If we over-fit our models, then we can get overly optimistic AUCs. Use cross-validation, the bootstrap or external validation
- Any new biomarker for should improve on the existing biomarkers (e.g. PSA for prostate cancer, CA-125 for ovarian cancer)

Prediction: Technical issues for cancer screening

- Specificity of a cancer screening test must be high. We are most interested in the LHS of the ROC curve (partial AUC)
- We can combine information on different predictors using regression models, such as logistic regression and a range of machine learning algorithms
- If we over-fit our models, then we can get overly optimistic AUCs. Use cross-validation, the bootstrap or external validation
- Any new biomarker for should improve on the existing biomarkers (e.g. PSA for prostate cancer, CA-125 for ovarian cancer)
- However, new biomarkers are likely to be evaluated on men who have been screened using the existing biomarkers

Prediction: Interpretation of screen-positive designs

- A screen-positive design is a study where the disease status is only known for those that are positive for a screening test
- This is a common design for cancer screening. For example, in prostate cancer testing men are often referred to biopsy on the basis of their PSA test value
- The interpretation of the AUC and apparent sensitivity/specificity becomes awkward as there is selection on those who have the reference test (e.g. biopsy)
- Remarkably, with these designs we can compare biomarkers using the relative sensitivity and relative false positive fraction (=1–specificity)

Prediction: Assessment of genetic biomarkers for cancer screening

- Sophisticated algorithms are available to determining which SNPs are associated with an outcome (e.g. lasso, elastic net)
- However, it is common to use a genetic risk score based on the sum of the number of risk alleles times the significant univariate log odds ratios
- The bad news is that, after adjustment for existing biomarkers, the genetic risk score often add little to the AUC:-(

Simulation in cancer screening: Calculation of cost-effectiveness

- For decision-making, health economists would like to compare the average lifetime costs and lifetime utilities for different screening scenarios
- Lifetime utilities are a generalisation of life-expectancy, where an individual in full health has a current utility of one, a person who has some disutility has a utility between zero and one, and a person who has died has a utility of zero
- Health economists are very useful at calculating costs and utilities — but they sometimes need help with modelling the natural history for a disease
- The lifetime calculations of cost-effectiveness can be complex if there is a complex natural history model

Simulation: Taxonomy

There are several model classes that we can use to calculate cost-effectiveness:

- Decision-tree analysis
- Loss of expected life calculations
- Markov models
- Microsimulation models

Simulation: Taxonomy

There are several model classes that we can use to calculate cost-effectiveness:

- Decision-tree analysis
- Loss of expected life calculations
- Markov models
- Microsimulation models

These models can be further classified by:

- Discrete versus continuous time
- Deterministic versus stochastic models

Simulation: Taxonomy

There are several model classes that we can use to calculate cost-effectiveness:

- Decision-tree analysis
- Loss of expected life calculations
- Markov models
- Microsimulation models

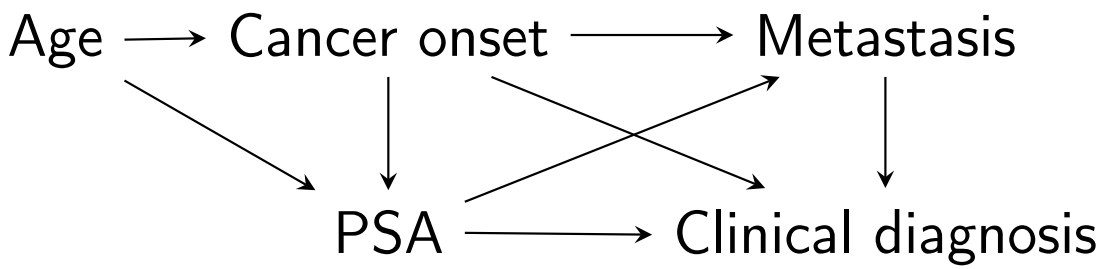
These models can be further classified by:

- Discrete versus continuous time
- Deterministic versus stochastic models

Simulation example: Prostate cancer testing

- We have extended an existing prostate cancer simulation model, initially developed by colleagues at the Fred Hutchinson Cancer Research Center (FHCRC)
- The simulations will be used to assess the cost-effectiveness of (i) different testing/screening scenarios, (ii) informal and formal compliance/costs and (iii) different types of tests
- Costs comes from the Swedish Institute for Health Economics
- Utilities come from a review from the Netherlands

FHCRC natural history model

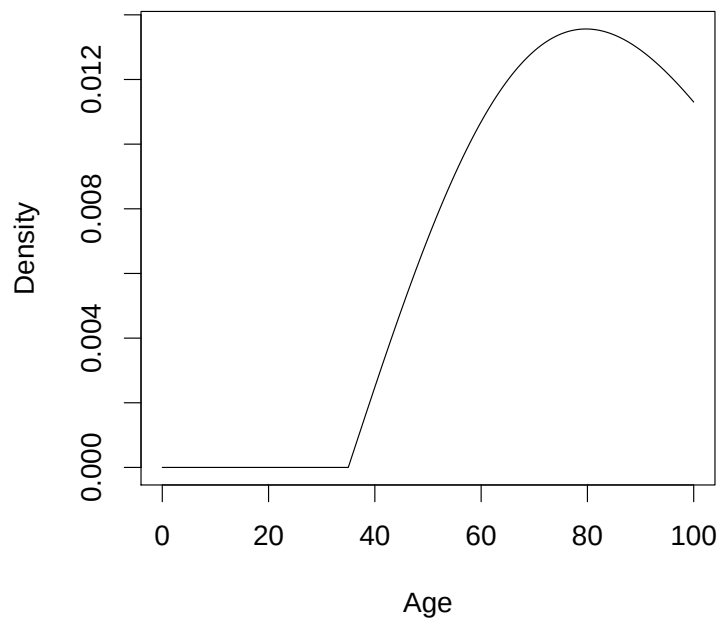


FHCRC: cancer onset

The hazard for the onset of cancer follows:

$$h_o(t) = \gamma_o t$$

where $t = \text{age} - 35$ and γ_o is a fixed parameter ($\hat{\gamma}_o = 0.0005$). The probability of no cancer onset is $S_o(t) = \exp\left(-\frac{\gamma_o}{2}t^2\right)$.



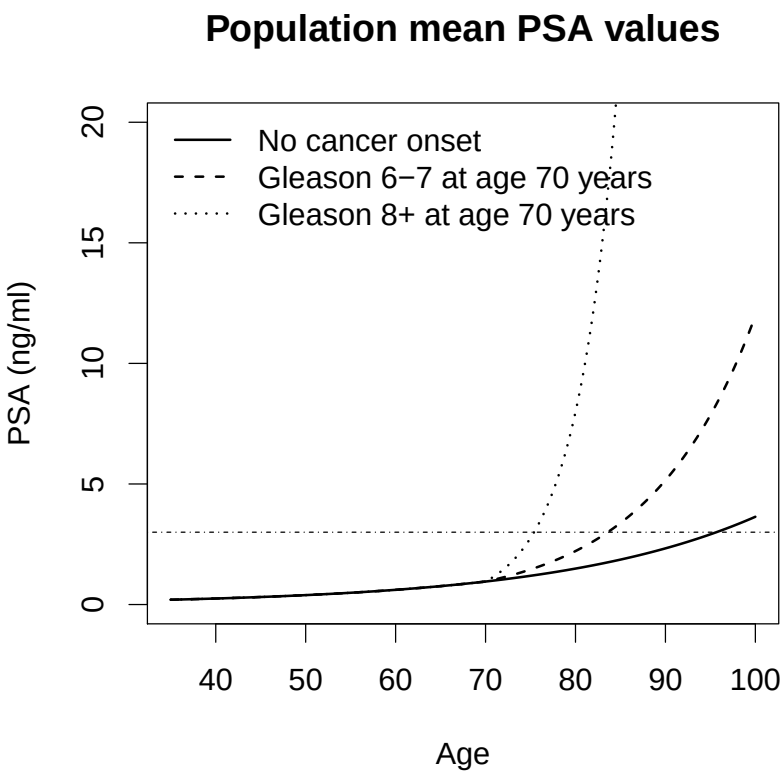
FHCRC: PSA

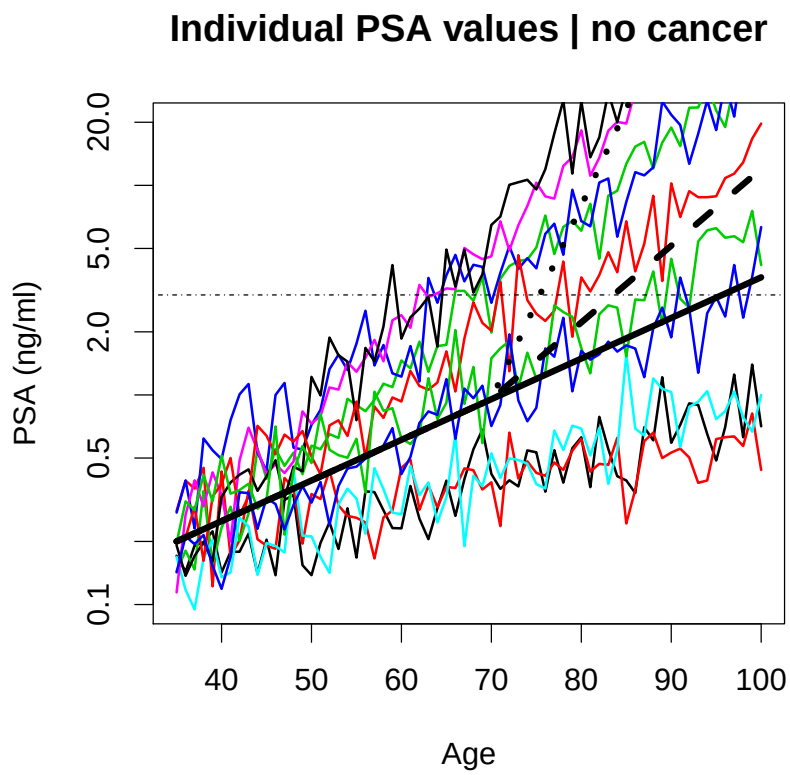
The longitudinal model for PSA is

$$\log(\text{PSA}) = \beta_0 + \beta_1 t + \beta_2 [t - t_o]_+ + \epsilon$$

where $[x]_+ = xI(x > 0)$ and t_o is the time of cancer onset. The parameters β_0, β_1 and β_2 are assumed to be random effects, and ϵ is measurement error.

FHCRC PSA model





Borrowing from ovarian cancer screening: Can we use repeated PSA values to predict the risk of prostate cancer?

Screening Based on the Risk of Cancer Calculation From Bayesian Hierarchical Changepoint and Mixture Models of Longitudinal Markers

Steven J. SKATES, Donna K. PAULER, and Ian J. JACOBS

The standard approach to early detection of disease with a quantitative marker is to set a population-based fixed reference level for making further individual screening or referral decisions. For many types of disease, including prostate and ovarian cancer, additional information is contained in the subject-specific temporal behavior of the marker, which exhibits a characteristic alteration early in the course of the disease. In this article we derive a Bayesian approach to screening based on calculation of the posterior probability of disease given longitudinal marker levels. The method is motivated by a randomized ovarian cancer screening trial in the United Kingdom comprising 22,000 women screened over 4 years with an additional 5 years of follow-up on average. Levels of the antigen CA125 were recorded annually in the screened arm. CA125 profiles of cases and controls from the U.K. trial are modeled using hierarchical changepoint and mixture models, posterior distributions are calculated using Markov chain Monte Carlo methods, and the model is used to calculate the Bayesian posterior risk of having ovarian cancer given a new subject's single or multiple longitudinal CA125 levels. A screening strategy based on the risk calculation is then evaluated using data from an independent screening trial of 5,550 women performed in Sweden. A longitudinal CA125 screening strategy based on calculation of the risk of ovarian cancer is proposed. Simulations of a prospective trial using a strategy based on the risk calculated from longitudinal CA125 values indicate potentially large increases in sensitivity for a given specificity compared to the standard approach based on a fixed CA125 reference level for all subjects.

KEY WORDS: Longitudinal CA125; Markov chain Monte Carlo; Mixtures; Ovarian cancer; Screening.

Bayes' theorem for predicting prostate cancer

For a biomarker Y and disease outcome D,

$$\Pr(D|Y=y) = \frac{\Pr(Y=y|D)\Pr(D)}{\Pr(Y=y|D)\Pr(D) + \Pr(Y=y|\bar{D})\Pr(\bar{D})}$$

Note that $\Pr(D) = \Pr(t > t_o)$ and, currently, Y is measured by log(PSA).

This calculation could be extended to include a man's PSA test history.

Extending the natural history model to incorporate multiple biomarkers

- PSA is a measure of time since cancer onset

Extending the natural history model to incorporate multiple biomarkers

- PSA is a measure of time since cancer onset
- SNPs are time-invariant, affecting either susceptibility or the rate of onset (cf. time since onset)

Extending the natural history model to incorporate multiple biomarkers

- PSA is a measure of time since cancer onset
- SNPs are time-invariant, affecting either susceptibility or the rate of onset (cf. time since onset)
- Different biomarkers will have varying functional relationships — many of which may be poorly characterised

Extending the natural history model to incorporate multiple biomarkers

- PSA is a measure of time since cancer onset
- SNPs are time-invariant, affecting either susceptibility or the rate of onset (cf. time since onset)
- Different biomarkers will have varying functional relationships — many of which may be poorly characterised
- As a simplification, we need only characterise the functional relationship between cancer and the biomarker panel

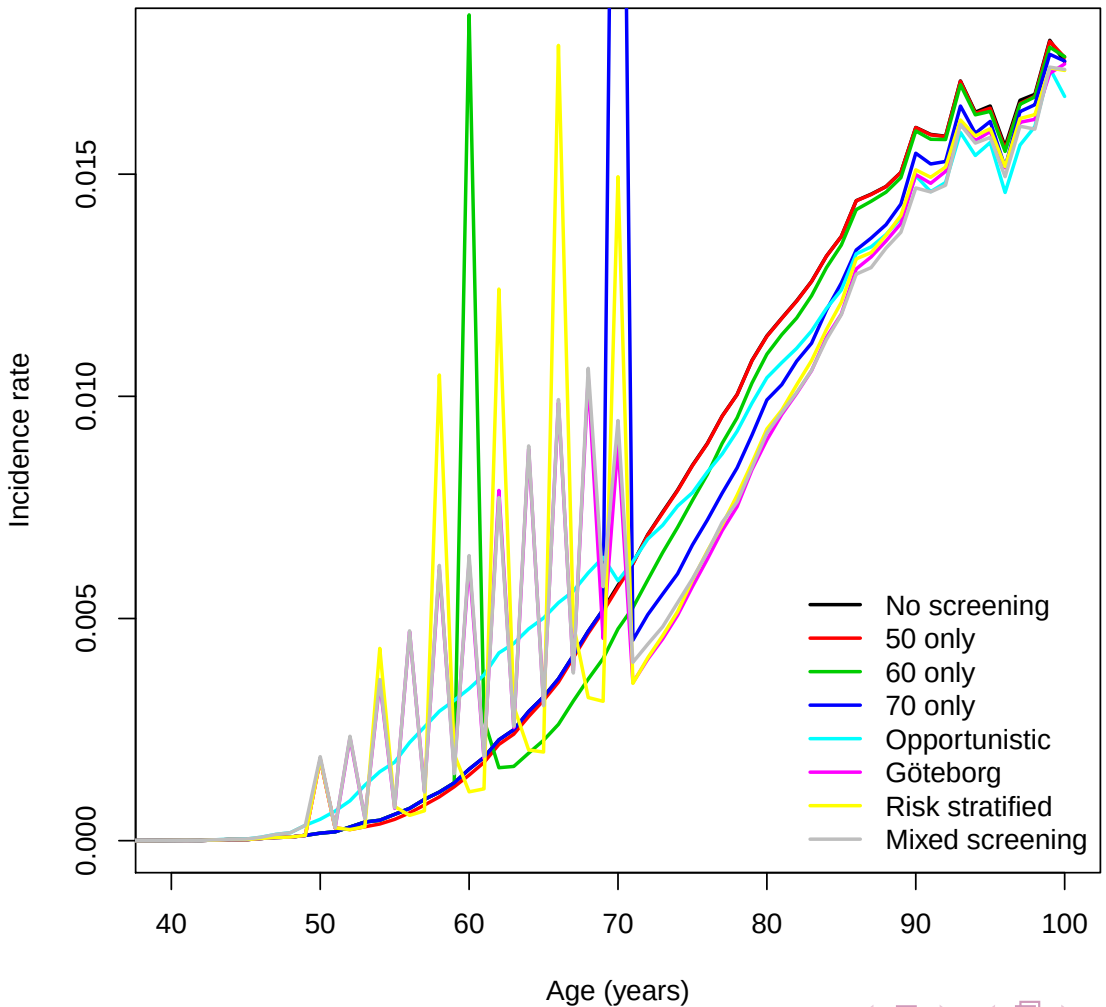
Extending the natural history model to incorporate multiple biomarkers

- PSA is a measure of time since cancer onset
- SNPs are time-invariant, affecting either susceptibility or the rate of onset (cf. time since onset)
- Different biomarkers will have varying functional relationships — many of which may be poorly characterised
- As a simplification, we need only characterise the functional relationship between cancer and the biomarker panel
- Complication: we usually have little or no information on longer-term outcomes for the new biomarkers

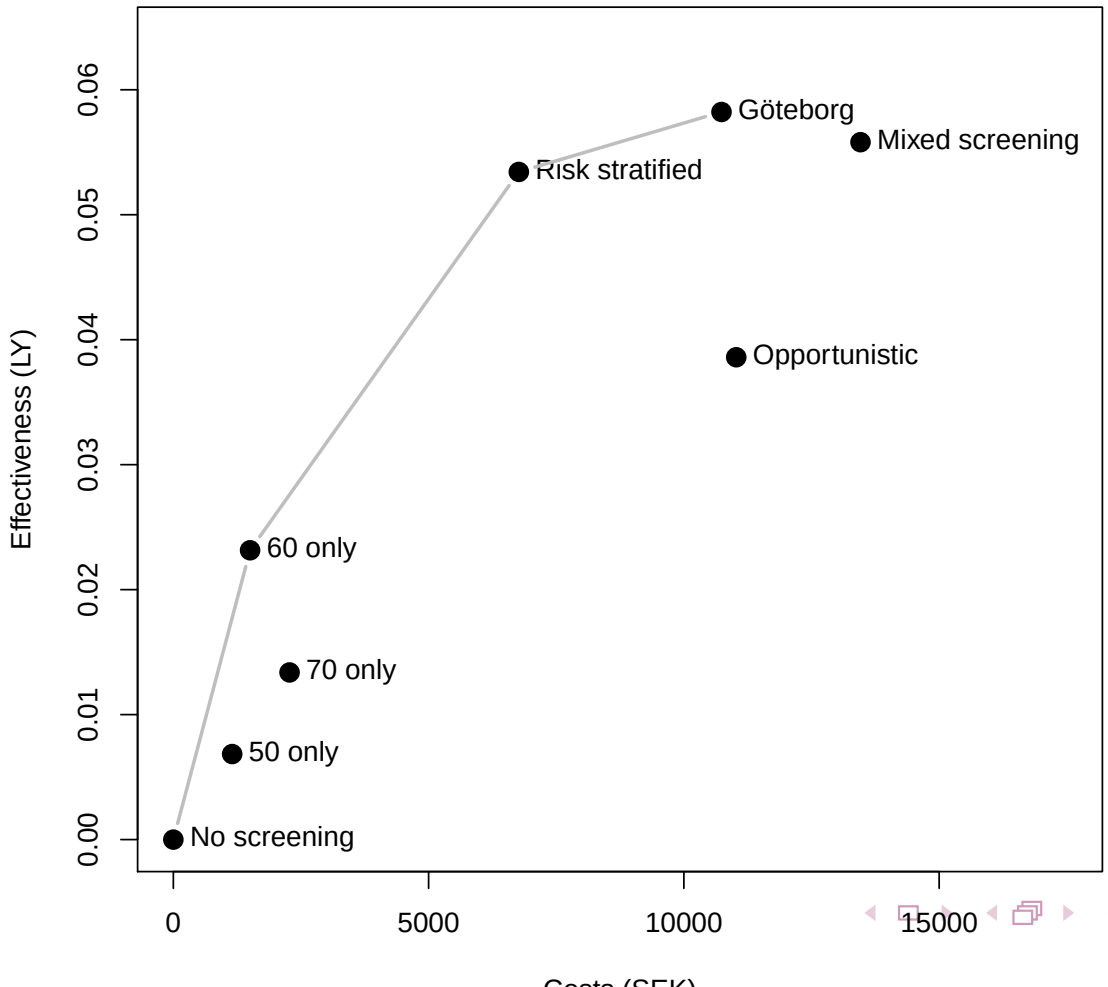
Implementation of the FHCRC simulation

- We have completed an open-source implementation of the FHCRC model. The R package uses various C++ libraries and scales well to multiple processors.
- See https://github.com/mclements/microsimulation/tree/develop.

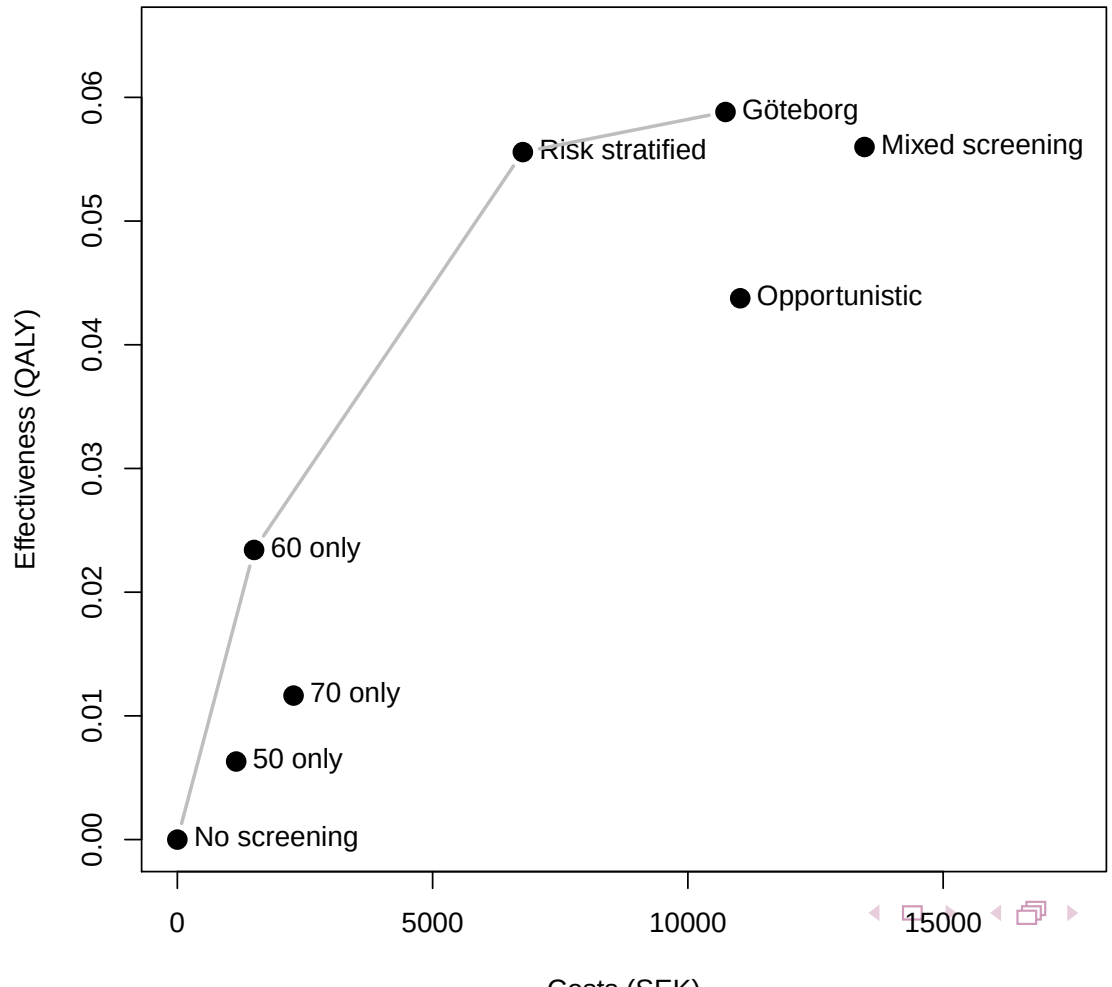
Example: Simulations for prostate cancer incidence under different testing/screening scenarios



Example: Cost-effectiveness under different testing/screening scenarios, costs versus life-years (discount=0%)



Example: Cost-effectiveness under different testing/screening scenarios, costs versus QALYs (discount=0%)



Returning to where we started: Can we test/screen better for prostate cancer?

- The STHLM3 diagnostic trial uses a paired, screen-positive design to assess whether a given biomarker panel has similar sensitivity and greater specificity for advanced prostate cancers

Returning to where we started: Can we test/screen better for prostate cancer?

- The STHLM3 diagnostic trial uses a paired, screen-positive design to assess whether a given biomarker panel has similar sensitivity and greater specificity for advanced prostate cancers
- To calculate cost-effectiveness for formal testing/screening using the biomarker panel, we will use the microsimulation model. This evidence is intended to support Stockholm County in deciding whether to introduce organised screening

Returning to where we started: Can we test/screen better for prostate cancer?

- The STHLM3 diagnostic trial uses a paired, screen-positive design to assess whether a given biomarker panel has similar sensitivity and greater specificity for advanced prostate cancers
- To calculate cost-effectiveness for formal testing/screening using the biomarker panel, we will use the microsimulation model. This evidence is intended to support Stockholm County in deciding whether to introduce organised screening
- The genetic risk score is a significant, yet moderately small, component of the biomarker panel for prostate cancer

Returning to where we started: Can we test/screen better for prostate cancer?

- The STHLM3 diagnostic trial uses a paired, screen-positive design to assess whether a given biomarker panel has similar sensitivity and greater specificity for advanced prostate cancers
- To calculate cost-effectiveness for formal testing/screening using the biomarker panel, we will use the microsimulation model. This evidence is intended to support Stockholm County in deciding whether to introduce organised screening
- The genetic risk score is a significant, yet moderately small, component of the biomarker panel for prostate cancer
- Given uncertainties in the longer-term outcomes, it is often challenging to assess the cost-effectiveness of a novel genetic marker on cancer screening