From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

Microsimulation modelling of prostate cancer screening in Sweden

Andreas Karlsson

Karolinska Institutet

Stockholm 2019

The front page illustrates men in a microsimulation reaching other-cause death before prostate cancer symptoms as a result of a screening intervention. The figures are produced by the author, if not otherwise stated.

Published by Karolinska Institutet.

ISBN 978-91-7831-440-9

Printed by Eprint AB 2019

Institutionen för Medicinsk Epidemiologi och Biostatistik

Microsimulation modelling of prostate cancer screening in Sweden

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i hörsal Petrén, Nobels väg 12 B, Karolinska Institutet, Solna

Måndagen den 13 Maj 2019, kl 13.00

av

Andreas Karlsson

Huvudhandledare:
Senior Lecturer Mark Clements
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Fakultetsopponent:
Dr Eveline Heijnsdijk
The Erasmus University Medical Center
Department of Public Health

Bihandledare:
Dr Alexandra Jauhiainen
AstraZeneca, Gothenburg
Biometrics, Early Clinical Development

Betygsnämnd:
Adjunct Professor Rosaria Galanti
Karolinska Institutet
Department of Public Health Sciences

Professor Erwin Laure
Royal Institute of Technology
Division of Computational Science and Technology

Associate Professor Niklas Zethraeus
Karolinska Institutet
Department of Learning, Informatics, Management and Ethics

Professor Henrik Grönberg
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics, and
Saint Göran Hospital
Department of Urology

Professor Tom Britton
Stockholm University
Department of Mathematics

Stockholm 2019

Everything should be made as simple as possible, but no simpler.

- simplified quote by Einstein

Abstract

Evidence-based health policy may require modelling of different interventions. The choice of model complexity is a trade-off, where simpler models may be easier to describe and calibrate, while complex models may better represent the disease dynamics and lead to more valid predictions. Modelling for health policy is inherently multidisciplinary, with relevant disciplines including epidemiology, clinical medicine, biostatistics, health economics and computer science.

As a motivation, we sought to assess the cost-effectiveness of a new prostate cancer screening test – STHLM3. The STHLM3 test uses a combination of biomarkers and self-reported data for prediction. The STHLM3 test can be used as a reflex test after a PSA test. To assess the cost-effectiveness, we needed a model that represents the natural history of prostate cancer. This model was then used to predict the short- and long-term effects of different prostate cancer testing interventions.

To achieve this, we developed a framework for event-oriented, discrete event simulation in R and C++ in **Study I**. The framework included common random numbers, which reduces the Monte Carlo error, and detailed in-simulation reporting for health economic evaluations. In **Study II**, we extended an older US prostate cancer model to better model for Gleason score. Model inputs included PSA testing, prostate cancer diagnosis, treatment, management and survival. The calibration of the natural history model included both screened and unscreened populations. We initially calibrated the Swedish "Prostata" model using maximum likelihood estimation with non-linear equality constraints. Subsequently in **Study IV**, we developed a method based on approximate Bayesian computations and Markov chain Monte Carlo methods. The hybrid method provided a more systematic approach to incorporate evidence at different scales while still using known likelihoods. For **Study III**, we further extended the calibrated model to include costs, health state values and discounting. We calculated the life-time expected costs and effectiveness under different test interventions. We found that the STHLM3 test was cost-effective in Sweden at a reflex PSA threshold of 2 ng/mL.

For broader conclusions, first, microsimulation is a challenging computational and scientific task, particularly for calibration and sensitivity analyses. Second, ironically, the depth of the Swedish health and population registers made it easier to invalidate complex models that had been well validated in other populations. The Swedish data can support efforts to improve existing models for cancer screening and, more broadly, other health interventions.

Strengths of our approach included: a flexible, lightweight, fast, scalable, open, microsimulation framework for health policy development; calibration of the natural history model to current incidence by Gleason grading and recent survival, whereas most other models have been calibrated to older, PSA-naïve populations; and incorporation of detailed data and estimates, making best use of the available Swedish health and population registers. Limitations of our approach include: imprecise estimates for the effect of prostate cancer testing on mortality; uncertainty in the validity of the natural history model for predictions outside of observed evidence; and uncertainty in the validity of the health state values. For future work, we plan to extend the Prostata model to include magnetic resonance imaging in combination with the newer prostate cancer screening tests.

List of scientific papers

These papers are referred to by their Roman numerals throughout, and are presented in full at the end of this thesis.

I. Andreas Karlsson, Niten Olofsson, Erwin Laure & Mark S. Clements.
A parallel microsimulation package for modelling cancer screening policies.
2016 IEEE 12th International Conference on e-Science, 323–330, 2016.

II. Andreas Karlsson, Alexandra Jauhiainen, Roman Gulati, Martin Eklund, Henrik Grönberg, Ruth Etzioni & Mark S. Clements.
A natural history model for planning prostate cancer testing: Calibration and validation using Swedish registry data.
PLOS One, **14**(2):0211918, 2019.

III. Andreas Karlsson, Shuang Hao, Alexandra Jauhiainen, K. Miriam Elfström, Lars Egevad, Tobias Nordström, Emelie Heintz & Mark S. Clements.
The cost-effectiveness of prostate cancer screening using the STHLM3 test.
Submitted

IV. Andreas Karlsson, Mark S. Clements & Alexandra Jauhiainen.
A hybrid ABC approach for calibrating microsimulation models.
Manuscript

The studies in this thesis were supported by the Swedish Research Council, Prostatacancerförbundet, Cancerfonden and the Swedish e-Science Research Centre.

Related paper

(not included in this thesis)

- Ola Spjuth, Andreas Karlsson, Mark Clements, Keith Humphreys, Emma Ivansson, Jim Dowling, Martin Eklund, Alexandra Jauhiainen, Kamila Czene, Henrik Grönberg, Pär Sparen, Fredrik Wiklund, Abbas Cheddad, Thorgerdur Palsdottir, Mattias Rantalainen, Linda Abrahamsson, Erwin Laure, Jan-Eric Litton & Juni Palmgren.
  E-Science technologies in a workflow for personalized medicine using cancer screening as a case study.
  J Am Med Inform Assoc, **24**(5):950–957, 2017

- Thorgerdur Palsdottir, Tobias Nordström, Andreas Karlsson, Henrik Grönberg, Mark Clements & Martin Eklund.
  The impact of different prostate specific antigen (PSA) screening intervals on Gleason score at diagnosis and the risk of experiencing false positive biopsy recommendations: A population based cohort study
  BMJ Open, **9**(3):e027958, Mar 2019.

Contents

List of abbreviations

ABC approximate Bayesian computation

AUC area under the curve

CAP Cluster Randomized Trial of PSA Testing for Prostate Cancer

CI confidence interval

CISNET Cancer Intervention Surveillance Modeling Network

DES discrete-event simulation

ERSPC European Randomized study of Screening for Prostate Cancer

FHCRC Fred Hutchinson Cancer Research Center

GDPR General Data Protection Regulation

GPL GNU General Public License

IARC International Agency for Research on Cancer

ICER incremental cost-effectiveness ratio

IRR incidence rate ratio

ISPOR International Society of Pharmacoeconomics and Outcomes Research

LHS Latin hypercube sampling

MCI Monte Carlo Interval

MCMC Markov chain Monte Carlo

MISCAN Microsimulation Screening Analysis

ML maximum likelihood

MRR mortality rate ratio

NCI National Cancer Institute

NICE National Institute for Health and Care Excellence

NPCR National Quality Registry for Prostate Cancer

PCa prostate cancer

PCBaSe Prostate Cancer data Base Sweden

PCPT Prostate Cancer Prevention Trial

PLCO Prostate, Lung, Colorectal and Ovarian Study

PSA prostate-specific antigen

QALY quality-adjusted life-year

SBPR Stockholm PSA and Biopsy Register

SE standard error

SEER Surveillance, Epidemiology, and End Results

USPSTF United States Preventive Services Task Force

1 Introduction

Prostate cancer is a highly prevalent cancer in older men and the most common cause of cancer death among Swedish males [1]. Costs due to prostate cancer testing, diagnostics and treatments are substantial. The large European Randomized study of Screening for Prostate Cancer (ERSPC) provided evidence that screening using the prostate-specific antigen (PSA) test could reduce prostate cancer mortality by 20% over 13 years [2]. However, there is uncertainty whether any potential benefits from prostate cancer screening will outweigh the harms or whether screening will be cost-effective [3]. Due to these uncertainties, PSA testing is not systematically organised in Sweden.

1.1 Aims of the Thesis

The overall aim was to provide tools and methods to predict costs and effectiveness of cancer screening in general and prostate cancer screening in Sweden specifically. As an application, we aimed to investigate the cost-effectiveness of the STHLM3 prostate cancer screening test.

To achieve this the following studies were performed:

- For **Study I**, we aimed to describe the design choices in our parallel discrete-event `microsimulation` R package for planning cancer screening.
- For **Study II**, our aim was to adapt and extend an older US model of the natural history of prostate cancer for the Swedish setting. We calibrated and validated our model using detailed longitudinal Swedish registry data. The resulting model is available as an R package called `prostata`.
- In **Study III** we used the outcomes from **Study I** and **Study II** to investigate the cost-effectiveness of a novel screening test for prostate cancer.
- In **Study IV** we wanted to describe an alternative method for calibration, improving on the calibration that was done in **Study II**.

2 Background

2.1 Prostate cancer

Our research relates to the early detection of prostate cancer. The prostate gland is a male reproductive organ, located around the urethra. The prostate is, together with other glands, responsible for the production of semen. The risk of developing cancer in the prostate increases with a man's age.

Prostate cancer is the most common cancer diagnosis and the most common cause of cancer death in Sweden, responsible for 33% of male cancer diagnoses and 22% of male cancer deaths in 2018 [4]. Prostate cancer incidence rates in Sweden are now comparable to rates in countries that had an early introduction of PSA testing (see Figure 2.1), while prostate cancer mortality rates in Sweden are higher than in most other countries (see Figure 2.2) [5].

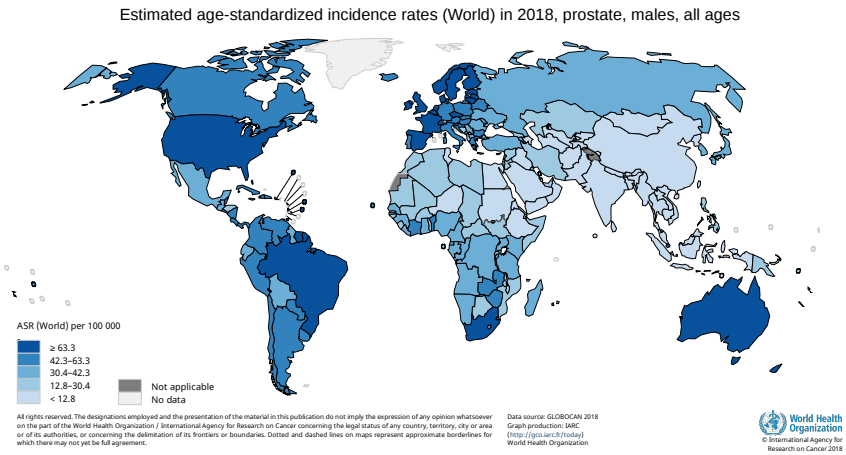

Figure 2.1: Age-standardised prostate cancer incidence per 100,000 man-years. *Source:* [4]

With over 100,000 prevalent cases in Sweden, the health burden and the costs on the health care system are substantial [1]. While a number of risk factors have been proposed for prostate cancer, including diet and occupational exposures, the only factors conclusively shown to increase risk of the disease are age, ethnicity and family history [6, 7]. Given the high prevalence of the cancer and limited opportunities for primary prevention, a good screening[1] test would reduce the health burden due to prostate cancer.

[1] We use "screening" in the general sense of testing in an asymptomatic population.

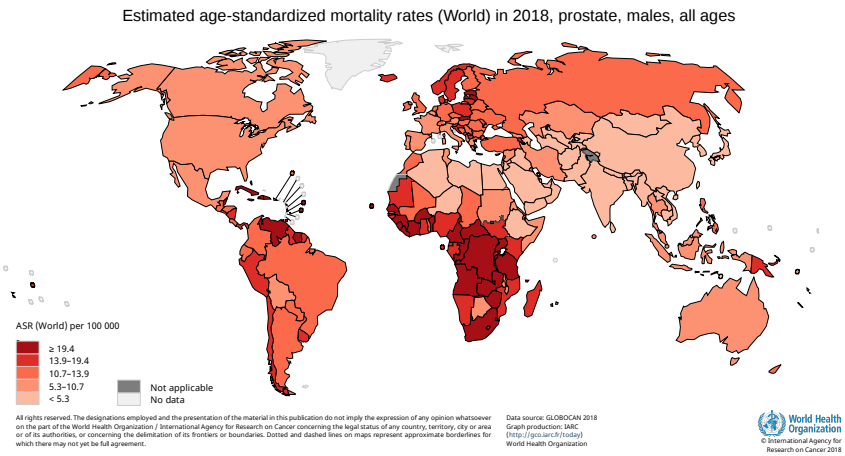

Figure 2.2: Age-standardised prostate cancer mortality rates per 100,000 man-years. *Source:* [4]

2.1.1 Gleason grade

The Gleason grading system is used by pathologists when assessing histological patterns in prostate cancer biopsies. The pathologist quantifies the cell morphology with a grade ranging from one to five. This is then summarised in a total score composed by the primary grade (i.e. the most common morphological pattern) and a secondary grade. If the primary grade also is the most severe grade, then the secondary grade is decided by the second most common pattern (ignoring patterns that occupy less than 5% of the tumour area). If there are patterns more severe than the primary grade, then the most severe pattern is used for the secondary grade independent of size. In 2005 the International Society of Urological Pathology changed the Gleason reporting such that the lowest reported cancer grade (with "extremely rare exception"), for needle biopsies was set to Gleason score 3+3 = 6. Patients with a prostate cancers with a Gleason score ≤ 6 usually have better prognosis than patients with a higher Gleason score [8, 9]. The 2005 change was done partly because of changing practices in reporting Gleason patterns, which had lead to a Gleason score inflation. Since then there have been further reports on continued Gleason inflation [10]. Changes in Gleason reporting, both through inflation and altered guidelines, pose a challenge for assessing the long-term prognosis by Gleason score.

2.1.2 Staging

The TNM system is used for classifying tumours. The *T* describes the size or extent of the tumour, *N* describes the involvement of lymph nodes and *M* if there are distant metastasis or not. For our model, M-staging and T-staging will be of most importance. The T-stages starts with T0, which indicates no evidence of prostate cancer, and ends with T4, which means that the tumour has invaded adjacent tissue [11, 12].

2.2 PSA screening

The PSA test was first used to monitor disease progression in prostate cancer patients, and was later taken up as the *de facto* screening test for prostate cancer in many countries, leading to rapid rises in prostate cancer incidence. The test characteristics for the PSA test in detecting prostate cancer are comparable to those for breast cancer mammography, with a sensitivity of 72% and a specificity of 93% when using a PSA threshold of 4 ng/ml [13]. However, the Swedish guidelines recommends a threshold of 3 ng/ml for men below age 70 years [12], which has led to an increased sensitivity and a reduced specificity [14]. Our group has recently reported that 50–75% of men aged 50 years and over living in Stockholm have had a PSA test within the last nine years [15].

There is uncertainty in how best to use the PSA test. The ERSPC trial indicated that PSA testing reduces prostate cancer mortality by around 20% after 13 years of follow-up [2]. This was, however, not observed in two other large randomised studies. The results from the Prostate, Lung, Colorectal and Ovarian Study (PLCO) Cancer Screening Trial in the United States suggest that organised PSA testing has little effect on prostate cancer mortality in a population where opportunistic testing has already reached high levels and where there is poor biopsy compliance [16] [1]. The recent Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) in the United Kingdom investigated the effects of a single screen. CAP found that, with 40% attendance at a single clinical visit, PSA testing lead to a non-significant estimated mortality reduction of 4% after ten years of follow-up [17].

PSA testing is common in most western countries, even though the balance between benefits and harms is uncertain [3, 18].

For the Swedish context, the Göteborg site of the ERSPC estimated a large effect size [19], however there is no statistical evidence for heterogeneity between the ERSPC sites (p=0.4) [2]. It is however noteworthy, that the mortality rates in the Nordic region are comparatively high [4], as described in the Prostate cancer section.

[1] Biopsy compliance is defined as having a biopsy following a positive PSA test.

2.3 Alternative screening tests

Efforts have been made to improve upon the PSA screening test. This has been suggested in the form of a reflex test, where a PSA test is first performed and depending on its result a secondary test may be performed.

The STHLM3 test is a new prostate cancer test based on five plasma protein biomarkers (including PSA, free PSA, intact PSA, hK2, MSMB and MIC1) and 232 single nucleotide polymorphisms (SNPs) together with clinical variables such as age, family, history, previous prostate biopsy and prostate examinations. The accuracy of the STHLM3 test has been reported in The Lancet Oncology [20]. The STHLM3 test reduced false positive biopsies by 42% and reduced diagnoses of small Gleason score 6 cancers by 17% whilst maintaining sensitivity for cancers with Gleason score 7 and higher, compared to using PSA $\geq$ 3ng/mL. Due to the paired screen-positive design of the study, the corresponding area under the curve (AUC)s can not be directly compared to AUCs from other risk panels.

Other prediction models for early detection of prostate cancer include the Four-kallikrein Panel (4K panel), the Prostate Health Index (PHI) and the Prostate Cancer Antigen 3 (PCA3) test. The 4K panel consists of kallikrein-related peptidase 2 (hK2), intact PSA, and free and total PSA. The four-kallikrein panel has shown an AUCs of 71.8% when predicting high-grade cancer (Gleason score $\geq$7). The Prostate Health Index includes total PSA value, free PSA value, and the PSA isoform [S2]proPSA. The PHI panel has shown similar AUCs as the 4K panel, with an AUC of 71.1% for detecting high-grade cancer (Gleason score $\geq$7) [21]. These panels have been evaluated in diagnostic trials, however a screening trial would be needed to provide evidence on the long-term outcomes.

2.4 Prostate cancer models

The Cancer Intervention Surveillance Modeling Network (CISNET) consortium uses statistical modelling to understand cancer interventions and their effects on incidence and mortality. CISNET was started in 2000 and includes models for breast, cervical, colorectal, esophagus, lung, and prostate cancer. Over the years CISNET has made substantial contributions to understanding the effects of mass screening [22, 23]. A particularly important contribution is the comparative modelling, where common inputs and outputs allows for comparisons of model assumptions. CISNET has also been successful in bringing together data from different data sources, including trial evidence. The consortium has, however, been constrained by the lack of longitudinal data from US data sources that are needed to set up and evaluate personalised screening programs. For

prostate cancer screening there are three microsimulation models used within CISNET, which were developed by the Erasmus Medical University, the University of Michigan and the Fred Hutchinson Cancer Research Center (FHCRC), respectively [24].

2.4.1 Erasmus Medical University

The Microsimulation Screening Analysis (MISCAN) program has been in continuous development at the Erasmus Medical University since the 1980s to model cancer screening for different cancer sites [25]. For prostate cancer, the cancer related event history is defined by a sequence of disease states and the ages at which these states are entered [24]. These life histories are generated by a semi-Markov process. The disease history are modelled in a preclinical phase and a clinical phase, where the preclinical phase does not correspond to clinical diagnosis, but the disease can be detected by screening. Because of this, the preclinical phase depends not only on the biological processes but also on the probability of cancer detection. Preclinical parameters are estimated from indirect evidence using maximum likelihood estimation using the Nelder-Mead algorithm. The MISCAN model includes 18 preclinical states, derived from combinations of T-stages (T1, T2 and T3+), Gleason grade (well, moderately, and poorly differentiated) and disease extent (loco-regional and distant). The cancer can progress from each preclinical detectable state to clinical diagnosis, where the cancer is diagnosed due to symptoms.

The screening is then superimposed on the simulated life histories. If a person has a preclinical disease state the early detection could then alter a person's life-history. The screening tests are modelled by its sensitivity and there is no direct representation of PSA in the MISCAN model. After clinical diagnosis, the MISCAN model uses stage and treatment specific survival.

2.4.2 Michigan

The Michigan model is a statistical mixture model of prostate cancer incidence and mortality [26, 27]. It was developed to analyse the US Surveillance, Epidemiology, and End Results (SEER) population and cancer registry data. It is used to understand, predict and optimise the population impact of cancer control processes in prostate cancer.

The model is composed of components by which predictions are made based on population data. The incidence component takes population data as inputs and predicts prostate cancer incidence by calendar year and age, both in the presence and absence of PSA testing. The survival component also takes population data as an input and models the relationship between a set of covariates (including age, year of diagnosis, cancer

stage and tumour grade) and a man's survival prognosis.

2.4.3 FHCRC

The FHCRC prostate cancer model has been described previously [28, 29]. The model includes prostate cancer onset as a function of age and continuous PSA growth with random effects as a function of age, time from cancer onset and Gleason score. It also includes clinical diagnosis and progression to metastatic cancer as a function of PSA as well as survival as a function of treatment, Gleason score, extent and age. This is the model that we extended and calibrated for the Swedish setting in **Study II**.

2.5 Cost-effectiveness of prostate cancer testing

Important modelling studies evaluating the cost-effectiveness of screening include the *Prostate Cancer Screening Options Appraisal* that was done for the British National Institute for Health and Care Excellence (NICE), in which clinical outcomes were modelled using microsimulations [30]. The NICE assessment concluded that PSA testing was not cost-effective.

More recently, Lao and colleagues [31] undertook a systematic review of cost-effectiveness for prostate cancer screening using the PSA test. They concluded that the cost per quality-adjusted life-year (QALY) was generally estimated to be over $275,000USD, which was not cost-effective in most jurisdictions. Several recent publications were not included in the review by Lao and colleagues, including cost-effectiveness analyses by [3] and [32] , who found evidence that prostate cancer testing could be cost-effective if carefully organised.

A NICE health technology assessment was undertaken for the Prostate Health Index (PHI) and the prostate cancer antigen 3 (PCA3) test [33]. This careful assessment came to the following conclusions: there were concerns with the precision of the PCA3 measurement; concerns with the storage and stability of the PHI samples; insufficient evidence to establish the clinical thresholds for both tests; and the authors' cost-effectiveness analyses suggested that neither test was cost-effective.

A recent review study concluded that although some model-based evaluations of prostate cancer screening have been found cost-effective, the evidence is lacking [34]. The study further emphasised the need for country-specific data.

For Sweden, a recent cost-effectiveness analysis by the National Board of Health and Welfare showed that PSA-testing for the ages 50–70 years would be cost-effective compared with the current opportunistic testing [35, 36]. The study did not, however, include the absence of PSA-testing as one of the comparators. It also assumed lower

levels of compliance for urology visits and biopsies for the comparator with opportunistic PSA-testing.

2.6 Modelling tools

Before we decided to develop the `microsimulation` R package, we reviewed the available simulation frameworks. Microsimulation models have been used in economics, demography and health; for a recent review of dynamic microsimulation models, see [37]. Perhaps surprisingly, there are moderately few general frameworks for continuous time microsimulations. One example is the proprietary `MODGEN` language, which is developed by Statistics Canada [38]. The `MODGEN` language has been re-implemented as an open source framework (`Openm++`, [39]), which uses different build tools.

Many discrete-event simulation (DES) frameworks and libraries focus on large, process-oriented simulations, where events are modelled as a series of processes (e.g. `SimPy`, `JADES`, and `CSIM`). In contrast, our task is focused on complex changes for a small set of events, which are well suited for event-oriented simulations. Several dynamic microsimulation models use a discrete time formulation (e.g [40, 41]). This, however, is less efficient than continuous time for modelling rare events.

There are comparatively few DES libraries available for `R` [42]: `simmer` supports process-oriented simulations, where the model is specified in `R` [43]; `MILC` models the natural history model of lung cancer and `MicSim` implements simple microsimulations, which are useful for teaching [44]. In contrast, `R` is frequently used for post-processing [40, 45, 46]. There are few open source microsimulation models for cancer screening.

For health economics specifically, I have a recent and positive experience with the `heemod` R package. The `heemod` package is a tool for health economic analysis with Markov models [47]. The Markov framework is arguably better suited for smaller models, where a detailed representation of the natural history is not required. Finally, the `hesim` package for health economic simulation modelling shows promise. It allows for multiple states with partitioned survival models and has a continuous time formulation [48]. The package is relatively recent and under intense development, but could prove useful also for more complex models.

3 Materials

3.1 Data sources

Several data sources were used during this thesis, including the Stockholm PSA and Biopsy Register (SBPR), the STHLM3 diagnostic trial, Prostate Cancer data Base Sweden (PCBaSe), life-tables and aggregated data on prostate cancer incidence and mortality (see Table 3.1).

The primary research database was the SBPR database [15], which consists of all men who lived in Stockholm county during the period 2003–2016. The SBPR uses an encrypted identifier to link to: PSA tests, including their PSA values (approximately 400,000 men with 1.5 million tests) and prostate biopsies (approximately 60,000 biopsies) from the laboratories serving the Stockholm county; the National Prostate Cancer Register (NPCR, with 20,000 diagnoses) and the National Cancer Register, where both databases have a coverage that is over 93% [49]; the Total Population Register, including migration to and from the Stockholm county; inpatient and outpatient hospitalisations; the National Death Register, including cause of death; and the Prescribed Drugs Register. Using information from SBPR, we modelled for PSA test uptake and re-testing, repeat testing following a negative biopsy, biopsy compliance and biopsy accuracy.

From the STHLM3 trial, we got the relative true and false positive fractions for the STHLM3 test for men with a PSA above a reflex threshold of 1 ng/mL from the baseline publication [20]. We re-analysed the STHLM3 data to calculate true and false positive fractions for reflex thresholds of 1.5 and 2 ng/mL, while keeping the STHLM3 threshold at the same level as the baseline publication.

PCBaSe is a national research database that links the National Quality Registry for Prostate Cancer (NPCR) with other health and population registers, including tumour characteristics, primary treatment and survival. We used PCBaSe to model survival from prostate cancer diagnosis by age, Gleason score, tumour extent and PSA values. Period-based relative survival estimates were provided by PCBaSe at 10 and 15 years after diagnosis for the period 1998–2014.

We also used aggregated data from life-tables provided by Statistics Sweden, and incidence and mortality rates for prostate cancer provided by Socialstyrelsen.

3.2 High performance Computing Resources

Microsimulation modelling of large populations can be computationally expensive. For a rare event like prostate cancer deaths, one would need to simulate a large population in order to contrast interventions (typically 10^6–10^8 individuals). Calibrating the model, i.e. fitting to observed data, is a particularly computationally demanding procedure. To do this, we used methods such as Nelder-Mead [50] and approximate Bayesian computation [51].

To access sufficient computational resources, we used high-performance computers at the Swedish Royal Institute of Technology's Center for High Performance Computing. This was facilitated through the Swedish e-Science Research Centre. These resources allow for both shared and distributed memory parallelisation. Without this type of resource, this project would not have been feasible.

Dataset name	Description	Approximate size	Studies
SBPR a.k.a. STHLM0	Population of men living in in Stockholm from 2003. Linked with registrations, deaths, migration, prescribed drugs etc.	400,000 men	**I–IV**
STHLM-3	Diagnostic study of prostate cancer screening protocol and biomarker development in 2013–2014.	60,000 men	**III**
PCBaSe	Survival at 10 and 15 years by PSA, grade and stage. PCBaSe links the national quality register on prostate cancer with the cause of death register.	80,000 cases	**II–IV**
Life-tables	Male life-tables by age and calendar year from Statistics Sweden.	Sweden, Stockholm county	**I–IV**
Prostate cancer incidence	Prostate cancer incidence by age and calendar year from Socialstyrelsen.	Sweden, Stockholm county	**II**
Prostate cancer mortality	Prostate cancer mortality by age and calendar year from Socialstyrelsen.	Sweden, Stockholm county	**II**

Table 3.1: The datasets used throughout studies I–IV.

4 Ethical considerations

The data sources we use are described in the section Data sources. The parameter inputs and calibration targets that we use for the model are derived from detailed individual-level data. These data are potentially identifiable and their analysis has required the appropriate ethical approvals. As part of those approvals, the data need to be securely stored and publication of the findings should not identify individuals. To ensure this, we consider the General Data Protection Regulation (GDPR), Swedish legislation and the KI rules in our use of individual-level data.

In contrast, the synthetic simulations use summary inputs and summary calibration targets, for which there are no ethical issues: the microsimulations are purely stochastic and do not have any meaning for any real individual. The simulated individuals are synthetic and are not sensitive data.

Another ethical aspect is our cost-effectiveness analysis, which uses measures such as age and health specific quality of life (these can even be discounted). The quality of life measure could be seen as a scalar for the value of a life-year, and is controversial in some context (e.g. in the United States). The use of cost-effectiveness thresholds can also be interpreted as setting a monetary value on life. Although these measures makes sense at a population level, they can easily become offensive if interpreted at an individual level. Care must be taken when communicating these type of results.

Besides aspects on security of personal level data and the communication of results, there are also ethical aspects on conducting open and reproducible research. To do this we have, as one of few, published our natural history model [52], and our articles are written using literate programming.

5 Methods

5.1 Screening dynamics

As may be evident from the Background section on PSA screening, the effects of screening for prostate cancer are debated. The controversy is partly due to classic challenges, such as whether a study assesses technical efficacy compared with real-world effectiveness, or due to issues with study design, such as contamination of the control arm. There are, however, inherent challenges in understanding screening because empirical evidence is only partially observed for a dynamic system. A mathematical formulation for screening was developed by Zelen and colleagues, with seminal articles by Day and Walter [53, 54]. Using a simplified model, we assume a constant intensity for latent, screen-detectable disease (e.g. $I = 0.01$), with an exponential time from cancer onset to symptomatic detection (e.g. with a mean of ten years), we could imagine screening at 55, 59, 63 and 67 years of age. The predicted incidence of symptomatic diagnoses and the prevalence of latent, screen-detectable disease at each of the screens are shown in Figure 5.1. We see that the incidence of symptomatic diagnoses and prevalence of latent cancers decline with repeated screens. Note that this part of the system is normally not observed and the incidence of clinical prostate cancer is expected to increase at older ages.

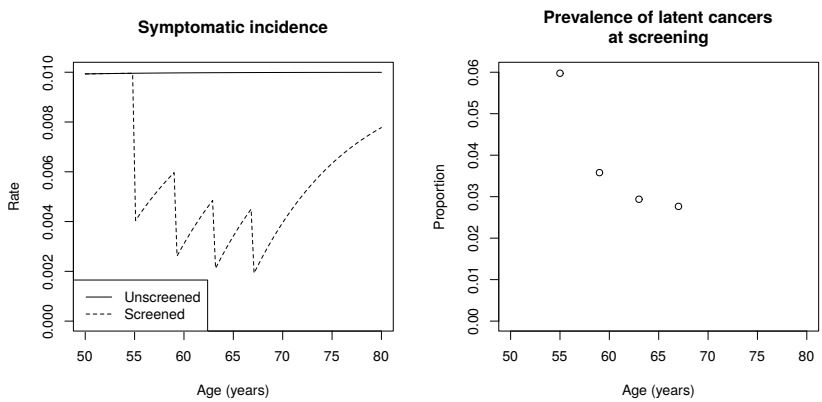

Figure 5.1: Simplified model for incidence and prevalence of pre-clinical disease for screening at ages 55, 59, 63 and 67 years. *Inspired by:* [53]

A second challenge is that the merits of a screening program consists not only of disease detection but also of the subsequent treatment and management. If one assumes that the prognosis is determined by the stage at detection, this can be used to model

disease specific survival. If an earlier detection leads to a shift in the stage at diagnosis, this can be used to model for a mortality benefit from screening [55]. However, if the model includes too few stages then the survival effect may "lack in resolution", leading to an under-representation of the survival benefit. This is the reason why we chose to model for T-stages, grouped by T1–T2 and T3–T4. To illustrate how this can lead to both overdiagnosis and treatment benefits, we constructed Figure 5.2. It shows two men: at the top, there is a man with a screening benefit under two counterfactual life-lines; and at the bottom, a man with an overdiagnosis under two counterfactual life-lines. As described in the caption, this figure also describes three time periods (sojourn time, lead time and delay time) that are specific to screening [53].

In summary, there are many factors that influence the effectiveness of early detection of prostate cancer and only a few have been evaluated by controlled clinical trials. These factors include the test characteristics, treatment modalities, population characteristics and detection strategies. These factors can also interact to influence the effectiveness of early cancer detection.

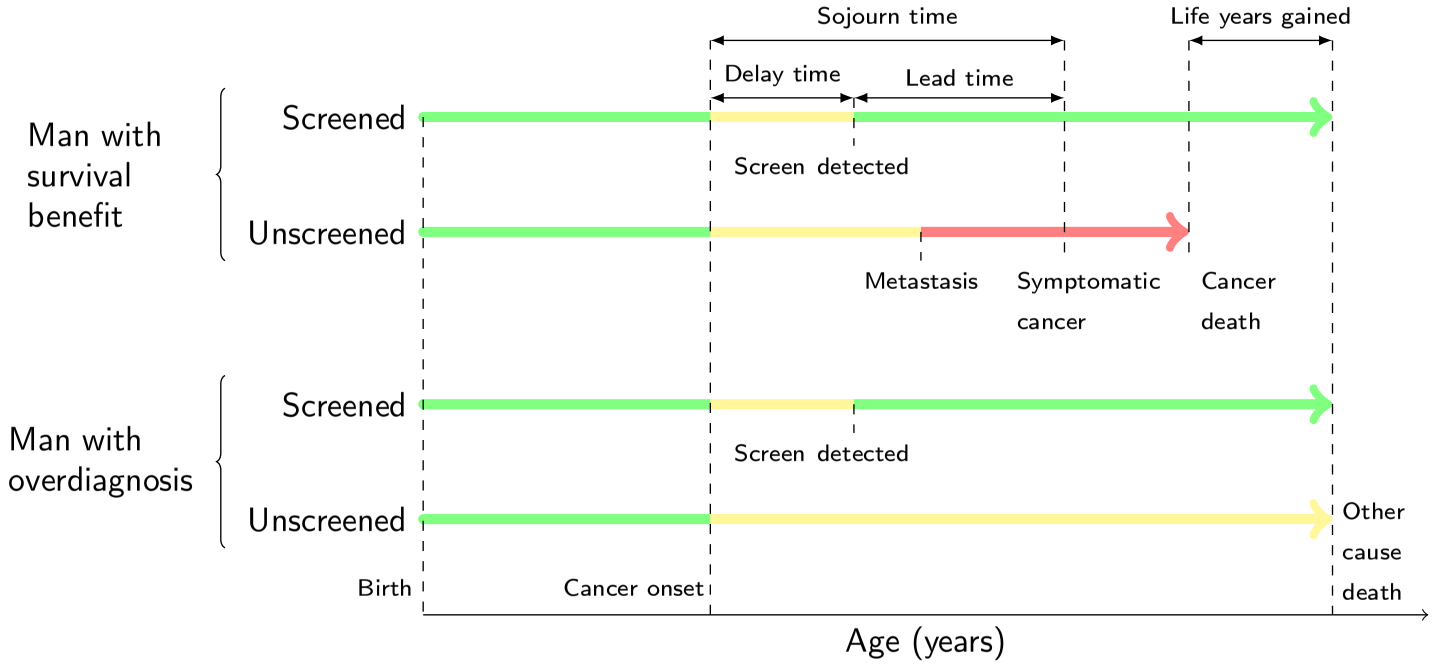

Figure 5.2: Example showing one man with screening benefits (top) and one man who experiences overdiagnosis (bottom). The life-lines for each man are shown under two scenarios: first under screening and then without screening. The man with the top two life-lines experiences early detection and cure under screening and would in the scenario without screening have developed terminal metastatic cancer. The man with the bottom two life-lines also experienced early detection through screening but would under the alternative scenario without screening never have developed symptomatic cancer. At the top we show some time intervals specific to screening: the period between detectable disease and clinical symptoms is called *Sojourn time*; the time between screen-detectable disease and an actual screen detection is called *Delay time* and the time between screen detection and a counterfactual clinical detection is called *Lead time* [53].

5.2 Model taxonomy

A system can either be described analytically, with e.g. a system of equations, or as a process implemented via a simulation model. If the system of interest is too complex to solve analytically, the modeller may then need to use simulation. If the model has time dependencies, the model is called *dynamic*; in contrast, a *static* (or steady-state) model calculates the system in equilibrium [37]. If the model also represents uncertainties it is called probabilistic or *stochastic* and if the model result is completely determined by its initial conditions it is called *deterministic*. The modeller also needs to decide whether to represent the modelled variables, including time, states and events, either as being *continuous* or *discrete*. For discrete events, the model formulation can be either *process-oriented* (when the events are modelled within a process), or *event-oriented* (when the events are modelled separately). Lastly, the population can be modelled using *groups* or using *individuals*.

For continuous time with discrete states, we can consider the transition intensities or rates from one state to another to be dependent on age, which describes a *Markov* process, or the rate may depend on the time in state, which describes a *semi-Markov* process. For our purposes, we will use continuous time with discrete states, with a combination of Markov and semi-Markov rates for individuals. This is a form of *microsimulation*. In health sciences and the econometrics, these units are often individuals and the microsimulation will generate stochastic life histories. We have implemented this model within a framework for *discrete-event simulation* (i.e. no changes in the system occur between consecutive events) using event-oriented simulations [56].

To reduce the Monte Carlo error, we use random number streams for different processes and use random number sub-streams for each individual, as proposed by Stout and colleagues [57].

5.3 Calibration, validation and sensitivity analysis

The validity of the model is important if the model is going to provide useful predictions. Some of the model parameters can be estimated directly from observable data. Examples of this include PSA re-testing and biopsy compliance. However, other model parameters are not directly observable. We can use *calibration* to ensure that the model parameters reflect the available data. We can calibrate either by using maximum likelihood estimation or Bayesian computations. Importantly, we can only observe part of the data that we ideally would like to have and typically we fit models to observed marginal distributions using *indirect estimation* [58] or *approximate Bayesian computation* [51].

The approximate Bayesian computation (ABC) approach has the advantage of allowing for prior (empirical) distributions for some parameters and, very usefully, sampling from the posterior distribution.

Moreover, we can compare the predictions with available data for model *validation*. As for calibration, this will often use observed marginal distributions, such as cancer rates by age and stage, to assess whether the model predicts observed patterns.

Given the marked uncertainty in the model parameters, it is important to understand how sensitive the predictions are to model assumptions. The classical approach in health economics is to use *probabilistic sensitivity analysis*: prior distributions for parameters are independently sampled 1000 times, and then predictions are recalculated to demonstrate the effect of uncertainty on the parameters. A more computationally efficient, yet challenging, approach is to take samples from the posterior distributions to represent the parameter uncertainty, and use those samples to show uncertainty in the predictions. Both of these approaches are computationally challenging. One alternative approach used by cervical cancer screening modellers at Harvard is to perform a grid search on the parameter space and keep the values that are close to the maximum likelihood estimate using a likelihood ratio test [59].

5.4 Computational methods

5.4.1 Nelder-Mead

Nelder-Mead is numeric method for derivative-free function optimisation $f : \mathbb{R}^k \to \mathbb{R}$ [50]. Nelder-Mead does not guarantee reaching a global optimum and therefore the point of convergence may depend on the initial values. The algorithm first starts by calculating a simplex, S. This is a geometrical shape, with $k+1$ vertices for $\in \mathbb{R}^k$. Then follows a sequence of steps, which varies between implementations. Generally this sequence involves ordering of vertices, calculating the centroid on the opposite side of the worst vertex, and a transformation step. During the transformation step, the geometrical shape of the simplex is changed and the proposed vertex is compared with the previous centroid point. If the new point better optimises the function then the proposed point becomes part of the new simplex. In **Study II** we used the Nelder-Mead algortihm provided in R's `optim()` function.

5.4.2 Calculating standard errors from the Hessian

The Hessian is a square matrix of the second-order partial derivatives of a scalar-valued function with respect to its parameters. We focus here on the log-likelihood function $l(\boldsymbol{\theta})$

with variables θ. The derivatives describe the curvature of the function over variables θ_i and θ_j. The Fisher information can be expressed as $\mathrm{I}(\hat{\theta}_{\mathrm{ML}})$, $\hat{\theta}_{\mathrm{ML}}$ are the maximum likelihood estimates, such that

$$\mathbf{I}(\hat{\theta}_{\mathrm{ML}}) = \left(-\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta) \right) \Bigg|_{\theta=\hat{\theta}_{\mathrm{ML}}}, \; 1 \leq i \leq k, \; 1 \leq j \leq k \tag{5.1}$$

where k is the length of the vector θ.

The observed Fisher information can be found by minimising the negative log-likelihood, and calculating the Hessian for the minimum. The standard error (SE) of the estimate, can then be calculated by taking the square roots of the inverse of the diagonal elements of the covariance matrix. For the i^{th} value of $\hat{\theta}_{\mathrm{ML}}$, the standard error is

$$\mathrm{SE}(\hat{\theta}_{\mathrm{ML},i}) = \sqrt{(\mathbf{I}(\hat{\theta}_{\mathrm{ML}})^{-1})_{ii}} \tag{5.2}$$

Calculating analytical derivatives for our simulation model was not possible, due to the nested random-effects and the hierarchical modelling of events. The Hessian were calculated through numerical differentiation using finite differences.

Numerical differentiation based on simulations from a stochastic, and therefore imprecise, function has some inherent challenges. Besides the obvious solution of increasing the simulation size even further for a better signal-to-noise ratio, we benefited from increasing the distance between the function evaluations, denoted Δ. As exemplified in Figure 5.3, the stochastic variation makes it difficult to calculate the partial derivatives for a Δ where the stochastic variation is large in comparison to the likelihood curvature.

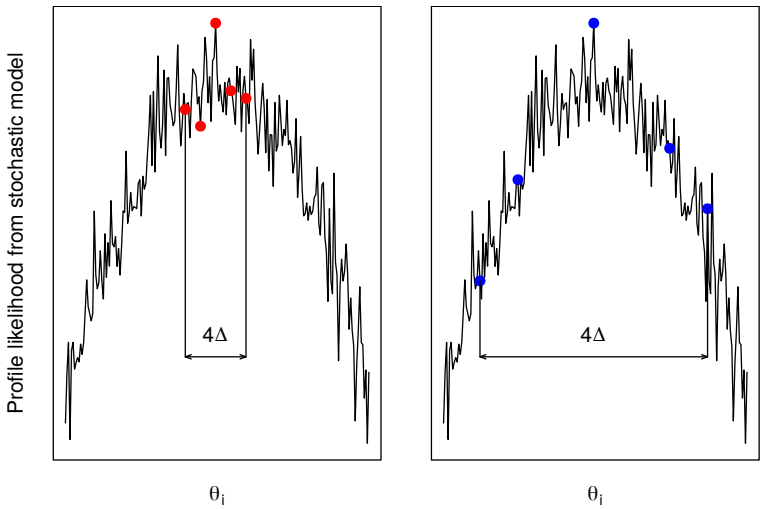

Figure 5.3: Example showing the challenge in numerically calculating the Hessian with imprecise function evaluations from a stochastic model. The Hessian from the left panel with a smaller Δ would not be consistent with a function optimum, whereas the Hessian from the right panel with a larger Δ would.

5.4.3 Markov chain Monte Carlo

The mechanism behind a Markov chain Monte Carlo (MCMC) is to construct and simulate a Markov chain whose equilibrium distribution is the distribution we want to sample from. A time homogeneous Markov chain will converge to a stationary (equilibrium) distribution, independent of the initial parameter values.

The Metropolis-Hastings algorithm is an MCMC algorithm that, based on the current location of the chain, picks a new candidate sample point, which then either is rejected or accepted. New candidate sample points are picked using a proposal density $Q(\theta^*|\theta^{(t-1)})$ [60]. In **Study IV** the proposal density was a multivariate normal distribution whose covariance matrix was estimated from a random sample of the parameter space in which the target distribution was evaluated. The steps in the Metropolis-Hastings algorithm are outlined as follows:

1. Propose a move to θ^* by drawing from $q(\cdot|\theta)$, where θ is the current state.

2. Calculate $r = \pi(\theta^*)q(\theta|\theta^*)/\pi(\theta)q(\theta^*|\theta)$

3. Move to θ^* with probability $\min(1,r)$, else stay at θ; go to 1.

5.4.4 Approximate Bayesian Computations

In Bayesian inference, we are interested in either computing or sampling from the posterior distribution of a set of parameters θ given data $\mathcal{D}$ generated from a model $\mathcal{M}$. The posterior distribution is deduced from the data likelihood $P(\mathcal{D}|\theta)$ and a prior distribution $\pi(\theta)$:

$$P(\theta|\mathcal{D}) = \frac{P(\mathcal{D},\theta)}{P(\mathcal{D})} = \frac{P(\mathcal{D}|\theta)\pi(\theta)}{P(\mathcal{D})} \propto P(\mathcal{D}|\theta)\pi(\theta) \tag{5.3}$$

To deduce, or simulate from, the posterior is not possible with the standard framework if the data likelihoods are hard or impossible to compute. ABC was proposed to simulate observations from posterior distributions without the use of likelihoods. It can also be useful when a full Bayesian analysis is computationally heavy. The rejection sampling version of ABC is summarized in the following steps [61]:

1. Generate θ from $\pi(\cdot)$
2. Simulate $\mathcal{D}'$ from model $\mathcal{M}$ with parameter θ
3. Calculate a measure of distance $\rho(S(\mathcal{D}'),S(\mathcal{D}))$ between $S(\mathcal{D}')$ and $S(\mathcal{D})$, where S is a summary statistic.
4. Accept θ if $\rho < \delta_\varepsilon$, and return to 1

For a given distance function ρ and a threshold δ_ε, accepted observations are independent and identically distributed from $f(\theta|\rho(S(\mathcal{D}),S(\mathcal{D}')) \leq \delta_\varepsilon)$ Ideally, S should be a *sufficient* statistic for θ, otherwise the resulting posterior will only be an approximation of the true posterior distribution [62].

5.4.5 ABC-MCMC

Although, rejection sampling in ABC is easy to implement and generates independent observations, it can be very inefficient as the acceptance rate of the samples is low [62]. Instead, a MCMC approach has been proposed in connection with ABC [61]. The method, outlined below, is based on the Metropolis-Hastings algorithm where $q(\theta^*|\theta^{(t-1)})$ is a proposal distribution.

1. Propose a move to θ^* by drawing from $q(\cdot|\theta)$, where θ is the current state.
2. Simulate data $\mathcal{D}'$ using model $\mathcal{M}$ and parameters θ.
3. If $\rho(S(\mathcal{D}'),S(\mathcal{D})) < \delta_\varepsilon$, go to 4, else stay at θ and return to 1.

4. Calculate $r = \pi(\theta^*)q(\theta|\theta^*)/\pi(\theta)q(\theta^*|\theta)$

5. Move to θ^* with probability $\min(1,r)$, else stay at θ; go to 1.

Various improvements to the ABC-MCMC approach have been proposed, including transformations of the summary statistics S [62] and ABC regression adjustments [63].

5.5 Cost-effectiveness analysis

To provide evidence to policy makers for deciding on which cancer screening interventions are most suitable, it is common to undertake a health economic evaluation. An evaluation that compares the lifetime expected costs and lifetime expected health utilities for different interventions is called a cost-utility analysis or a *cost-effectiveness analysis*. For a screening intervention k, the expected values for the effectiveness and costs can be calculated by simulating the life histories for $i = 1,\ldots,n$ individuals and then calculating the average, such that:

$$\text{Effectiveness}_k = \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \frac{dU_{ik}(t)}{(1+\delta)^t}$$

$$\text{Costs}_k = \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \frac{dC_{ik}(t)}{(1+\delta)^t}$$

where we have time dependent cumulative utilities $U_{ik}(t)$ and cumulative costs $C_{ik}(t)$ for person i, where δ is the discounting rate (e.g. $\delta = 0.03$). The costs and utilities are typically taken from reviews of the literature or analyses of register data. The costs can be taken from a *health sector* or *societal* perspective and the utilities can be calculated using years of life lost or using quality-adjusted life-years.

For decision analysis, two scenarios can be compared using an incremental cost-effectiveness ratio (ICER). The ICER is a ratio of the cost difference and the effectiveness difference from the compared scenarios. For a plot of the effectiveness against the costs, the convex hull of values provides a *cost-effectiveness frontier*, which suggests which scenarios or combination of scenarios are most effective given a cost-effectiveness threshold [64].

For **Study III**, costs were taken from a review by the Swedish National Board of Health and Welfare, as described in the section Cost-effectiveness of prostate cancer testing [36]; see the full paper for details. The cost inputs for this study were largely based on the southern region in Sweden. Particularly influential costs, e.g. the cost of a biopsy, were compared with the cost per patient for diagnosis-related disease groups ("N75O manliga genitalia px O", [65]).

Health state values were taken from a review by Heijnsdijk and colleagues [66]. These values combine evidence from different populations using different study designs and instruments.

6 Summary of the results

In summary, we provide evidence, methods and tools for planning cancer screening. In **Study III**, we present evidence for policy planning of prostate cancer screening in Sweden. By using the microsimulation framework from **Study I** and the natural history model from **Study II** we performed a cost-effectiveness analysis of the STHLM3 test. Leveraging our experiences from **Study II**, we present a method for calibrating microsimulation models in **Study IV**.

6.1 Study I

In **Study I**, we presented our parallel discrete-event microsimulation R package. We motivated our design choices, such as the use of `Rcpp` for speed, the `SSIM` library for handling events and the `RngStreams` library for common random numbers. We also investigated how the simulations could be mapped onto several processor cores followed by a reduction operation to collect the results in a single report. We investigated the efficiency of four alternative implementations, specifically using `R`'s parallel package, `OpenMP`, `MPI` using the Rmpi package and a hybrid implementation using both `OpenMP` and `MPI`. While simulating a cohort of 10^7 individuals, we measured the speed for all implementations on a single compute node and for the distributed memory implementations, `MPI` and `OpenMP/MPI`, on up to 16 nodes. We observed close to linear scaling for all implementations within a compute node. `MPI` was the fastest implementation for the first three nodes. The hybrid `OpenMP/MPI` continued to scale, also after the first three nodes, until approximately 10 nodes and at 16 nodes and 128 cores it had an efficiency of approximately 25%. Finally, we showcase the use of the `microsimulation` package with a case study.

Besides the scientific publication, an equally important output of **Study I** is the software. The `microsimulation` R package allows other researchers with an interest in modelling cancer screening to leverage these features. However, the flexibility and speed comes at the price of accessibility. This package targets modellers who are willing to specify their models in `C++` using the provided classes and methods. Whereas the original FHCRC code, was written in `C` with little to no abstraction of the simulation framework. In contrast, we provide a framework, readily available, for modellers wishing to investigate cancer screening and its cost-effectiveness using microsimulation.

The microsimulation R package is open source and available under a GNU General Public License (GPL) 3 license at:

`https://github.com/mclements/microsimulation`

6.2 Study II

For **Study II**, we adapted and extended the FHCRC prostate cancer screening model. The model simulates PSA growth together with individual life histories, including cancer onset, and progression from localised to metastatic disease by Gleason score. Using the SBPR and PCBaSe databases, we included additional T-stages and Gleason grades in the model. We also calibrated the survival by PSA, M-stage, Gleason grade and age. Using results from ERSPC, we calibrated the model to the reported incidence rate ratio (IRR) and validated the model against the reported mortality rate ratio (MRR). We also used Swedish prostate cancer incidence and mortality rates for validation.

We also predicted the effects of dynamically changing the current opportunistic PSA testing pattern for regular PSA screening. We predicted a reduction of the prostate cancer incidence with an IRR 0.86 (95% Monte Carlo Interval (MCI): 0.86–0.86) and an increase of prostate cancer mortality with a MRR of 1.02 (95% MRR: 1.01–1.02), when introducing 8-yearly screening in men aged 55–69 over 20 years.

The resulting natural history model, "Prostata", is in itself an important output of **Study II**. The Prostata model is available as an R package for modelling prostate cancer screening for Sweden. The R user can change hundreds of parameters and simulate 18 different screening scenarios. The `prostata` R package can be used to provide evidence on cost-effectiveness of prostate cancer screening in Sweden.

The prostata R package is available under an open source (GPL 3) license at:
`https://github.com/mclements/prostata`

6.3 Study III

We investigated the cost-effectiveness of the STHLM3 screening test for prostate cancer. The test uses blood analysis of proteins, a genetic risk score and clinical variables to test for prostate cancer. The test process for STHLM3 includes: first a PSA test; men with a PSA above a reflex threshold (1 ng/mL in the trial) had their risk for Gleason ≥ 7 predicted, and finally men with a risk above 10% were referred to a urologist who after a digital rectal exam could perform a biopsy. For men with PSA between 3 and 10 ng/mL, this process reduced the number of benign biopsies by 44% (95% confidence interval (CI): 35–54%) and the number of biopsies resulting in a Gleason 6 cancer by 17% (95% CI: 7–26%), while maintaining the same number of Gleason 7+ cancers. We used the reported results for a reflex threshold of 1 ng/mL and we re-analysed the STHLM3 trial data for reflex thresholds of 1.5 and 2 ng/mL.

Using the natural history model from **Study II**, we predicted that PSA testing every four

years would result in 71 fewer deaths and 652 more life-years per 10,000 men, compared with no screening. When using the STHLM3 test at a reflex threshold of 1.5 mg/nL, we predicted 4 more deaths and 22 fewer life-years, but also a reduction in biopsies by 2235 per 10,000 men. This results in a slight increase of 4 QALYs for 10,000 men, compared with PSA testing.

We also evaluated the cost-effectiveness of the STHLM3 test under a four yearly screening program for ages 55–69 year. The STHLM3 test was more cost-effective at higher reflex thresholds when comparing to PSA testing. At reflex thresholds of 1, 1.5 and 2 ng/mL the STHLM3 test had ICERs of 170,000, 60,000 and 5000 EUR/QALY, respectively.

The STHLM3 test, when used as a reflex test for PSA $\geq$ 2.0 ng/mL, was predicted to have a low cost per QALY compared with PSA testing for Sweden. However, using the STHLM3 test for lower reflex thresholds was predicted to have high costs per QALY for Sweden.

6.4 Study IV

For models to be useful, they need to be calibrated against observed data targets. This calibration may be difficult for calibration targets on different scales and when data likelihoods are not available. In **Study IV**, we propose a method for addressing some of the challenges we experienced in **Study II**. Our methods incorporate ABC within a standard Bayesian framework to allow for calibration to data with partially unknown likelihoods. We then use MCMC for sampling from the posterior distribution. For calibrating the model described in **Study II**, we needed to set acceptance regions for some of the calibration targets.

We compared our suggested ABC hybrid method with a full ABC approach and the likelihood approach with quadratic constraints presented in **Study II**. We found that the full ABC approach with a negative log-likelihood cut-off of 450 did not, as expected, optimise the likelihood well. The likelihood approach with quadratic constraints did optimise the likelihood better, with resulting negative log-likelihood of 302, but required somewhat arbitrary scale factors for the quadratic scale factors. Initially, we experienced difficulties in sampling from the parameter space with the ABC-hybrid method, due to sharp peaks in the likelihood. We solved this by scaling the likelihood before sampling and then re-weighting the posterior distribution. For the ABC-hybrid we reached a negative log-likelihood of 270, while fulfilling the other constraints. The benefits of the ABC-hybrid method came at the cost of increased computational requirements.

7 Discussion

The question on how to screen for prostate cancer has proven to be challenging, which motivates the use of more complex models. However, these models rely on assumptions whose validity should be assessed. The evidence we provide should, therefore, be considered together with evidence from other modelling studies and the controlled trials.

One of the challenges with microsimulation modelling is assessing whether the model is valid for a specific research question. We have spent considerable time improving the model to evaluate the cost-effectiveness of the STHLM3 test. For the STHLM3 test, we extended the FHCRC model to include more detailed modelling of Gleason scores. Our subsequent fitting of survival under current screening to the NPCR also led to a marked improvement in the modelling of differential survival. We further improved the survival model by allowing for T stages in the natural history model. The question remains: how do we know whether we have a good model for prediction? Moreover, a further challenge is how to validly represent uncertainty for a large and complex model, as outlined in **Study II** and **Study IV**.

7.1 Limitations

There are several important limitations of our approach. First, survival may not fully represent the more recent Gleason grading, where the longer-term survival was represented by cases from the late 1990s. Conversely, it would have been useful to have good survival data from an unscreened population, however data on Gleason grading prior to 1998 were not systematically collected and any such data would be based on older Gleason grading. Second, the model is not well calibrated at older ages. This limitation leads to difficulties in the interpretation of the epidemiological outcomes over a life-time (e.g. over-diagnosis and prostate cancer deaths). Third, register data are often less reliable at older ages, including cause of death in older men. Fourth, our modelling of the effectiveness of prostate cancer testing is critically dependent on one trial (ERSPC). If the point estimates for the incidence and mortality rate ratios for PSA testing are biased, then our predictions will also be biased.

Fifth, we have assumed that policy should be based on out-of-sample predictions from the observed randomised trial to life-time predictions. From our model, the short-term estimates of costs and effectiveness are qualitatively different to the life-time predictions. Consequently, the choice of time-line will lead to different conclusions. However, we also know that the long-term predictions are imprecise and possibly biased.

Sixth, from **Study III**, there is an extensive discussion on the uncertainty in the cost

inputs, particularly for the biopsies. There are substantial discrepancies between the listed outpatient price in Stockholm and the price used in the governmental report [36], which in turn is similar to the national reimbursement provided to the hospital. We decided to base our costs on the government report and addressed this uncertainty in a sensitivity analysis. This uncertainty warrants further investigation.

Seventh, there is uncertainty in the health state values used for **Study III**. The health state values were developed by Heijnsdijk and colleagues [66] and are frequently cited. However, the original estimates of the health state values were based on multiple instruments with different study designs from multiple populations.

7.2 Strengths

Our approach has a number of strengths. We started out with a prostate cancer natural history model that was well validated in the US population. Using the detailed Swedish health and population registers, we carefully extended and calibrated the model. Using those contemporary data, we expect that the calibrated model will represent the effects of modern diagnostic procedures and treatment well. Our model is particularly well suited, and arguably the best model available, to the evaluation of reflex tests, such as STHLM3 and the Prostate Health Index test.

We have aimed to follow the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) guidelines for DES outlined by Karnon and colleagues [67]. Particularly important *best practices* we use are: stochastic modelling of time-to-event as fixed times in the disease states could easily be misleading when designing screening policies; we do not solely rely on guidelines but use observed clinical practice from SBPR; we take life-time perspective with continuously modelled risks of disease progression and other cause death; we use variance reduction techniques such as common random numbers; and we quantify uncertainty using a probabilistic sensitivity analysis.

Finally, we used some of the target data for validation rather than calibration. For example, prostate cancer incidence and mortality rates were used for validation. Those data would probably aid in any improvement of the model calibration, however our approach provides a more objective evaluation of the model validity.

7.3 Our contribution to open science

The development of a natural history model represents a considerable effort by the investigators. It is therefore tempting for researchers to not make their models open and

available to others. This can be motivated by arguing that a closed source model, financed through occasional consultancies, allows for more development resources and better control of the model development and its use.

On the other hand, such a model is difficult for other researchers to validate and could be critiqued for being a "black box". It is also not clear to us, that a closed model will develop better over time. We are grateful towards Ruth Etzioni and Roman Gulati for sharing the original FHCRC model with us. The choice to extend upon the FHCRC model was simple: it was the only open source CISNET model. Continuing on this path, we have released both the `microsimulation` and `prostata` R packages under an open source license, with the hope that others might use, validate and extend them.

7.4 Generalisability

We have experience with comparing and adapting *one* natural history model to the Swedish setting. The uncalibrated model did not represent incidence, mortality or survival well. This could be due to differences in the Swedish population or perhaps in part due to more detailed data.

In theory there is a clear distinction between model structure and model inputs. However, for these extensive models, some inputs are not easily available, e.g. cancer onset distributions, in what is otherwise a data-rich setting. The model generalisability is therefore confounded by those uncertain data inputs.

We have made several adaptions and extensions of the Prostata model to better represent prostate cancer in Sweden. Most of those model inputs can be easily replaced, e.g. the PSA threshold of 3 ng/mL, treatment patterns, biopsy compliance patterns, and the population structures. Other parts, such as calibrated parameter values, would require detailed data and there would be some effort to re-calibrate the model for a new context.

The cost-effectiveness analysis of the STHLM3 test in **Study III** was performed from the Swedish context, where the natural history model, the cost inputs and the background health state values were specific to Sweden. Interpretations for other countries should be done with caution.

7.5 The right tool for the job

Cancer screening is a complex process, where latent processes and lead-time bias make attempts to simplify the model structure more prone to errors. The lack of contemporary untested populations further complicates attempts to understand the screening effects. Nevertheless, evaluations of cancer screening for the Swedish National Board of Health

and Welfare have often been performed using simpler deterministic, so called "cohort", models. Although this model structure may be useful for transferring directly observed trial results to the Swedish population structure, any extrapolation would rely heavily on assumptions. The accuracy of these simpler models remain unclear, since they are not available for validation by other researchers.

The question of model complexity is also interesting on a more philosophical level. The main use of these models are out-of-sample predictions, e.g. life-time risk, novel tests or altered testing patterns, which are difficult to validate against data. Instead they rely on the assumption that the modelled mechanistic properties remain similar also for the extrapolation.

7.6 Concluding remarks

Microsimulation modelling of the natural history of prostate cancer is computationally and scientifically challenging. It requires detailed data and an understanding of the healthcare pathways as well as statistical and computational resources. Once a model is well calibrated to a population it can be used for answering questions on detection, diagnosis and treatment. In particular, early detection is difficult to model in other ways.

This thesis describes the development of tools and methods for microsimulation modelling for cancer screening in general, and prostate cancer screening in Sweden in particular. We provide predictions for the dynamic effects on prostate cancer incidence and mortality in Sweden, should screening be introduced with the current opportunistic PSA testing patterns as a baseline. Prostate cancer screening has been assessed many times, in many countries, but remains an open question.

Assessing the prediction uncertainty is a difficult task. Ideally, model assumptions should be investigated on a larger scale by multimodel initiatives like CISNET. The comparative modelling by CISNET may, at least in its earlier years, have described uncertainty in the model assumptions. However, longer-term collaborations may lead to cross-pollination between the models, where the predictions from the different models are more precise (i.e. lower variance), but there may be shared bias across the models.

Assessing the uncertainty in model inputs is more easily done. This could be done either by deriving the standard errors from the Hessian as in the probabilistic sensitivity analysis in **Study III** (partly based on the standard errors from **Study II**) or using the methods we develop in **Study IV**.

One way to address the question of early detection of prostate cancer is to use a test with a higher specificity, e.g. the STHLM3 test. We investigated the cost-effectiveness of the STHLM3 test, at its current price, and found it cost-effective if used in a group with

higher risk. Specifically, the STHLM3 test was found cost-effective when used as a reflex test following a PSA of 2 ng/mL, but not when used after a PSA of 1.5 ng/mL or lower.

8 Future perspectives

8.1 Other applications for the Prostata model

Given its careful calibration and validation, the Prostata model is now well suited to answer a range of other research questions. In particular, we plan to assess the cost-effectiveness of magnetic resonance imaging (MRI) at a urological consultation prior to a prostate biopsy and in combination with an MRI-guided biopsy. Moreover, there is considerable interest in whether MRI will be cost-effective when combined with the newer prostate cancer screening tests.

Sweden uses a lower PSA threshold than many other countries. In Sweden, a PSA threshold of 3.0 ng/mL is used for men below age 70, without familial risk. In contrast e.g. the United States uses a PSA threshold of 4 ng/mL. One interesting question would be to investigate how this policy choice affects the effectiveness and costs of testing with PSA. Another question of interest would be to investigate how the STHLM3 test would perform in comparison to a PSA threshold of 4 ng/mL.

8.2 Potential model structure development

For the Prostata model to represent the tumour biology as well as possible, some future development could be considered.

First, we could incorporate more detailed modelling of tumour size and extent of spread. One approach would be to include further T-stages. Second, we could investigate whether the model predictions are affected by Gleason de-differentiation. There has been some debate in the literature as to whether the Gleason grade is largely fixed at cancer onset or whether it de-differentiates with time following cancer onset. De-differentiation is not directly observable, Gleason scores are imprecisely measured, and the modelling and indirectly observed evidence is inconclusive [68]. Third, we would like to merge more recent changes for the FHCRC model into the Swedish Prostata model, including changes to the onset distribution and the shape of cancer survival. Fourth, we could revise the PSA sub-model using the PSA values available from the SBPR. Fifth, we could investigate the accuracy for the symptomatic diagnoses in the older ages, also using SBPR.

We also see an opportunity for inputs from the Prostata model to be incorporated back into the CISNET prostate cancer models. In particular, the Swedish data on Gleason scores and prostate cancer survival are more contemporary than data from the US SEER registries.

8.3 Applications to other domains

The simulation framework is well suited to moderately complex disease modelling. Candidate diseases for the framework could include: breast cancer, using detailed natural history models developed by Keith Humphreys and colleagues [69]; cervical cancer screening, to take advantage of the National Cervical Cancer Register; and models of diabetes care.

9 Sammanfattning på svenska

För att kunna basera riktlinjer inom vården på evidens, måste man ibland utvärdera interventionerna med hjälp av modeller. Valet av hur komplex en modell bör vara är en avvägning, där en enklare modell kan vara lättare att beskriva och kalibrera, medan en mer komplex modell bättre kan representera sjukdomsdynamiken och leda till mer giltiga prediktioner. Att modellera riktlinjer inom vården är en tvärvetenskaplig uppgift, som kräver kunskap inom flera discipliner t.ex. epidemiologi, medicin, biostatistik, hälsoekonomi och datavetenskap.

Detta illustreras i **Studie III**, där vi företog oss att bedöma kostnadseffektivitet hos ett nytt test (STHLM3-testet) för tidig upptäckt, s.k. screening, av prostatacancer. STHLM3-testet använder en kombination av biomarkörer och självrapporterad data för att förutsäga risken för prostatacancer. STHLM3-testet kan användas som ett reflextest efter ett PSA-test. För att kunna avgöra hur väl testet fungerar för screening av prostatacancer, behövde vi en modell av naturalförloppet hos prostatacancer. Modellen användes sedan till att prediktera korta och långsiktiga effekter av olika screeninginterventioner.

För kunna göra detta utvecklade vi först, i **Studie I**, ett ramverk för att simulera diskreta händelser med kontinuerlig tid i `R` och `C++`. I ramverket ingår bland annat hantering av gemensamma slumptal för att minska Monte Carlo-felet och detaljerade rapporter för hälsoekonomiska utvärderingar. I **Studie II** utökade vi en existerande amerikansk modell av naturalförloppet hos prostatacancer. Vi använde svenska indata för bland annat PSA-tester, diagnoser av prostatacancer, behandling, vårdförlopp och överlevnad. Kalibreringen av naturalförloppet inkluderade båda PSA-testade och icke testade svenska populationer. Till att börja med kalibrerade vi "Prostata" modellen genom att maximera en *likelihood* tillsammans med icke-linjära restriktioner. Efter det har vi i **Studie IV**, utvecklat en metod baserad på approximativa Bayesianska beräkningar och Markovkedje-Monte Carlo metoder. Vår metod använder ett mer systematiskt tillvägagångssätt för att väga samman bevis på olika skalor och drar samtidigt nytta av de *likelihoods* som är tillgängliga. I **Studie III** inkluderade vi mått på kostnader och livskvalitet och beräknade förväntad livslängd och kostnader under en livstid för olika screeninginterventioner. Vi fann att STHLM3-testet var kostnadseffektivt i Sverige om det används som ett reflextest för PSA högre än 2 ng/ml.

En lärdom är att mikrosimulering är både en beräkningstung och vetenskapligt utmanande uppgift, särskilt för kalibrering och känslighetsanalyser. En annan är att, ironiskt nog, så gör detaljrikedomen i de svenska sjukvårdsregistren att det är lätt att hitta brister i modeller som är välvaliderade i andra populationer. Svenska data har därför potential att förbättra de redan befintliga modellerna för cancerscreening.

Styrkor i vårt tillvägagångssätt är: ett flexibelt, snabbt, skalbart och öppet ramverk för att utveckla riktlinjer inom cancerscreening; vi har kalibrerat modellen av naturalförloppet till nuvarande svensk incidens och överlevnad noggrannare för Gleasongrader, medan de flesta andra modeller har kalibrerats till äldre, PSA-naiva populationer; vi har införlivat detaljerad information och dragit bästa möjliga nytta av de tillgängliga svenska sjukvårdsregistren. Begränsningar i vårt tillvägagångssätt är: osäkerhet kring storleken på effekten som screening har på dödligheten i prostatacancer; osäkerhet i validiteten av prediktionerna utanför observerade data och osäkerhet i hur korrekta livskvalitetmåtten är. I framtiden planerar vi att använda Prostata-modellen för att utvärdera användandet av magnetisk resonanstomografi i kombination med nya screeningtest för prostatacancer.

10 Acknowledgements

◇ **Supervisors & mentor**
Thank you!

Mark Clements for your knowledge, courage and genuine enthusiasm. I have learnt so much from you. Besides your many skills in statistics, modelling, coding$^{\lambda}$ and archaic text editors – I have the deepest respect for you as a scientist. I have found knowledge, altruistic motives in combination with integrity to be too rare in science. You set a great example.

Alexandra Jauhianien for your statistical skills, resourcefulness and tireless support. Had it not been for your guidance I would have been calibrating the model throughout eternity. I really appreciate all the effort that you have put in and for making my trips to Göteborg so pleasant.

Henrik Grönberg for your deep clinical knowledge and the vivid scientific environment that you provide. I also have deep respect for your drive and ambition, that is currently improving prostate cancer detection.

Erwin Laure for our guiding meetings and for sharing your knowledge on high performance computing. Those skills have been essential throughout my thesis.

My mentor **Cecilia Lundholm** for your level-headed thinking and good advice. Having a coffee with you have often made the world look a little better.

◇ **Colleagues & friends**
I would also like to express my gratitude to all the brilliant researches and dear friends, that I got to know during these years.

To past and present **Biostat PhD** students: **Linda Abrahamsson**, **Nurgul Batyrbekova**, **Rose Bosire**, **Hannah Bower**, **Elisabeth Dahlqwist**, **Wenjiang Deng**, **Sandra Eloranta**, **Pablo Gonzalez Ginestet**, **Shuang Hao**, **Gabriel Isheden**, **Anna Johansson**, **Myeongjee Lee**, **Daniela Mariosa**, **Zheng Ning**, **Henrik Olsson**, **Þorgerður Pálsdóttir**, **Nikolaos Skourlis**, **Rickard Strandberg**, **Peter Ström**, **Chen Suo**, **Robert Szulkin** and **Marco Trevisan**.

I would also, like to add: **Kat Bokenberger**, **Julien Bryois**, **Flaminia Chiesa**, **Sophie Debonneville**, **Isabella Ekheden**, **Miriam Elfström**, **Malin Ericsson**, **Tong Gong**, **Alessandra Grotta**, **Ida Karlsson**, **Anna Kähler**, **Frida Lundberg**, **Behrang Mahjani**, **Nelson Ndegwa Gichora**, **Anna Plym**, **Cecilia Radkiewicz**, **Sven Sandin**, **Arvid Sjölander**, **Agnieszka Szwajda**, **Nghia Trung Vu**, **Fei Yang** and **Shuyang Yao**. You guys made my time at MEB great!

To my Pre-Dissertation Committee: **Shuang Hao**, **Rickard Strandberg** and **Peter Ström**. You accepted without any hesitation for which I'm very grateful – I know that this takes a lot of effort!

To **Bénédicte Delcoigne** for being such a strong and inspiring officemate. Now it's me trying to remember why we put ourselves through this. I hope we stay in touch and maybe go ice-skating next winter.

To **Johan Zetterqvist** for countless evenings with interesting discussions and for organising the advanced R group with me.

To **Elisabeth Dahlqwist** for being the complete opposite to me. Your continued success provides good food for thought.

To **Maya Alsheh Ali** for all the good company when working late.

Thank you **Therese Andersson** for being such a good friend and for all the support during my thesis writing. I owe you one.

Xingrong Liu – thank you for the company along the way!

To **Henric Winell** for our discussions on Emacs, R, overclocking and bios modding. Not all appreciates how cool this really is.

To **Gabriel Isheden** for all the interesting conversions we have had. I will take all my future financial advice from you.

Til **Tor-Arne Hegvik** jeg synes du er utenordentligt morsom og håper å råke deg hyppigere fremover.

To **Caroline Weibull** for being such a supportive person, you really improve the work atmosphere.

To **Erin Gabriel** for providing such energy and passion about our work. Do you think I will get a 'time and relative dimension in space' now?

To **Keith Humphreys** for all discussions on societal trends and for being such a caring colleague. I'm glad that you will be chairing my defence – I can think of no one better.

To **Alexander Ploner** for your wisdom, good advice and shared love of R and open source.

Robert Karlsson – thank you for being a good friend and for all the coding help. You are the person I go to when google falls short.

Camilla Ahlqvist – thank you so much for everything. For an absent-minded person like me, this would not have been possible without your support. I will have to pressure someone for your home address – retiring is no excuse for missing a dissertation party.

To **Marie Jansson** I think that it took me less than a week to figure out that it is you how make the world go round in the Biostat group. Thank you so much for all the help over the years!

To the **PubMEB** crew of years long past: **Elisabeth Dahlqwist**, **Laura Ghirardi**, **Johanna Holm**, **Andreas Jangmo**, **Jiayao Lei**, **Robert Karlsson**, **Favelle Lamb**, **Henrik Olsson**, **Shihua Sun**, **Emilio Ugalde**, **Vilhelmina Ullemar**, **Dylan Williams** and **Anne Örtqvist**.

To my friends in the **Running Club**, particularly **Anders Forss**, **Caitlin Clementine** and **Ulrika Zagai**. The running kept me sane and prevented certain obesity from my abuse of the vending machine.

To the **Prostate Cancer Group**, thank you for being such a knowledgeable and supportive group. A special mention to **Peter Ström**, **Henrik Olsson**, **Þorgerður Pálsdóttir**, **Kimmo Kartasalo**, **Alessio Crippa** and **Martin Eklund**. I have enjoyed working with you and hope to do that again!

◇ **Family**

To **Anna** my partner in crime and life – thank you so much for all the support. I'm sorry for all the late nights, we can go to France ~~or~~ and Italy whenever you like now.

To **Oskar**, **Sara**, **Axel**, **Arvid**, **Magnus** and **Mia** for always making me feel so welcome. I really look forward to relaxing with you at Hamnö this summer!

Thank you **Mamma** and **Pappa** for being so encouraging and always believing in me. This one was for you. To **Anna**, **David**, **Dedde** and **Karoline** for being the best family one could ever wish for. To **Alfred**, **Matilda**, **Alma** and **Arthur**: thank you for reminding me that scientific gains are built over generations. Your joy in learning has inspired me and I hope that this thesis might one day inspire you to keep learning even when you are as mature as your uncle.

Bibliography

[1] Christensen N Hansen HL Hertzum-Larsen R Johannesen TB Kejs AMT Khan S Ólafsdóttir E Petersen T Schmidt LKH Virtanen A Engholm G, Ferlay J and Storm HH. **Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.1 (28.06.2018)**. http://www.ancr.nu, 2019. Online; accessed 2019-03-25. 1, 3

[2] Fritz H. Schröder, Jonas Hugosson, Monique J. Roobol, Teuvo L. J. Tammela, Marco Zappa, Vera Nelen, Maciej Kwiatkowski, Marcos Lujan, Liisa Määttänen, Hans Lilja, Louis J. Denis, Franz Recker, Alvaro Paez, Chris H. Bangma, Sigrid Carlsson, Donella Puliti, Arnauld Villers, Xavier Rebillard, Matti Hakama, Ulf-Hakan Stenman, Paula Kujala, Kimmo Taari, Gunnar Aus, Andreas Huber, Theo H. van der Kwast, Ron H. N. van Schaik, Harry J. de Koning, Sue M. Moss, Anssi Auvinen, and ERSPC Investigators. **Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up**. *Lancet*, **384**(9959):2027–2035, December 2014. 1, 5

[3] E. a. M. Heijnsdijk, T. M. de Carvalho, A. Auvinen, M. Zappa, V. Nelen, M. Kwiatkowski, A. Villers, A. Páez, S. M. Moss, T. L. J. Tammela, F. Recker, L. Denis, S. V. Carlsson, E. M. Wever, C. H. Bangma, F. H. Schröder, M. J. Roobol, J. Hugosson, and H. J. de Koning. **Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data**. *Journal of the National Cancer Institute*, **107**(1):366, January 2015. 1, 5, 8

[4] Lam F Colombet M Mery L Piñeros M Znaor A Soerjomataram I Bray F Ferlay J, Ervik M. **Global Cancer Observatory: Cancer Today**. https://gco.iarc.fr/today, 2018. Lyon, France: International Agency for Research on Cancer. Online; accessed 2019-03-20. 3, 4, 5

[5] F Bray, J Lortet-Tieulent, J Ferlay, D Forman, and A Auvinen. **Prostate cancer incidence and mortality trends in 37 European**

countries: an overview. *Eur J Cancer*, **46**(17):3040–3052, November 2010. 3

[6] R. M. MERRILL, D. L. WEED, AND E. J. FEUER. **The lifetime risk of developing prostate cancer in white and black men**. *Cancer Epidemiol Biomarkers Prev*, **6**:763–8, 1997. 10. 3

[7] ROBERT KARLSSON, MARKUS ALY, MARK CLEMENTS, LILLY ZHENG, JAN ADOLFSSON, JIANFENG XU, HENRIK GRÖNBERG, AND FREDRIK WIKLUND. **A Population-based Assessment of Germline HOXB13 G84E Mutation and Prostate Cancer Risk**. *European Urology*, July 2012. 3

[8] J. I. EPSTEIN, W. C. ALLSBROOK, M. B. AMIN, AND L. L. EGEVAD. **The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma**. *Am. J. Surg. Pathol.*, **29**(9):1228–1242, Sep 2005. 4

[9] L. EGEVAD, B. DELAHUNT, J. R. SRIGLEY, AND H. SAMARATUNGA. **International Society of Urological Pathology (ISUP) grading of prostate cancer - An ISUP consensus on contemporary grading**. *APMIS*, **124**(6):433–435, Jun 2016. 4

[10] D. DANNEMAN, L. DREVIN, D. ROBINSON, P. STATTIN, AND L. EGEVAD. **Gleason inflation 1998-2011: a registry study of 97,168 men**. *BJU Int.*, **115**(2):248–255, Feb 2015. 4

[11] **TNM. Union for International Cancer Control.** https://www.uicc.org/resources/tnm. Online; accessed 2019-03-28. 5

[12] **Prostatacancer, Nationelltvårdprogram. Version: 4.1**. https://www.cancercentrum.se/samverkan/cancerdiagnoser/prostata/vardprogram/. Online; accessed 2019-04-03. 5

[13] K. MISTRY AND G. CABLE. **Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma**. *J Am Board Fam Pract*, **16**:95–101, 2003. 2. 5

[14] BENNY HOLMSTRÖM, MATTIAS JOHANSSON, ANDERS BERGH, ULF-HÅKAN STENMAN, GÖRAN HALLMANS, AND PÄR STATTIN. **Prostate specific antigen for early detection of prostate cancer: longitudinal study**. *BMJ (Clinical research ed.)*, **339**:b3537, 2009. 5

[15] T. NORDSTRÖM, M. ALY, M. S. CLEMENTS, C. E. WEIBULL, J. ADOLFSSON, AND H. GRÖNBERG. **Prostate-specific antigen (PSA) testing is prevalent and increasing in Stockholm County, Sweden, Despite no recommendations for PSA screening: results from a population-based study, 2003-2011**. *Eur. Urol.*, **63**(3):419–425, Mar 2013. 5, 11

[16] G. L. ANDRIOLE, E. D. CRAWFORD, R. L. GRUBB, S. S. BUYS, D. CHIA, T. R. CHURCH, M. N. FOUAD, C. ISAACS, P. A. KVALE, D. J. REDING, J. L. WEISSFELD, L. A. YOKOCHI, B. O'BRIEN, L. R. RAGARD, J. D. CLAPP, J. M. RATHMELL, T. L. RILEY, A. W. HSING, G. IZMIRLIAN, P. F. PINSKY, B. S. KRAMER, A. B. MILLER, J. K. GOHAGAN, P. C. PROROK, J. E. BURING, D. ALBERTS, H. B. CARTER, G. CHODAK, E. HAWK, H. MALM, R. J. MAYER, S. PIANTADOSI, G. A. SILVESTRI, I. M. THOMPSON, C. L. WESTHOFF, J. P. KAHN, B. LEVIN, D. DEMETS, J. R. O'FALLON, A. T. PORTER, M. M. ASHTON, AND W. C. BLACK. **Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up**. *J. Natl. Cancer Inst.*, **104**(2):125–132, Jan 2012. [PubMed Central:PMC3260132] [DOI:10.1093/jnci/djr500] [PubMed:22228146]. 5

[17] R. M. MARTIN, J. L. DONOVAN, E. L. TURNER, C. METCALFE, G. J. YOUNG, E. I. WALSH, J. A. LANE, S. NOBLE, S. E. OLIVER, S. EVANS, J. A. C. STERNE, P. HOLDING, Y. BEN-SHLOMO, P. BRINDLE, N. J. WILLIAMS, E. M. HILL, S. Y. NG, J. TOOLE, M. K. TAZEWELL, L. J. HUGHES, C. F. DAVIES, J. C. THORN, E. DOWN, G. DAVEY SMITH, D. E. NEAL, AND F. C. HAMDY. **Effect of a low-intensity PSA-based screening intervention on prostate cancer mortality: the CAP randomized clinical trial**. *JAMA*, **319**(9):883–895, 03 2018. 5

[18] R. PATAKY, R. GULATI, R. ETZIONI, P. BLACK, K. N. CHI, A. J. COLDMAN, T. PICKLES, S. TYLDESLEY, AND S. PEACOCK. **Is prostate cancer screening cost-effective? A microsimulation model of prostate-specific antigen-based screening for British Columbia, Canada**. *Int. J. Cancer*, **135**(4):939–947, Aug 2014. 5

[19] JONAS HUGOSSON, REBECKA ARNSRUD GODTMAN, SIGRID V. CARLSSON, GUNNAR AUS, ANNA GRENABO BERGDAHL, PÄR LODDING, CARL-GUSTAF PIHL, JOHAN STRANNE, ERIK HOLMBERG, AND HANS LILJA. **Eighteen-year follow-up of the Göteborg**

Randomized Population-based Prostate Cancer Screening Trial: effect of sociodemographic variables on participation, prostate cancer incidence and mortality. *Scandinavian Journal of Urology*, **0**(0):1–11, 2017. PMID: 29254399. 5

[20] HENRIK GRÖNBERG, JAN ADOLFSSON, MARKUS ALY, TOBIAS NORDSTRÖM, PETER WIKLUND, YVONNE BRANDBERG, JAMES THOMPSON, FREDRIK WIKLUND, JOHAN LINDBERG, MARK CLEMENTS, LARS EGEVAD, AND MARTIN EKLUND. **Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study**. *The Lancet Oncology*, **16**(16):1667 – 1676, 2015. 6, 11

[21] T. NORDSTROM, A. VICKERS, M. ASSEL, H. LILJA, H. GRONBERG, AND M. EKLUND. **Comparison Between the Four-kallikrein Panel and Prostate Health Index for Predicting Prostate Cancer**. *Eur. Urol.*, **68**(1):139–146, Jul 2015. 6

[22] CAROLYN M RUTTER, ALAN M ZASLAVSKY, AND ERIC J FEUER. **Dynamic microsimulation models for health outcomes: a review**. *Medical Decision Making*, **31**(1):10–18, 2011. 6

[23] JEANNE S MANDELBLATT, KATHLEEN A CRONIN, STEPHANIE BAILEY, DONALD A BERRY, HARRY J DE KONING, GERRIT DRAISMA, HUI HUANG, SANDRA J LEE, MARK MUNSELL, SYLVIA K PLEVRITIS, ET AL. **Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms**. *Annals of Internal Medicine*, **151**(10):738–747, 2009. 6

[24] CISNET. **Prostate Cancer Model Profiles**, 2019. Accessed: 2019-03-25. 7

[25] F. LOEVE, R. BOER, G. J. VAN OORTMARSSEN, M. VAN BALLEGOOIJEN, AND J. D. HABBEMA. **The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening**. *Comput Biomed Res*, **32**:13–33, 1999. 1. 7

[26] **CISNET - Michigan Prostate Cancer Model**. Accessed: 2016-09-02. 7

[27] A. TSODIKOV, A. SZABO, AND J. WEGELIN. **A population model of prostate cancer incidence**. *Stat Med*, **25**:2846–66, 2006. 16. 7

[28] R. GULATI, L. INOUE, J. KATCHER, W. HAZELTON, AND R. ETZIONI. **Calibrating disease progression models using population data: a critical precursor to policy development in cancer control**. *Biostatistics*, **11**(4):707–719, 2010. 8

[29] ROMAN GULATI, JOHN L GORE, AND RUTH ETZIONI. **Comparative effectiveness of alternative prostate-specific antigen–based prostate cancer screening strategies: model estimates of potential benefits and harms**. *Annals of Internal Medicine*, **158**(3):145–153, February 2013. 8

[30] J. CHILCOTT, S. HUMMEL, AND M. MILDRED. **Options Appraisal: Screening for prostate cancer. Report to the UK National Screening Committee.** Technical report, Sheffield University, 2010. [www.screening.nhs.uk]. 8

[31] CHUNHUAN LAO, CHARIS BROWN, PAUL ROUSE, RICHARD EDLIN, AND ROSS LAWRENSON. **Economic evaluation of prostate cancer screening: a systematic review**. *Future Oncology (London, England)*, **11**(3):467–477, 2015. 8

[32] J.A. ROTH, R. GULATI, J.L. GORE, M.R. COOPERBERG, AND R. ETZIONI. **Economic analysis of prostate-specific antigen screening and selective treatment strategies**. *JAMA Oncology*, **2**(7):890–898, 2016. 8

[33] A. NICHOLSON, J. MAHON, A. BOLAND, S. BEALE, K. DWAN, N. FLEEMAN, J. HOCKENHULL, AND Y. DUNDAR. **The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation**. *Health Technol Assess*, **19**(87):1–191, Oct 2015. 8

[34] S. SANGHERA, J. COAST, R. M. MARTIN, J. L. DONOVAN, AND S. MOHIUDDIN. **Cost-effectiveness of prostate cancer screening: a systematic review of decision-analytical models**. *BMC Cancer*, **18**(1):84, 01 2018. 8

[35] SOCIALSTYRELSEN. *Screening för prostatacancer. Rekommendation och bedömningsunderlag.* Socialstyrelsen, Stockholm, 2018. Accessed: 2018-10-29. 8

[36] SOCIALSTYRELSEN. *Screening för prostatacancer med PSA- prov. Hälsoekonomisk analys.* Socialstyrelsen, Stockholm, 2018. Accessed: 2018-06-18. 8, 25, 32

[37] JINJING LI AND CATHAL O'DONOGHUE. **A survey of dynamic microsimulation models: uses, model structure and methodology**. *International Journal of Microsimulation*, **6**(2):3–55, 2013. 9, 20

[38] STATISTICS CANADA. *MODGEN Version 10.1.0: Developer's Guide*. Statistics Canada, 2012. 9

[39] S. GRIBBLE AND A. CHERKASSKY. **OpenM++: open source microsimulation platform**, 2016. Accessed: 2016-05-26. 9

[40] OLIVER MANNION, ROY LAY-YEE, WENDY WRAPSON, PETER DAVIS, AND JANET PEARSON. **JAMSIM: A Microsimulation Modelling Policy Tool**. *Journal of Artificial Societies and Social Simulation*, **15**(1):8, 2012. 9

[41] LENNART FLOOD, FREDRIK JANSSON, THOMAS PETTERSSON, TOMAS PETTERSSON, OLLE SUNDBERG, AND ANNA WESTERBERG. **SESIM III– a Swedish dynamic microsimulation model**. *Handbook of SESIM, Ministry of Finance, Stockholm*, 2005. 9

[42] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. 9

[43] BART SMEETS AND IÑAKI UCAR. **simmer: Discrete-Event Simulation for R**, 2016. Accessed: 2016-05-26. 9

[44] SABINE ZINN. **The MicSim Package of R: An Entry-Level Toolkit for Continuous-Time Microsimulation**. *International Journal of Microsimulation*, **7**(3):3–32, 2014. 9

[45] FRANS WILLEKENS. *Multistate Analysis of Life Histories with R*. Springer, New York, NY, 2014. 9

[46] SABINE ZINN, JAN HIMMELSPACH, ADELINDE M UHRMACHER, AND JUTTA GAMPE. **Building Mic-Core, a specialized M&S software to simulate multi-state demographic micro models, based on JAMES II, a general M&S framework**. *Journal of Artificial Societies and Social Simulation*, **16**(3):5, 2013. 9

[47] ANTOINE FILIPOVIĆ-PIERUCCI, KEVIN ZARCA, AND ISABELLE DURAND-ZALESKI. **Markov Models for Health Economic Evaluation: The R Package heemod**. *ArXiv e-prints*, feb 2017. R package version 0.9.2. 9

[48] DEVIN INCERTI. *hesim: Health-Economic Simulation Modeling and Decision Analysis*, 2017. R package version 0.1.0. 9

[49] LOTTI BARLOW, KERSTIN WESTERGREN, LARS HOLMBERG, AND MATS TALBÄCK. **The completeness of the Swedish Cancer Register – a sample survey for year 1998**. *Acta Oncologica*, **48**(1):27–33, 2009. 11

[50] J. A. NELDER AND R. MEAD. **A Simplex Method for Function Minimization**. *Computer Journal*, **7**(4):308–313, 1965. 12, 21

[51] S. TAVARÉ, D. J. BALDING, R. C. GRIFFITHS, AND P. DONNELLY. **Inferring coalescence times from DNA sequence data**. *Genetics*, **145**(2):505–518, Feb 1997. 12, 20

[52] JAUHIAINEN ALEXANDRA CLEMENTS, MARK AND ANDREAS KARLSSON. **Prostata package for R**, 2018. Accessed: 2018-02-18. 15

[53] N. E. DAY AND S. D. WALTER. **Simplified models of screening for chronic disease: estimation procedures from mass screening programmes**. *Biometrics*, **40**:1–14, 1984. 1. 17, 18, 19

[54] S. D. WALTER AND N. E. DAY. **Estimation of the duration of a pre-clinical disease state using screening data**. *Am J Epidemiol*, **118**:865–86, 1983. 6. 17

[55] R. D. ETZIONI, R. J. CONNOR, P. C. POROK, AND S. G. SELF. **Design and analysis of cancer screening trials**. *Stat Methods Med Res*, **4**:3–17, 1995. 1. 18

[56] AVERILL LAW. *Simulation Modeling and Analysis (McGraw-Hill Series in Industrial Engineering and Management)*. McGraw-Hill Education, 5 edition, 1 2014. 20

[57] NATASHA K. STOUT AND SUE J. GOLDIE. **Keeping the noise down: common random numbers for disease simulation modeling**. *Health Care Management Science*, **11**(4):399–406, December 2008. 20

[58] C. GOURIEROUX, A. MONFORT, AND E. RENAULT. **Indirect Inference**. *Journal of Applied Econometrics*, **8**(3):S85–S118, 1993. 20

[59] J. J. KIM, K. M. KUNTZ, N. K. STOUT, S. MAHMUD, L. L. VILLA, E. L. FRANCO, AND S. J. GOLDIE. **Multiparameter calibration of a natural history model of cervical cancer**. *Am J Epidemiol*, **166**:137–50, 2007. 2. 21

[60] STEVE BROOKS, ANDREW GELMAN, GALIN JONES, AND XIAO-LI MENG. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011. 23

[61] PAUL MARJORAM, JOHN MOLITOR, VINCENT PLAGNOL, AND SIMON TAVARÉ. **Markov chain Monte Carlo without likelihoods**. *Proceedings of the National Academy of Sciences*, **100**(26):15324–15328, 2003. 24

[62] D. WEGMANN, C. LEUENBERGER, AND L. EXCOFFIER. **Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood**. *Genetics*, **182**(4):1207–1218, Aug 2009. 24, 25

[63] M. A. BEAUMONT, W. ZHANG, AND D. J. BALDING. **Approximate Bayesian computation in population genetics**. *Genetics*, **162**(4):2025–2035, Dec 2002. 25

[64] M.G.M. HUNINK, M.C. WEINSTEIN, E. WITTENBERG, M.F. DRUMMOND, J.S. PLISKIN, P.P. GLASZIOU, AND J.B. WONG. *Decision Making in Health and Medicine*. Cambridge medicine. Cambridge University Press, 2014. 25

[65] **Kostnad per patient**. https://skl.se/ekonomijuridikstatistik/statistik/kostnadperpatientkpp/kp Online; accessed 2018-10-01. 25

[66] E. A. HEIJNSDIJK, E. M. WEVER, A. AUVINEN, J. HUGOSSON, S. CIATTO, V. NELEN, M. KWIATKOWSKI, A. VILLERS, A. PAEZ, S. M. MOSS, M. ZAPPA, T. L. TAMMELA, T. MAKINEN, S. CARLSSON, I. J. KORFAGE, M. L. ESSINK-BOT, S. J. OTTO, G. DRAISMA, C. H. BANGMA, M. J. ROOBOL, F. H. SCHRODER, AND H. J. DE KONING. **Quality-of-life effects of prostate-specific antigen screening**. *N. Engl. J. Med.*, **367**(7):595–605, Aug 2012. 26, 32

[67] J. KARNON, J. STAHL, A. BRENNAN, J. J. CARO, J. MAR, AND J. MOLLER. **Modeling using discrete event simulation: a report of the**

ISPOR-SMDM Modeling Good Research Practices Task Force–4. *Value Health*, **15**(6):821–827, 2012. 32

[68] G. DRAISMA, R. POSTMA, F. H. SCHRODER, T. H. VAN DER KWAST, AND H. J. DE KONING. **Gleason score, age and screening: modeling dedifferentiation in prostate cancer**. *Int J Cancer*, **119**:2366–71, 2006. 10. 37

[69] LINDA ABRAHAMSSON AND KEITH HUMPHREYS. **A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density**. *Statistical Methods in Medical Research*, **25**(4):1620–1637, 2016. 38