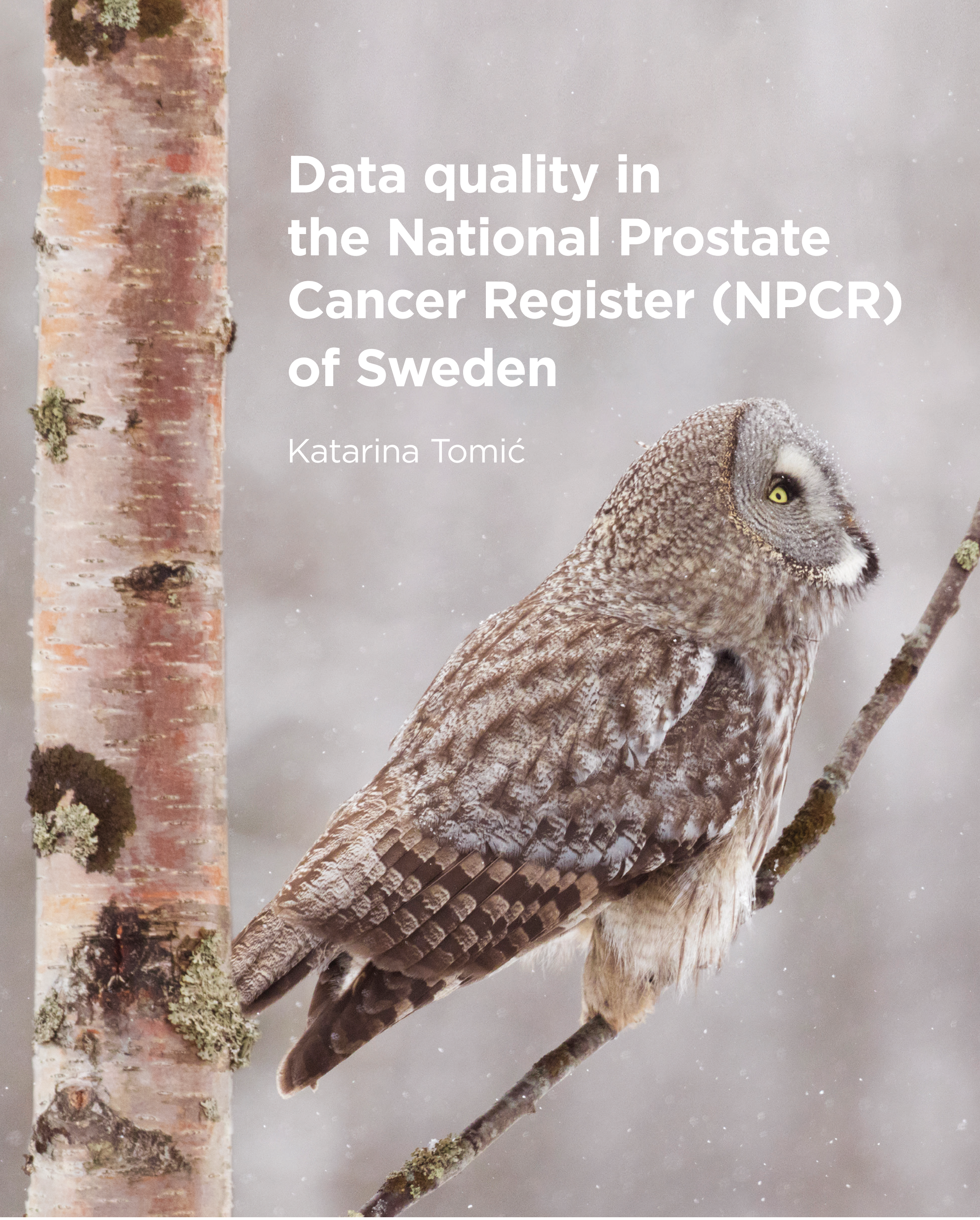


Data quality in the National Prostate Cancer Register (NPCR) of Sweden

Katarina Tomić

UMEÅ UNIVERSITY

UMEÅ UNIVERSITY

Data quality in the National Prostate Cancer Register (NPCR) of Sweden

Katarina Tomić

Department of Surgical and Perioperative Sciences
Urology and Andrology
Umeå 2018

Dissertation for PhD
ISBN: 978-91-7601-831-6
ISSN: 0346-6612
Umeå University Medical Dissertations, New Series No 1945
Cover photo: *Strix nebulosa* © Copyright Radiša Tomić
Electronic version available at: http://umu.diva-portal.org/
Printed by: UmU Print Service, Umeå University
Umeå, Sweden 2018

To my family

Table of Contents

Abstract

Background: Data in quality registers are increasingly used for quality assurance of health care, benchmarking, and research. If valid conclusions are to be drawn from such studies, it is vital that register data have high quality. The aim of this thesis was to assess data quality in the National Prostate Cancer Register (NPCR) of Sweden, a nationwide register that since 1998 captures 98% of all cases of Prostate cancer (Pca) in Sweden. The proportion and characteristics of Pca cases not registered in NPCR was investigated in paper I. Four dimensions of data quality were evaluated for NPCR in paper II: completeness, timeliness, comparability, and validity. Proportion and characteristics of Pca cases registered in NPCR but with unknown risk category were investigated in paper III. Finally, the association between Socioeconomic Status (SES) and Pca diagnosis, treatment, and mortality was studied in paper IV.

Material and methods: Data quality of NPCR was studied by cross-linkages between NPCR and other health care registers and demographical databases by use of the Swedish personal identity number. Validity was further studied by re-abstraction of patient health care records, followed by comparison of re-abstracted and original register data.

Results: Men not registered in NPCR, who constituted around 2% of all cases in the Swedish Cancer Register, differed only modestly in characteristics from cases in NPCR, indicating that NPCR is generalizable for all men with Pca in Sweden. Data quality in NPCR was high overall, with high completeness compared to the Swedish Cancer Register with registration mandated by law and few Pca cases were detected by use of death certificates. There was timely registration, and good comparability with registration forms and coding routines that were compliant with international guidelines. Data validity was high with high agreement and correlation for key variables. Men with unknown risk category had, compared to men with known risk category, more often concomitant bladder cancer, higher comorbidity, and lower Pca mortality. Men with high SES had, compared to men with low SES, higher probability of Pca detected during health checkup, shorter waiting times for prostatectomy, and higher probability of curative treatment for intermediate and high-risk cancer. Pca mortality was lower in men with high SES than in men with low SES for high-risk cancer.

Conclusion: These results indicate that data quality in NPCR is high and that NPCR is population-based. There were consistent differences in

diagnostic and therapeutic activity according to SES despite an equal access tax-financed healthcare system in Sweden.

Abbreviations

ADT	Androgen Deprivation Therapy
ATC	Anatomical Therapeutic Chemical classification system
CCI	Charlson Comorbidity Index
CI	Confidence Interval
GnRH	Gonadotropin-Releasing Hormone
IARC	International Agency for Research on Cancer
ICD	International statistical Classification of Diseases and related health problems
INCA	Information Network for Cancer care
IQR	Inter Quartile Range
LISA	Longitudinal integration database for health insurance and labor market studies
MAR	Missing at Random
MCAR	Missing Completely at Random
MNAR	Missing Not at Random
NCCN	National Comprehensive Cancer Network
NPCR	National Prostate Cancer Register of Sweden
OR	Odds Ratio
Pca	Prostate cancer
PCBaSe	Prostate Cancer data Base Sweden
PREM	Patient Reported Experience Measures
PROM	Patient Reported Outcome Measures
PSA	Prostate Specific Antigen
RCC	Regional Cancer Center
SES	Socioeconomic Status
TNM	Tumor Node Metastasis classification
UICC	International Union Against Cancer
WHO	World Health Organization

List of papers

I. Tomic, K., Berglund, A., Robinson, D., Hjälm-Eriksson, M., Carlsson, S., Lambe, M., & Stattin, P. (2015). Capture rate and representativity of the National Prostate Cancer Register of Sweden. Acta Oncologica, 2015; 54(2), 158-163.

II. Tomic, K., Sandin, F., Wigertz, A., Robinson, D., Lambe, M., & Stattin, P. (2015). Evaluation of data quality in the National Prostate Cancer Register of Sweden. European Journal of Cancer, 2015; 51(1), 101-111.

III. Tomic, K., Westerberg, M., Robinson, D., Garmo, H., & Stattin, P. (2016). Proportion and characteristics of men with unknown risk category in the National Prostate Cancer Register of Sweden. Acta Oncologica, 2016; 55(12), 1461-1466.

IV. Tomic, K., Ventimiglia, E., Häggström, C., Robinson, D., Lambe, M., & Stattin, P. Socioeconomic status and diagnosis, treatment, and mortality in men with prostate cancer. Nationwide population-based study. International Journal of Cancer. 2018; In press. DOI 10.1002/ijc.31272.

Reprints were made with permission from the respective publisher.

Sammanfattning på svenska

Kvalitetsregister består av insamlade data om personer med en viss sjukdom eller personer som har genomgått en viss behandling. Dessa register används idag alltmer för att kvalitetssäkra sjukvården, jämföra vårdformer och vårdgivare, och för forskningsändamål. Data i kvalitetsregister måste ha hög kvalitet för att korrekta slutsatser ska kunna dras i dessa typer av studier. Syftet med denna avhandling var att undersöka datakvalitén i Nationella prostatacancerregistret (NPCR), som sedan 1998 registrerar 98% av alla fall av prostatacancer (Pca) i Sverige. I avhandlingen undersöktes karaktäristika bland män med Pca som inte registrerats i NPCR i delarbete 1. Fyra olika dimensioner av datakvalitet studerades i delarbete 2: täckningsgrad (hur stor del av alla fall med Pca som NPCR registrerar), inrapporteringshastighet (hur snabbt nya fall registreras), jämförbarhet (huruvida standardiserade registreringsmetoder och sjukdomskodning används) och validitet (huruvida registrerade data är korrekta). Karaktäristiska undersöktes i delarbete 3 bland män som registrerats i NPCR men för vilka riskkategorin hos cancern var okänd, det vill säga inte kunde bedömas på grund av inkomplett data gällande minst en av variablerna Gleason-summa, TNM-stadium, och/eller PSA-värde. Slutligen studerades sambandet mellan socioekonomiskt status (SES) och diagnos, behandling, och överlevnad av Pca i delarbete 4.

Datakvalitén i NPCR studerades genom registerlänkning av NPCR med andra nationella sjukvårdsregister och demografiska databaser. Information om varje man i dessa olika register länkades med hjälp av personnummer. Validitet studerades genom så kallad re-abstraktion, det vill säga att delar av NPCR återskapades genom att utvalda patientjournaler omregistrerades av oberoende rapportörer (tredje part), i detta fall sjuksköterskor, i en kopia av registret. Denna registerkopia jämfördes sedan med originalregistret för att undersöka om det fanns några systematiska fel i inrapportering eller kodning av information.

Överlag var datakvalitén hög i NPCR. Täckningsgraden var hög jämfört med Cancerregistret, där registrering av alla cancerfall i Sverige är lagstadgad och endast ett fåtal fall av Pca diagnostiserades med hjälp av dödsorsaksintyg. Inrapporteringshastigheten var också hög, exempelvis registrerades 76% av alla nya Pca fall inom 6 månader och 95% inom 12 månader. Det skedde även en förbättring av inrapporteringshastigheten över hela landet under den studerade tidsperioden, 2008–2012. Även jämförbarheten var hög i NPCR, då registreringsrutiner och kodning följde internationell standard, med användning av Gleason-summa för tumördifferentiering och TNM-systemet för tumörklassificering. Validiteten var hög, med hög

överensstämmelse och/eller korrelation för nyckelvariabler relaterade till diagnos, utredning och behandling. Exempelvis var korrelationen 1.0 för variabeln operationsdatum, korrelationen 0.97 för PSA och 97% av alla Gleason-summor överensstämde. Studien av icke-registrerade män visade på att dessa män överlag var jämförbara med de män som registrerats i NPCR avseende tumörkaraktäristika, behandling och överlevnad. Detta antyder att NPCR är generaliserbart för alla män med Pca i Sverige. Män i NPCR med okänd riskkategori hade högre komorbiditet, speciellt avseende blåscancer, och de hade lägre dödlighet i Pca jämfört med män med känd riskkategori. Dödligheten av andra orsaker skiljde sig inte markant mellan grupperna. Sista delarbetet visade att män med hög SES hade högre sannolikhet att få en Pca diagnos till följd av en hälsoundersökning. De hade även kortare väntetider till behandling med operation eller strålning, högre sannolikhet att få kurativ behandling vid mellan- och högriskcancer, och lägre dödlighet vid högriskcancer. Dessa skillnader var mest markanta vid jämförelse av män med hög respektive låg inkomst, men skillnaderna var statistiskt säkerställda även när högutbildade män jämfördes med lågutbildade.

Resultaten i denna avhandling visar att datakvalitén i NPCR är hög och att NPCR är ett populationsbaserat register eftersom varken män med Pca som inte registrerats i NPCR eller män i NPCR med okänd riskkategori hade några signifikanta avvikelser från övriga män i NPCR. Det fanns genomgående skillnader i diagnostik och behandling beroende på SES, trots att sjukvården i Sverige är skattefinansierad och allmänt tillgänglig.

Background

Cancer registers

Cancer registers track all cancer cases in a country and are used to study trends in number of new cancer cases (incidence) and number of cancer-caused deaths (mortality), to assess the development of cancer diseases over time, as well as for research. Furthermore, cancer registers can be used to control activities such as screening, to evaluate efficiency of cancer prevention, to detect differences across geographical areas, as well as for international comparisons (1, 2). Key statistics from cancer registers include incidence, mortality, and number of individuals living with cancer (prevalence), as well as time-trends. As cancer registers also hold sensitive data about individuals, registers are most commonly operated by governmental agencies or other organizations with similar function. The permission to use register data for research or other purposes is often subject to ethical approval by a board appointed by the register holder in addition to an approval from an ethics review board.

Cancer quality registers

To complement general-purpose cancer registers where data is generic, cancer quality registers, also known as clinical cancer registers, contain more comprehensive, in-depth information on e.g., means of diagnosis, work-up and staging, cancer stage and histopathological differentiation, and treatment. This makes cancer quality registers a rich source of information for quality assurance and quality improvement through feedback to clinicians, including assessment if procedures recommended by national guidelines are used and benchmarking between health care providers. Cancer quality registers can also be used for surveillance of adverse events and epidemiologic research such as assessment of differences in diagnosis, treatment, and mortality in specific subgroups, as well as analysis of the impact of social and environmental factors on cancer risk. The usefulness of cancer quality registers depend on high data quality, which is required if valid conclusions are to be drawn based on these registers (3).

Swedish national registers

Sweden has a long tradition of population based registers and as early as 1749, *Tabellverket,* a register covering the entire population of Sweden was established. A large number of registers are today constituting important resources on areas such as population, health and medical care, social insurance, social services, education and research, household finances, labor market, prices and consumption, public finances, environment, energy, democracy, etc.

Swedish quality registers

Sweden is world leader in development of health care quality registers. Swedish registers provide a wealth of data and are a unique resource for improvement of health care and for research (4). Some registers are organized around a certain diagnosis, such as the National Diabetes Register (5), whereas others are based on certain treatments, e.g., the Swedish Hip Arthroplasty Register (6). The number of registers available for research is growing, in Sweden and in other countries. This trend is timely coordinated with the development of digital tools for online data registration and validation, as well as advances in statistical methods to compensate for differences between cohorts in register-based research studies. Thanks to the combination of increased number of registers, better tools for registration and data validation, and improved statistical methods, register-based research is gaining momentum (7, 8).

In 2012, the Swedish government launched a five-year effort to further develop quality registers in Sweden. The aims included improved data quality, improved analysis and feedback to support quality improvement efforts, increased access to, and use of, quality registers for research and innovation, increased openness and access to data, including for patients, and improved healthcare provided on more equal terms (7). Four certification levels were introduced for assessment of quality registers. These levels are associated with increasing requirements regarding register organization, operation and use, including data collection, feedback to health care providers, and research, as well as aspects related to validation of data quality (8). In more details:

- *Register candidates* have been identified as highly relevant, established connection with patient groups, etc. appointed a steering group, approved the register design, and have an associated public body.

- *Level 3* registers have, in addition, a register design according to national standard, the ability to register data, statistics that can be summarized centrally, approved feedback and analysis capabilities, and have initiated work with feedback to reporting units.

- *Level 2* registers have, in addition, coverage of a large proportion of cases, online feedback to reporting units, open reporting of data, e.g., in annual reports, and have been used for quality improvement work and research.

- *Level 1* registers have, in addition, been used to improve health care, for research, and to secure research funding in competition, have information about results for patients online, and have validated data quality.

In 2016, 13 quality registers had reached the highest certification level, including the National Prostate Cancer Register (NPCR).

Cancer registers and cancer quality registers internationally – a comparison

Swedish registers have in common that they are powerful tools and unique resources for information and research in order to provide better public health and welfare. Sweden together with other Nordic countries possess unique health care registers and databases with high quality, making it possible to be world-leading in register-based research (9). An important tool in Sweden that makes linkages of registers possible is the unique Swedish personal identity number. It enables almost 100% coverage in many Swedish registers (10).

In contrast 85% of the world population lacks quality cancer registers and one-third of the European population is in the same situation (11). In Europe, cancer registration is challenged by large differences in cancer register completeness and data quality. Further challenges include insufficient harmonization and comparability of procedures and data, differences in legislation that limits the use of cancer registers in international networks, collaboration, and research. These differences reflect the large variations among European nations in economical, regulatory, social, and cultural aspects.

This issue has been addressed at European Union level in the recent years with some promising results. Challenges has been investigated, a set of best practices and recommendations has been formulated, along with policy frameworks and tools for collaborations and information sharing. Despite these advances, with cancer now being the second most common cause of death in Europe, one third of the European population still lacks quality cancer registration, most of these living in the regions with lowest resources and health status. Cancer registers have a key role for this growing problem and continued development of cancer registers is thus needed, but equally important is that the knowledge obtained from the registers is transformed into actions.

Data quality

Cancer quality registers have many uses, and it is thus crucial that register data is of high quality and monitored. Information generated by cancer registers must be comparable and reproducible in order to accomplish validity. For example, conclusions in research studies are only valid if the study is based on complete and high quality data. There are four standard key dimensions of data quality (12, 13):

Completeness – the proportion of cases registered by the population-based register as a fraction of the total number of cases in the population.

Timeliness – the rate of which registration is performed, commonly studied for registration of new cases.

Comparability – whether standardized collection practices and codification of cancer data such as tumor characteristics are used, which can be assessed through a comprehensive review of the used registration routines. Comparability is essential for comparison and interpretation of data.

Validity – the proportion of the identified cancer cases with a certain property that actually have this attribute, i.e., whether the registered data actually is correct. Validity can be evaluated through comparisons within the register, cross-linkages to relevant registers that contain the same data, or to specific subsets of cases. Other methods to evaluate validity include re-abstraction and recoding from original data sources such as patient records, analysis of missing information, and internal consistency evaluation methods.

The International Agency for Research on Cancer (IARC) have defined guidelines for evaluation of data quality, including methods to assess these four key dimensions of data quality (14). Data quality assurance is a term used to describe a broad set of planned and systematic procedures to guarantee the quality of register data. These procedures take place before, during, and after data collection. Examples of procedures are training of health care professionals, including staff for data registration, in registration routines and systems, as well as follow-up, e.g., annual reporting of register data quality to involved registration staff. These procedures can be complemented by automated error checking for data registration that can prevent implausible values to be entered for particular variables where the plausible ranges are known. The most common causes of errors in registration (leading to incomplete records and/or data with poor validity) have been surveyed and a framework with procedures to improve data quality has been proposed (15).

One overall objective with any quality register is to achieve high completeness as this improves data quality and thus also the results of studies based on the register. A second goal is to ensure that there is no systematic non-registration, e.g., significantly lower completeness for certain clinics or health care regions, or worse, for cases with certain characteristics. The latter to ensure that all patients receive health care on equal terms, and that research reports based on register data are not biased due to exclusion of certain subgroups of cases. Timeliness can be more important for health care quality improvement and rapid follow-up than for research, where inclusion up recent data may be less important in a large register that spans many years and thousands of cases.

Missing data

Virtually all registers have missing data that potentially can affect data analysis (16, 17). Correct assessment and handling of data quality is important to avoid bias in research. Cancer quality registers can have missing data for multiple reasons, e.g., due to data missing from sources such as patient records, due to incorrect registration and coding, or that the register structure is updated over time to include more detailed information about the disease or treatment, resulting in incomplete data for older records.

Ethical aspects

Data in cancer quality registers are sensitive as they contain detailed health data about individuals that cannot be made publicly available for integrity reasons. It is important that the register and staff working with it maintain confidentiality. In Sweden, all health care quality registers must have a public authority responsible for management of personal data about individuals (*Swedish: Centralt personuppgiftsansvar*). In creation of laws and regulations, the rights of individual patients must be weighed against the benefits of the population at large. A scenario where only the patient and the treating medical staff have knowledge of a cancer would imply integrity, but at the expense of not being able to use the experience of other cases and contribute to the common knowledge about the disease (18).

Ethical questions in the use of quality register data in research have become increasingly important as modern databases evolve. With population-based cancer quality registers and the ability to link to other population-based registers, researchers can obtain detailed information about characteristics, not only about groups, but also of individual cases. It is thus critical that patient integrity is preserved at all times. This includes both laws and regulations for use of data in cancer quality registers (19) as well as routines and mechanisms for information security, e.g., pseudo-anonymizing cases by replacing any national identification numbers with identifiers that cannot be traced to a particular person (20). However, to preserve anonymity a register cannot use a code key that enables the personal identity number to be retrieved from anonymous identifiers, and data linkages must be planned with care to avoid that these reveal data about individual persons.

To handle the trade-offs between integrity and usefulness, access to quality register data for research purposes is subject to approval by an ethical board that assesses the risks and benefits of the proposed study. In applying for access, researchers must present a research plan and in advance describe what variables they need to access and for what purposes.

Aims of the thesis

The general aim of this thesis was to investigate the data quality in the National Prostate Cancer Register (NPCR) of Sweden. The specific aims for each paper were:

Paper I: To investigate completeness and whether Prostate cancer (Pca) cases not registered in NPCR differ in comorbidity, management, and mortality compared to cases registered in NPCR. Two categories of men with Pca not registered in NPCR were studied, men registered in the Swedish Cancer Register only and men detected by death certificate only.

Paper II: To investigate the validity of data in NPCR by comparing data in a re-abstraction of 731 patient charts with the original registration and by comparing registration in NPCR through cross-linkages with data in other health care registers.

Paper III: To investigate the proportion and characteristics of men in NPCR with unknown Pca risk category by identifying men who had unknown risk category due to missing diagnostic variables and to compare these cases with men who had known risk categorization regarding age, treatment, socioeconomic factors, comorbidity, and morbidity.

Paper IV: To investigate the differences according to Socioeconomic Status (SES) in diagnosis, diagnostic work-up, treatment, and mortality in men with Pca in NPCR.

Material and methods

Study settings and registers

The study population in this thesis is Swedish men with Pca. The registers used in the four papers to study quality in NPCR are the Swedish Cancer Register, the National Patient Register, the Prescribed Drug Register, the Cause of Death Register, and the Longitudinal Integration Database for Health Insurance and Labor Market Studies (LISA). These registers, summarized in Table 1, were all linked through the Prostate Cancer data Base Sweden (PCBaSe).

Table 1: Overview of registers used in this thesis

Register	Year of creation	Record structure	Validation studies	Register Holder
The National Prostate Cancer Register (NPCR) of Sweden	1998	Diagnosed cases of Pca in alive men, including data related to diagnosis, work-up, and treatment.	Paper II	Uppsala County Council
The Swedish Cancer Register	1958	Data on all reported cancer cases in Sweden, including personal data about the patient, diagnosis date, diagnosis details (clinical and morphological), and tumor extent.	3.7% underreporting according to Barlow et al. (21).	The National Board of Health and Welfare

The National Patient Register	1964 (nation-wide since 1987)	In-patient care data: patient information, admission and discharge dates, and medical data, including diagnosis and surgeries.	The positive predictive value varied between diagnoses, but was generally 85-95% (22).	The National Board of Health and Welfare
The Prescribed Drug Register	2005	All fillings of prescribed drugs, including information about the patient, the drug, and the prescribed drug. Excludes drugs administered during hospitalization.	Twenty disease-specific validation studies according to Wallerstedt et al. (23).	The National Board of Health and Welfare
The Cause of Death Register	1961 (exists older register for 1952-1960)	Data on all deaths among people registered in Sweden (inside and outside the country), including cause of death.	96% completeness (cases having specific cause of death recorded) and 77% agreement (86% for Pca) (24).	The National Board of Health and Welfare
LISA	2004	Record for each person aged 16 and above, including employment, disposable income, and education. Additional records for all companies and work places.	None	Statistics Sweden

The National Prostate Cancer Register (NPCR) of Sweden

The first Swedish regional Pca register was set up in 1987 in the South-East healthcare region. In 1998, all six health care regions in Sweden joined to register all incident cases of Pca, International Statistical Classification of Diseases and Related Health Problems codes (ICD), ICD-10 C619, and ICD-9 185 and NPCR was formed (1). The overall aim of NPCR is to assure good Pca care for all men with Pca, regardless of age and place of living. The register is thus used to document and benchmark the health care for men with Pca, which is essential for national quality improvement in clinical work. The national steering group for NPCR includes representatives from all six health care regions in Sweden and is led by Professor Pär Stattin, who is the NPCR chairman. The completeness of NPCR is 98% in comparison to the Swedish Cancer Register to which registration is mandatory and regulated by law (25).

Data registered in NPCR describe diagnostic work-up, tumor characteristics, and treatment. Four registration forms are currently used in NPCR: one form for diagnostic data, one for subsequent work-up and primary treatment, and separate forms for radical prostatectomy and radiotherapy. For diagnosis, examples of variables are date of diagnosis, diagnostic unit, and means of diagnosis. Tumor characteristics are described according to the Tumor Node Metastasis (TNM) classification. Differentiation is reported using the Gleason classification, including indicators of extent of cancer; number of biopsies obtained at diagnostic biopsy and number of cores with cancer, and total extent of the cancer in millimeters. Serum level of Prostate Specific Antigen (PSA) at date of diagnosis is also recorded. Primary treatment delivered within six months of date of diagnosis is recorded. Men who receive curative treatment are also asked to complete questionnaires regarding Patient Reported Outcome Measures (PROM) and Patient Reported Experience Measures (PREM) before start of treatment, as well as one, three, and five years after treatment (21, 26). Waiting times for different phases of diagnosis and follow-up are also recorded.

Annual reports from NPCR are publicly available and includes key data quality aspects of the register itself as well as characteristics of cancer cases. Data quality is studied in terms of completeness and timeliness of registration. Cancer characteristics include various aspects of diagnosis, follow-up and treatment of Pca, as well as patient waiting times during diagnosis and follow-up.

The annual report is complemented with real-time reports (*Swedish: Koll på läget*) for urology and oncology that are available online for involved health care professionals. In the real-time reports, a set of selected quality indicators are reported for each clinic, including timeliness in reporting of new Pca cases, patient waiting times after referral, and whether certain treatments and methods of diagnosis have been applied, e.g., active surveillance for very low-risk cancer and curative treatment for localized high-risk cancer. Another online report (*Swedish: RATTEN*) allows anyone to interactively browse the above described data from NPCR on Pca care in Sweden, and study data from certain years, counties, and health care providers (http://www.npcr.se/RATTEN).

Since 2007, all data in Swedish cancer quality registers are recorded with the Information Network for Cancer care (INCA) platform. Registration is performed by staff at each reporting unit and checked by staff at Regional Cancer Centers (RCC). The organization of the reporting with six RCCs, one in each health care region, enables close contact between register and reporting units and simplifies corrections and error checking in the registered data. The INCA platform was created to enable all cancer quality registers to use a common platform and thus collaborate in a structured manner, to enable fully digital registration without the use of paper forms, as well as to allow real-time extraction of data from each clinic for regional and national comparisons (27).

The Swedish Cancer Register

The Swedish Cancer Register was created in 1958, is nation-wide, and approximately 50,000 new cases of cancer are registered in Sweden each year. The Board of National Health and Welfare *(Swedish: Socialstyrelsen)* is responsible for the register. Registration is mandated by law and is since the mid 1980's performed by the six RCCs, previously Regional Oncological Centers. Cancer registration in Sweden is based on two mandatory independent reports, one from the responsible clinician who diagnosed the cancer case and the other from the pathologist who made the histopathological examination of the cancer tissue. In addition to the Swedish Cancer Register, the six RCCs host 28 cancer quality registers. Data in the Swedish Cancer Register is structured according to data about the patient (age, sex, place of residence, personal identity number), medical data (date and basis of diagnosis, reporting hospital and pathology/cytology department, tumor site, histological type, and stage), as well as follow-up data (date and cause of death, or date of migration) (28). Overall register quality is high, with approximately 99% of cases morphologically verified (29). An assessment of the completeness, using a comparison to the National Patient Register, concluded that the underreporting was around four percent. This underreporting, which is acceptable for most uses in research and health surveillance, was found to vary largely among clinics, increase with patient age, and also be overrepresented for diagnosis without verification by histology or cytology (21).

The National Patient Register

The National Patient Register covers all in-patient care in Sweden. The register was initiated in 1964 and registration became mandatory 1987. Day surgery is recorded since 1997 and all psychiatric and outpatient care delivered by non-primary health care units since 2001, when private health care providers also were included. These groups of cases are registered in the In-Patient and Out-Patient registers that are both part of the National Patient Register. The register is updated monthly since 2015 from each of the 21 Swedish county councils. Key variables include patients' personal identity numbers, hospital, diagnoses (main and contributing), and procedures (30).

Data quality controls are enforced for key variables - if the amount of incorrect data is above a threshold, new data are requested from the reporting clinic (30). The validity of the register was found to be high, varying from 85% to 95% among different diseases (29) and these numbers were later confirmed in an independent study (22). The National Patient Register is used to validate NPCR data regarding radical prostatectomy (KEC00, KEC01, KEC10, and KEC20) and surgical castration (KFC00, KFC10, and KFC15). In this thesis, discharge diagnoses from the In-Patient Register was used to assess comorbidity of Pca cases. Comorbidity was assessed according to the Charlson Comorbidity Index (CCI) that was developed to classify comorbid conditions that may alter the risk of mortality in longitudinal research (31). The CCI is a weighted sum of several comorbid factors such as diabetes, cardiovascular diseases, and various types of cancers, etc., where each disease adds a weight of 1, 2, 3, or 6 points depending on its severity.

The Prescribed Drug Register

The Prescribed Drug Register includes all filled prescriptions in Sweden since July 2005. Data include the prescribed drug, amount and daily dose, and date of prescription and date of filling (32). The register is limited to outpatient care, and thus excludes drugs administered to inpatients at hospitals and non-prescription drugs over the counter. The Prescribed Drug Register is extensively used for research (33). For register studies of Pca, it is commonly cross-linked to investigate treatments with androgen deprivation therapy with Gonadotropin-Releasing Hormone (GnRH) analogues, code

L02AE in the Anatomical Therapeutic Chemical Classification System (ATC), anti-androgens (ATC code L02BB), including Bicalutamide (ATC code L02BB03).

The Cause of Death Register

The Cause of Death Register records causes of death for all persons registered in Sweden from 1991, with an extension since 2012 to also include non-residents. Causes of death are encoded using ICD codes. Around one percent of all registered causes of death miss a death certificate and for slightly less than three percent, the cause of death is inconclusive (34).

Since 1911 cause of death determination has been mandatory in Sweden and the National Board of Health and Welfare publishes a statistical summary report “Cause of death” every year. A death certificate is since 1991 divided in two parts issued by a doctor, one death certificate (*Swedish: Dödsbevis*) that is sent to the Swedish Tax Agency as well as the Population Register, and another, cause of death certificate (*Swedish*: *Dödsorsaksintyg*). The latter is sent to the National Board of Health and Welfare that constitutes the base for cause of death statistics (34).

The validity of Pca as cause of death has been found to be high, with three percent more causes classified as deceased from Pca in the Cause of Death Register compared to reviewed medical records, as well as 86% overall agreement (35). In another study, an independent cause of death committee reviewed relevant medical data including death certificates according to a standardized algorithm. The overall agreement between cause of death recorded in the death certificates and determined by the committee was 96% (36).

The Longitudinal Integration Database for Health Insurance and Labor Market Studies (LISA)

The Longitudinal Integration Database for Health Insurance and Labor Market Studies (LISA), administered by Statistics Sweden, registers annually since 1990 the education level, income, and form of employment of all individuals in Sweden aged 16 or above. Additional information in this database includes details about companies and workplaces. With its rich content about education, employment, and alternative employment (studies, parental leave, unemployment, illness, etc.), the LISA database can be used to study socioeconomic factors associated with diseases (37). In this thesis, the LISA database was used to study the impact of education level and income on diagnosis, treatment, and mortality in Pca cases.

Data linkage

Cancer registers are often linked with other nation-wide health care registers and demographic databases. In such data linkage (record linkage), records in two or more datasets that describe the same individual are identified. This process is simplified if the same identification code is used in all registers, e.g., a national personal identity number, like in Sweden (10) but this is not always the case as such identifiers are not available in all countries. Integrity and protection of personal data is an important topic in data linkage for cancer quality registers, as identification codes used to link registers commonly include personal data and uniquely identifies a person.

Prostate Cancer data Base Sweden (PCBaSe)

In 2008, NPCR was linked to a number of other population-based registers by use of the Swedish personal identity number. The resulting database, Prostate Cancer data Base Sweden (PCBaSe), contains data about treatments, filling of prescribed drugs, education and income, and causes of death, allowing studies of treatment patterns, socioeconomic aspects, and mortality among men with Pca (Figure 1). For each Pca case, PCBaSe contains two (period 1987-1995) or five Pca-free control cases (period 1995-present) matched by year of birth and county of residence (38).

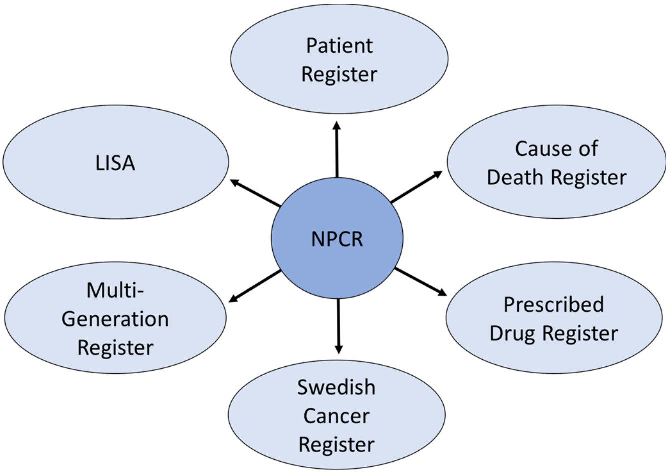


Figure 1. Overview of Prostate Cancer data Base Sweden (PCBaSe) that is based on men in the National Prostate Cancer Register (NPCR) of Sweden. In PCBaSe, NPCR has been enriched by information from a number of other health care registers and demographic databases through cross-linkages by use of the person identity number available for all Swedish residents.

The cross-linkages in PCBaSe are performed every three years allowing for more cases and longer follow-up. PCBaSe also includes two control series of men free of Pca at the date of diagnosis for the cases, as well as information on brothers of men diagnosed with Pca. This database allows for studies with case-control, cohort, or longitudinal case-only design, as well as assessment of aetiological factors, pharmaceutical prescriptions, and long-term outcomes (26).

Study design

This thesis is based on observational studies of men with Pca in Sweden during the defined study periods with the aim of evaluating data quality in NPCR. The four studies in the thesis are summarized in Table 2 and described in more detail below.

Table 2. Thesis at a glance. Summary of the study setting and design, and overall research questions in papers I-IV

Paper	**Aim**	**Study setting and design**	**Research question**	**Calendar years of Pca cases**
I	To compare characteristics of men with Pca registered versus not registered in NPCR	Comparison between NPCR and the Cancer Register	Have non-registered as compared to registered men in NPCR differences in Pca characteristics?	1998-2009
II	To study data quality of NPCR	Key quality dimensions	What is the completeness, timeliness, comparability, and validity of NPCR?	2009, 2008-2012
III	To study proportion and characteristics of men with unknown risk category	Comparison between known and unknown risk category	Have men in NPCR with unknown risk category versus known risk category different Pca characteristics?	1998-2012
IV	To study the pattern of care and outcome for men according to socioeconomic status	Comparison between education levels and also between income quartiles	Is socioeconomic status associated to Pca characteristics, treatment, and survival?	2007-2014

Paper I This study included all cases diagnosed with Pca in the Swedish Cancer Register between 1998-2009. Men diagnosed and registered before 1998 in the Swedish Cancer Register were excluded. Men registered between 1998 and 2009 with Pca as underlying cause in the Cause of Death Register, but who had neither been registered in NPCR nor in the Cancer Register, i.e. “death certificate only diagnosis” were identified by cross-linking to the Cause of Death Register and the Cancer Register, and were studied as a separate group. For a subgroup of men in the Cause of Death Register diagnosed in 2006-2009 who had a record of Pca treatment in the Prescribed Drug Register or in the In-Patient Register, the time from start of androgen deprivation therapy to date of death, i.e. time of treatment was used as a proxy for start of follow up in survival analysis.

Paper II This study evaluated the four dimensions of data quality in NPCR by comparing data from 731 re-abstracted patient charts with their original registration in NPCR in 2009 and by comparing registration in NPCR with data in other health care registers including the National Patient Register and the Prescribed Drug Register. Survival time was defined as the time from date of Pca diagnosis to the date of the death, emigration, or end of follow-up on December 31 2011, whichever event came first.

Paper III This study investigated the characteristics of men in NPCR with unknown Pca risk category by comparing these with men with known risk category. Risk category in NPCR is a composite variable based on the variables serum levels of PSA, clinical T-stage, N-stage, M-stage, and Gleason score and a modified version of the 2010 National Comprehensive Cancer Network (NCCN) risk categorization is used (Table 3). Cases with unknown risk category were missing PSA, TNM, and/or Gleason score, making it impossible to categorize their risk.

Table 3: Risk categories used in Paper III and IV

	T	N	M	Gleason score	PSA (ng/ml)
Metastasis					
either	*any*	*any*	M1	*any*	*any*
or	*any*	*any*	*any*	*any*	PSA > 100
Regional metastasis					
either	*any*	N1	not M1	*any*	PSA ≤ 100
or	T4	*any*	not M1	*any*	PSA ≤ 100
or	*any*	*any*	not M1	*any*	50 ≤ PSA ≤ 100
High-risk					
either	T3	not N1	not M1	*any*	PSA < 50
or	not T4	not N1	not M1	8-10	PSA < 50
or	not T4	not N1	not M1	*any*	20 ≤ PSA < 50
Intermediate-risk*					
either	not T3-T4	not N1	not M1	7	PSA < 20
or	not T3-T4	not N1	not M1	≤ 6	10 ≤ PSA < 20
Low-risk					
	T1-T2	not N1	not M1	≤ 6	PSA < 10

Modified version of The National Comprehensive Cancer Network (NCCN) risk classification (39).

* Men who had Gleason score 7 (4+3) cancer in more than 50% of the biopsies were classified as high-risk.

Paper IV This study assessed the differences according to SES in risk category at presentation, diagnostic work-up, cancer treatment, and mortality in men with Pca in NPCR. Risk categories were defined as shown in Table 3. SES was categorized according to income quartile, education level (low with < 10 years school, medium with 10 – 12 years schooling, and high with university of college studies), and marital status (married, never married, divorced/widowed, and unknown).

Statistical methods

The various statistical methods that have been applied in the four papers in this thesis are summarized below.

Paper I Univariable and multivariable logistic regression models with odds ratios and 95% Confidence Intervals (CI) were used to investigate the association of demographic factors and Pca characteristics to registration in NPCR. A competing risk analysis assessed Pca mortality and death due to other causes. Survival time was defined as the time from date of Pca diagnosis to the date of the death, emigration, or end of follow-up on 31 December 2011, whichever event came first.

Paper II The validity of NPCR data was assessed by agreement between original and re-abstracted data, with calculation of the exact agreement complemented with Pearson correlation for numerical variables and use of Cohen's kappa as a correlation metric for ordinal variables. The chi-square test was used to determine the significance of differences in timeliness, between health care regions as well as over time.

Paper III Univariable and multivariable logistic regression models with odds ratios and 95% Confidence Interval (CI) were used to investigate the association between unknown risk category and other factors. Absolute risk of Pca mortality and risk of death from other causes were assessed using competing risk analysis. The method of chained equations was used in the imputation study, resulting in 100 datasets with imputed values for unknown risk category. The imputation model was based on age and year of diagnosis, mode of detection, TNM, Gleason score, World Health Organization (WHO) grade (before 2000), serum PSA, risk category, any comorbid factors registered up to ten years prior to Pca diagnosis, type of concomitant cancer, Pca treatment, survival time from Pca diagnosis, and cause of death.

Paper IV Univariable and multivariable logistic regression models with risk ratios and 95% CI were used to investigate if men with low SES differed in Pca diagnosis, treatment, and mortality compared to men with high SES. In a competing risk analysis, absolute risk of Pca mortality and death due to other causes was assessed.

Ethical considerations

All studies were approved by the Research Ethics Board at Umeå University Hospital, Umeå, Sweden. All clinics that report to NPCR inform patients about this reporting, and patients may decline registration (opt-out principle) (40). Information to patients about use of NPCR data and their rights is also available online (http://www.npcr.se/hem/undersida-2/).

Main Results

Paper I

This study of men with Pca included 100,849 men registered in NPCR, 2,198 men registered in the Swedish Cancer Register but not in NPCR, and 1,929 men registered in neither register but diagnosed through death certificates only. There were substantial regional differences in completeness, with 7% of men in Stockholm registered in the Swedish Cancer Register only, compared to 1% or lower in all other regions, with the Northern region the lowest with 0.6%.

Men registered in the Swedish Cancer Register but not registered in NPCR were slightly older than men in NPCR, median age 72 versus 71 years, were less often treated with radical prostatectomy, 15%, versus 27%, had similar ten-year cancer mortality, 23% (95% CI 20-25) versus 24% (95% CI 24-25), as well as similar competing cause mortality, 28% (95% CI 26-31) versus 30% (95% CI 30-30). Men identified by death certificate only were older and had high comorbidity.

Treatment in the three groups of men for the last part of the study period (years 2006-2009) is shown in Table 4. There were minor differences regarding Androgen Deprivation Therapy (ADT) between men in NPCR and men registered in the Swedish Cancer Register only and men in NPCR were more often treated with prostatectomy. Men identified by death certificate only mainly received treatments indicating advanced Pca.

Table 4. Treatment for men diagnosed with prostate cancer in 2006-2009 and registered in the National Prostate Cancer Register (NPCR) of Sweden, men registered in the Swedish Cancer Register only, and men identified by death certificate only

	NPCR **N = 36,967**	**Swedish Cancer Register only** **N = 1,020**	**Death certificate only** **N = 698**
Treatment			
*Androgen deprivation therapy	13,691 (37%)	312 (31%)	395 (57%)
Radical prostatectomy	10,111 (27%)	151 (15%)	0 (0%)
**Other treatments	13,165 (36%)	557 (55%)	303 (43%)

* Information on Androgen Deprivation Therapy (ADT) was retrieved from The Prescribed Drug Register, which started on 1 July 2005. ADT included Gonadotropin-Releasing Hormone (GnRH) analogues and anti-androgens.

** Other treatments include radiotherapy, conservative treatment (i.e. no active treatment including active surveillance and deferred hormonal treatment), and transurethral resection of the prostate. Treatment had not been registered for 1,333 men.

Paper II

This study of data quality included 731 men diagnosed with Pca in 2009 for whom data from medical charts were re-abstracted and compared to their original registration in NPCR. Data on treatment were also validated by record linkage of NPCR with the Prescribed Drug Register and the National Patient Register. In the former register, fillings of GnRH analogues and anti-androgens were compared with NPCR data, and in the latter, records of radical prostatectomies and bilateral orchiectomies were studied.

The mean value for completeness of the 48 evaluated variables was 90% (range 60-100%). Timeliness increased substantially over time, with 95% of cases reported within 12 months of diagnosis in 2012, an increase from 77% in 2008. Comparability was good as NPCR was found to comply with national and international coding routines. Agreement and correlation between original and re-abstracted data was high overall, with 17 variables having agreement above 90% and correlation above 0.9. For four variables, agreement was below 80% and correlation below 0.8, whereas for the remaining 27 variables, agreement and correlation was in between these ranges. For example, the correlation for serum PSA was 0.97, the agreement for T-stage was 83%, for N-stage 95%, and for M-stage 73% (96% when adjusting for changes in the used International Union Against Cancer (UICC) classification, and for both radical prostatectomy and radiotherapy, the agreement was 96%. More than 95% of the androgen deprivation therapies registered in NPCR had a corresponding prescription filling. Two common causes for non-agreement was that the original data was unknown or that two adjacent categories were mixed up for categorical variables, both causes illustrated in Figure 2 for T-stage. The figure also shows, for N-stage where correlation was very low although agreement was very high, how a few miscategorized cases can have large impact on correlation for variables with few categories.

	T-stage	Re-abstraction								
		T0	T1a	T1b	T1c	T2	T3	T4	TX	Missing
Original	T0	0	0	0	0	0	0	0	0	0
	T1a	0	12	0	0	0	0	0	0	0
	T1b	0	1	6	0	1	0	0	2	0
	T1c	1	1	3	296	23	4	0	2	10
	T2	0	0	1	36	162	15	0	2	5
	T3	1	0	0	5	9	104	5	2	2
	T4	0	0	1	0	0	2	13	1	1
	TX	0	0	0	1	0	0	0	1	0
	Missing	0	0	0	0	0	0	0	0	0

Exact agreement = 83.3%, correlation = 0.749

	N-stage	Re-abstraction			
		N0	N1	NX	Missing
Original	N0	2	0	17	0
	N1	0	0	10	1
	NX	2	7	686	6
	Missing	0	0	0	0

Exact agreement = 95.0%, correlation = 0.085

Figure 2: Illustration of the agreement as well as all mismatching cases for ordinal variables T-stage and N-stage.

The correlation was high overall for continuous variables. Figure 3 shows common types of outliers: date of treatment decision (Figure 3, left) was in one case registered with three years difference from original data (a likely keyboard mistake by the person who performed the registration), and a few patients had implausible prostate volumes that exceeded 500 ml (Figure 3, right). These type of errors, although rare, have disproportionate influence on correlation. For example, when removing the two outliers with prostate volume greater than 500 ml, correlation was increased from 0.13 to 0.95.

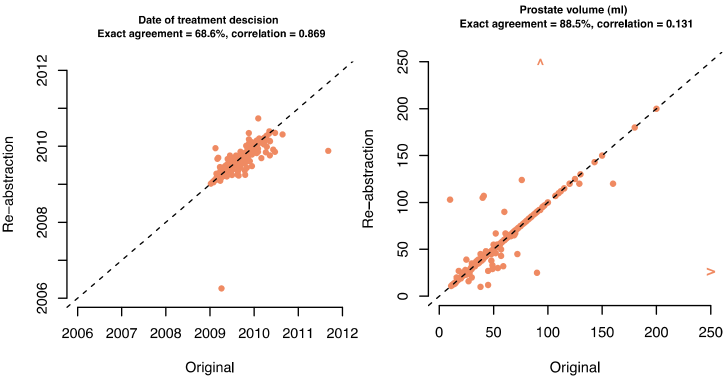


Figure 3: Correlation for continuous variables date of treatment decision (left) and prostate volume (right).

Paper III

This study included 3,315 men with Pca registered in NPCR with unknown risk category who were compared to 126,076 men in NPCR with known risk category. In men with unknown risk category, data were missing for serum PSA in 41%, for T-stage in 26%, and Gleason score in 13%, with 18% having two missing variables, and 2% all three variables missing (Figure 4).

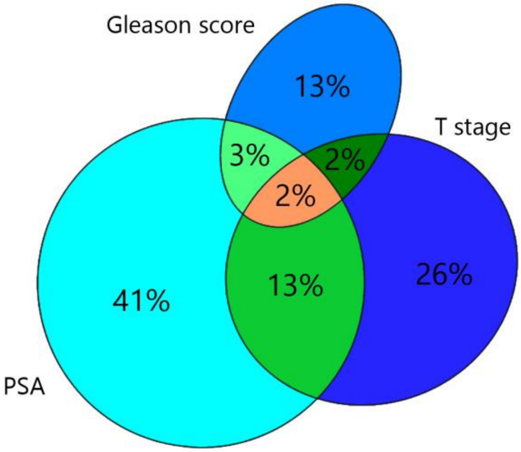
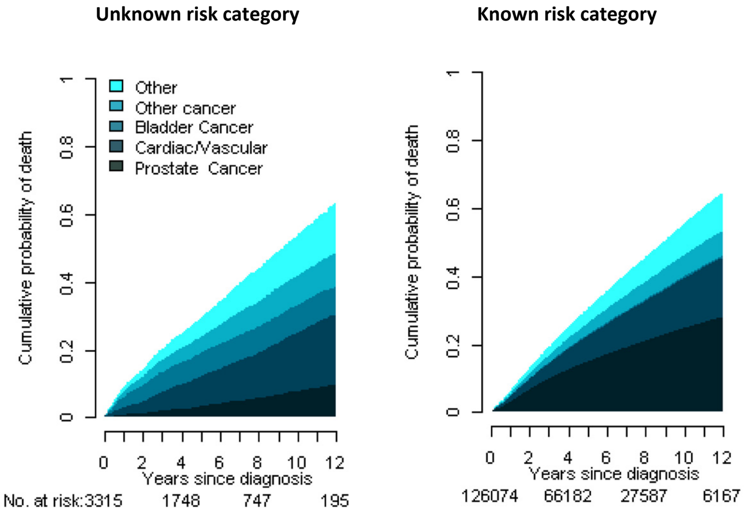


Figure 4: Distribution of missing data causing unknown risk categorization in the National Prostate Cancer Register (NPCR) of Sweden.

Men with unknown risk category were almost the same age at diagnosis as men with known risk category, mean 71 years versus 70 years, and were less often diagnosed following asymptomatic PSA testing, 13% versus 29%. Men with unknown risk category had higher comorbidity, 34% CCI 2 or higher versus 13%, in particular concomitant bladder cancer, 19% versus 1%, they less often received androgen deprivation therapy 9% versus 36% and more often conservative treatment 53% versus 26%. As shown in Figure 5, Pca mortality 12 years after diagnosis was lower in men with unknown risk category, 12% (95% CI 10-14%) versus 30% (95% CI 30-30%) but overall mortality was similar 57% (95% CI 54-59%) versus 57% (95% CI 57-58%).


Unknown risk category

Known risk category

Figure 5: Competing risk analysis of prostate cancer mortality and mortality from other causes for men with unknown (left) and known risk category (right) in the National Prostate Cancer Register (NPCR) of Sweden.

In the multivariable logistic regression model (Figure 6), unknown risk category was more likely for men with a concomitant bladder cancer diagnosis, for men receiving curative treatment or unspecified treatment, as well as for men with unspecified mode of detection and high comorbidity. Unknown risk category was not more likely for men with high age.

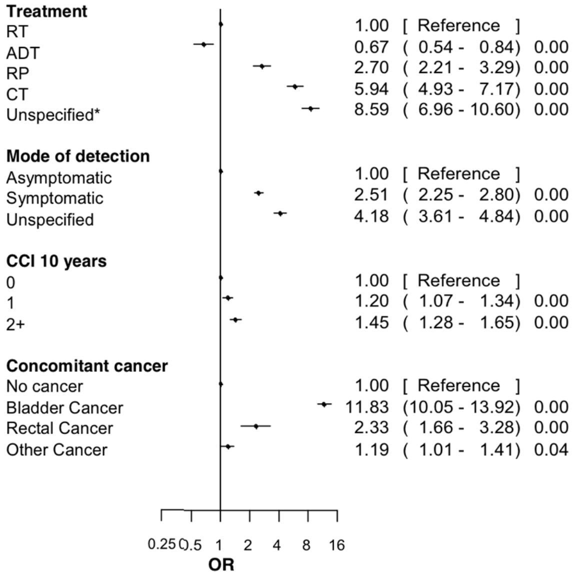


Figure 6. Multivariable risk of unknown risk category in the National Prostate Cancer Register (NPCR) of Sweden.

In the imputation analysis, the median risk proportion across all datasets where 49% (Inter Quartile Rate (IQR) 48-50) low-risk, 32% (IQR 30-33) intermediate-risk, 12% (IQR 11-12) high-risk, 6% (IQR 5-8) regionally metastatic, and 1% (IQR 1-2) distant metastatic Pca. This suggests that men with unknown risk category primarily had lower risk tumors.

Paper IV

This study of SES and Pca included 74,643 men registered in NPCR. These men were divided into three groups of educational status, low with < 10 years mandatory school, medium with 10-12 years schooling, and high with university or college education. They were also grouped into four income quartiles and by marital status. Of the men in the study, 54% were below age 70 at time of diagnosis, 74% had no known comorbid conditions (CCI 0), and 58% had low- or intermediate-risk Pca. At time of diagnosis, 66% were married and 35%, 39%, and 25% had low, intermediate, and high education level, respectively.

Men in the highest income quartile were, compared to men in the lowest income quartile, more likely to be diagnosed following health checkup, Odds Ratio (OR) 1.60 (1.45-1.77), less likely to be diagnosed with high-risk or more severe Pca, OR 0.57 (0.54-0.60), and less likely to have to wait more than three months for prostatectomy, OR 0.77 (0.69-0.86). Compared to men with lowest incomes, men with highest incomes were more likely to receive curative treatment for intermediate and high-risk Pca, OR 1.77 (1.61-1.95) and less likely to have positive margins after prostatectomy, OR 0.80 (0.71-0.90) (Figure 7). Similar, but less prominent differences were found for men with high education level compared to men with low education.

For men with no comorbidities and intermediate-risk cancer or regionally metastatic disease, there was a significantly lower Pca mortality for men with highest income compared to those with lowest. Men with high SES had lower all-cause mortality than men with low SES.

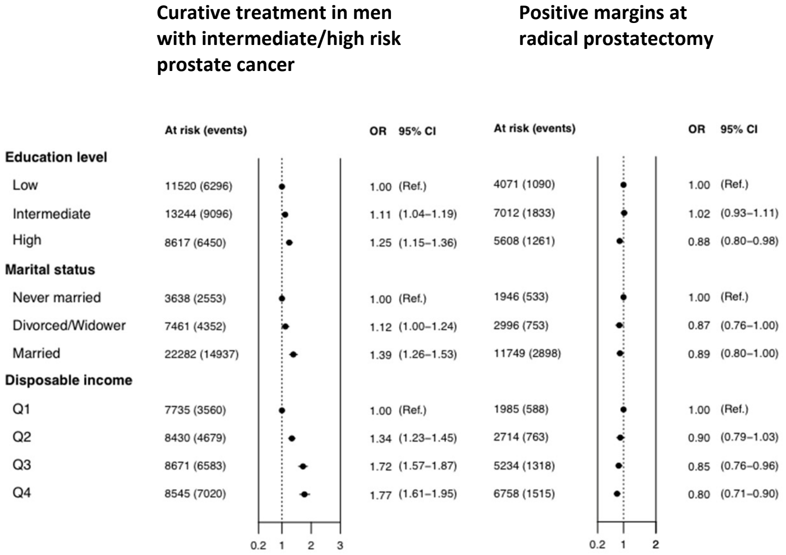


Figure 7. Odds Ratios (OR) for treatment strategies and treatment execution, 95% Confidence Intervals (95% CI) according to educational level and income.
Curative treatment: either radical prostatectomy or radiotherapy.
Q1 Lowest quartile of disposable income, Q4 highest quartile of disposable income.

Discussion

Strengths and limitations of register-based research

Register-based research differs from other methods such as randomized trials, and there are certain benefits and drawbacks of using cancer quality registers in research.

One benefit of register-based research is that data are already collected and thus the time consuming and costly task of data collection is avoided. Another benefit of register-based research is that the sample size is large, often with many thousands of cases recorded, sometimes even covering a complete population from a country. Such register completeness also allows detailed studies of subgroups e.g., living in a particular area, having certain comorbidities, or particular socioeconomic attributes. Register-based research can also be used for studies that would be unethical with randomized controlled trials, e.g., effects of use of certain drugs among pregnant women. Yet another benefit of using registers is that data has been gathered independent of the particular research study, which removes potential recall bias. Building on registers that tracks a population over decades also enables studies of diseases with long latency periods that only manifest themselves after many years, which would be unfeasible with randomized controlled trials. With data linkage across multiple registers, it is also possible to accurately adjust for some confounders, e.g., income, education, hospitalizations, filling of prescribed drugs, etc., where register-data often has higher validity than self-reported data (41).

One limitation with pre-collected data in registers is that required information can be unavailable, inaccurate, or misclassified (41), i.e., data selection and quality is outside the control of the researchers (42). Furthermore, register-based studies are also sensitive to data quality, e.g., differences in coding or introduction of updated coding systems can result in poor comparability. High data quality as well as knowledge about data quality in the used registers is required for valid conclusions. Similarly, missing data can be a limitation for several reasons, it may be unclear what the missing data actually represents, there can be under coverage, and statistical interpretation of results is cumbersome if missing data has non-random patterns.

One shortcoming of register-based observational studies compared to random control trials is that cases cannot be randomly allocated to cohorts in a register-based study. This can lead to confounding, e.g., when comparing different treatment options as treatment was not randomly allocated in cases in the register. Various statistical techniques can in part compensate for the non-random allocation in a register-based study, but random control trials are considered the golden standard for evaluation of treatment alternatives. Furthermore, as quality registers are based on information extracted from medical records, important confounding factors such as smoking and weight are commonly not recorded (43).

Another issue is left-truncation, as registers are truncated from the start of registration. This results in overestimating the incidence in the first years of registration unless only new diagnoses are included, and that the prevalence will be underestimated for slowly progressing diseases. In register-based research, the available data are often only proxies for what researchers really want to know, e.g., registers can be used to answer how much prescribed drugs were filled by patients, but not if or when they actually took their medications (44). Finally, the rich set of data available in quality registers may tempt researchers to focus too much on the available data, and not give enough attention to hypothesis formulation (41).

Paper I

In this paper, completeness was evaluated by data linkages to other registers. The results regarding completeness are much dependent on the completeness of the other registers used for the linkages, in this case, the Swedish Cancer Register, National Patient Register, Cause of Death Register, and Prescribed Drug Register. The quality of these have been shown to be good in earlier studies (21).

In the reverse linkage, 246 men out of 100,849 (0.2%) in NPCR where missing from the Swedish Cancer Register, which also indicate that completeness was high in NPCR, as registration to the Swedish Cancer Register in mandated by law and this register has high completeness.

As illustrated by the cases neither registered in the Swedish Cancer Register nor in NPCR but detected through death certificate only, sensitivity (avoidance of false negatives) affects register completeness (45). Most of these men received treatment for Pca and had thus been diagnosed with Pca and should have been registered in the Swedish Cancer Register as well as in NPCR. Among those not receiving any treatment, there could be both men who died before start of Pca treatment and those for whom Pca was diagnosed after death, through autopsy. The former group fulfil the NPCR inclusion criteria but the latter does not. Efforts to increase register sensitivity can be harmful to specificity (avoidance of false positives). In a quality register setting, sensitivity corresponds to completeness, and is defined as the number of registered cases divided by the total number of cases with the disease. The specificity is the disease-free population divided by the disease-free population plus any false positives, the latter being incorrectly registered cases that are disease-free. It is desirable to have both high sensitivity and specificity, but there is usually a trade-off between these. For example, if not yet verified persons in whom cancer is suspected after pathological analysis were to be registered, the result could be a quality register with many false positives, i.e., cases that do not have cancer. In this example, sensitivity is increased at the expense of specificity.

Our study identifies two significant factors for non-registration, geographical location and type of health care provider. In our study, clinics from larger cities were overrepresented in terms of non-registration. Similarly, private health care providers were shown to register a lower proportion of their patients, which also has been demonstrated in previous studies of quality register completeness (21). These factors could explain why Stockholm, the largest Swedish city with a high degree of private clinics, had the lowest completeness in our study.

There are different means to improve quality register completeness. Process improvements include aspects such as training of registration staff, e.g., through annual workshops as done in Sweden, as well as continuous feedback to staff. Simplification of the registration process includes improved user manuals and registration software aid for complicated variables, e.g., having common options pre-defined. Automated registration, i.e., direct import of register data from electronic health records is a promising method to improve completeness (46). This could enable fully automatic quality register registration (47) and remove encoding errors, but would at the same time remove additional plausibility checks by the registration staff. In Sweden, this approach would be complicated due to different electronic health record system used in different counties, and/or private clinics.

Paper II

In our study, comparability was found to be good as coding routines and registration forms were found to comply with international guidelines. Compared to the other dimensions of data quality, comparability is relatively easy to achieve for a quality register as there are standard classification systems that can be used, e.g., the TNM classification. However, caution is needed due to updates and changes in classification, as exemplified in our study where NPCR was updated to use a more recent TNM version from time of original registration to data re-abstraction. When classification and coding schemes are updated, registers can either leave old data as is or recode data for increased comparability. Recoding can introduce bias and errors by modification of historical data. Conversely, keeping original coding for old register data may over time result in a register with very low comparability - a data graveyard.

Timeliness varied between health care regions and improved over the duration of the study for all regions. This improvement coincided with the introduction of the INCA online registration platform that replaced paper-based registration. Automatic registration can be useful also to improve timeliness, as it would remove the burden of duplicated registration.

For evaluation of validity, the register cross-linkage method used in this thesis requires that the registers used for linkages in term have high validity. In this validation study, validity of the linked registers had been investigated in previous research and found to be high. For example, the Prescribed Drug Register has shown approximately 90% accuracy for discharge diagnosis (48) (49), the National Patient Register had a predicted positive value 85-95% in general in a meta-study (22), and for men with Pca, the Cause of Death register is very accurate (35, 36). An in-depth study of completeness of Swedish registers, as well as their work with validation was described by Emilsson et al. (50).

Closely related to automatic record abstraction are the various types of automated checks for valid data that can be applied in the quality register data registration software. Such checks could eliminate encoding errors, e.g., for categorical variables where the registration staff select among pre-defined options rather than entering free text. Furthermore, registration software systems could include range checks to avoid implausible values to be entered for continuous variables, in paper II exemplified by unrealistically large prostate volumes and errors in recorded date of treatment by three years.

Care must be taken when using standard correlation metrics such as Pearson correlation and/or Cohen's kappa for assessment of quality register validity. In paper II, correlation was very low for categorical variables with few options although a large majority of cases agreed and most non-agreeing cases belonged to an uninformative category (commonly "unknown"), as exemplified with M-stage and N-stage. Furthermore, outliers can drastically reduce the correlation of categorical variables, as illustrated in the results section with prostate volumes and dates. Ambiguous data recording instructions, e.g., how many decimals, if any at all, to include for variables such as serum PSA can also result in very low agreement, although correlation is still high. Similarly, small deviations in variables representing dates (of diagnosis, referral, treatment start, etc.) give very low agreement, although this has little impact on data quality. Aspects such as these must be considered when interpreting agreement and correlation results during quality register validation studies.

One potential drawback of this study is that cases for re-abstraction were not selected completely randomly for feasibility reasons, which could have introduced selection bias. To avoid this bias and ensure a good mix, cases were selected from private and public care providers, small and large units, and from three of the six geographical regions. In each visited unit, at least 20 re-abstractions of cases diagnosed in 2009 was performed. All public health care units and most private clinics have more than 20 cases per years, so this threshold did not introduce any major bias. Regarding potential regional differences, the only noticeable pattern regarding data validity was lower completeness in the Stockholm region that has many small private clinics.

Paper III

Increasingly fine-grained risk classifications are used, e.g., the recent introduction of a very low-risk category in NPCR, which also requires additional data to classify risk, in particular lower risk categories. This specialization along with differences between risk classification systems could reduce comparability between quality registers. Similarly, comparability over time is reduced, as older cases lack the additional data required to classify them according to more recent risk categorizations.

There are multiple approaches to handle missing data in clinical studies (51). One option is to exclude all cases with missing data, that is, to perform a complete case analysis. This simple method works well if few cases are missing data and no particular pattern exists for these. However, complete case analysis will reduce power if a large proportion of cases are excluded, and may introduce a selection bias if the group with missing data have different characteristics from the whole population, which is a common scenario (52). An alternative to complete case analysis is the indicator method, where cases with missing data are analyzed in a separate category (51). Another option, which is increasingly popular, is imputation, i.e., to estimate values for the missing data, and subsequently include cases with missing data in the full analysis (53).

Our imputation study showed that men with unknown risk category had lower risk category overall compared to men with known risk category. The implication of this is that lower risk category will be slightly underrepresented if complete-case analysis is performed and all men with unknown (but lower) risk category are excluded. Potential selection bias would be introduced only for complete-case analysis.

The comparison with cancer quality registers in other countries illustrated how the systematic work with data quality performed by NPCR and other Swedish registers (including dual reporting from clinician and pathologist, efficient communication and feedback between NPCR and reporting health care providers, etc.) can reduce the proportion of missing data for risk categorization by a factor of ten. This quality work also eliminated higher proportion of unknown risk category in subgroups, e.g., older men and/or men living in deprived areas as were exemplified in similar studies (54, 55).

A limitation of the imputation method is that imputation is most suitable if data is Missing at Random (MAR), i.e., data is missing for reasons that can be detected from other observed data. In case data is missing for reasons that are not observed, i.e., Missing Not at Random (MNAR), imputation can sometimes be performed but is more complicated. When data is Missing Completely at Random (MCAR), e.g. due to incorrect measurement equipment, complete case analysis gives equivalent results to imputation. An in-depth discussion of complete case analysis versus multiple imputation is given by White and Carlin who concluded that although no method is superior in the general case, multiple imputation performs better across a wider range of settings (51). Given knowledge of the used quality register, registration procedures, etc. it is possible to study each variable and try to determine the reason for missing data (56). In many cases, it can be complex to analyze why data is missing, and non-intuitive to determine correlations between missing and non-missing data. Thus, it is not possible to say in the general case whether imputation can be used for a particular data set. Understanding of both the data set at hand and the role of imputation is required.

Paper IV

Both income and education have been widely used in previous studies on social gradients in health care delivery and mortality (57). In our study, these measures yielded similar risk estimates with somewhat stronger associations for income. In the absence of data, we were unable to address the effects of other factors such as life style, health beliefs and awareness, and health care seeking behavior. In our study, men with low income received poorer treatment for Pca. Income is affected by retirement, with less marked differences after retirement, and as a majority of men in this study were above age of retirement this likely attenuated the association between income and pattern of care and mortality. Income is also affected by choice of occupation and age at retirement. In this study, age was compensated for in the analysis through logistic regression. An alternative explanation not investigated further in this work, is that men with low income could have received early retirement pension and/or other forms of sickness compensation. As comorbidity was higher among men with low income, it could have been that other diseases hindered these men to work (full-time), thus lowering their income.

Cancer characteristics include both biological features and stage at diagnosis. Previous studies suggested earlier detection in men with high SES, more frequent diagnosis following PSA testing, and thus potentially socioeconomic inequalities in survival (58, 59). A meta-analysis by Klein et al. surveyed 46 studies of Pca mortality and SES and concluded that 75% of these indicated lower survival in lower SES groups, which could in part be explained by a combination of comorbidity, stage at diagnosis, and treatment (59). Regardless of the reasons for inequalities in Pca care according to SES, our study highlights the importance of adherence to guidelines to ensure optimal and equal care for all cancer patients.

Conclusion

Data quality in NPCR is high and the register is truly nationwide and population-based. The small proportion of non-recorded cases differed insignificantly from registered cases, although completeness differed significantly among health care regions. Further, incompletely registered cases with unknown risk category were few and differed only marginally from completely registered men with known risk category. These findings indicate that NPCR has high data quality and is a reliable data source for research, benchmarking, and follow-up. Men with high SES received, compared to men with low SES, earlier diagnosis and better treatment of Pca and had lower Pca mortality among high-risk cases.

This thesis also highlights a few considerations when working with data quality evaluations – the need to understand register variables in-depth, including what they represent, and their use in clinical work as well as research. Some caveats associated with use of agreement and correlation to measure data validity for categorical variables with few options and continuous variables with significant outliers were also identified.

Future perspectives

The methods used in this thesis to validate register data quality are general and can be reused to evaluate data quality in other registers. Re-abstraction is time-consuming but can be performed based on patient health care records only. It is a good method for an initial validation study of a register as common mistakes in registration are identified, which can form a basis for future logical controls and other automatic methods to improve validity. Data linkages are powerful and potentially faster to perform than re-abstraction, but rely on existence of other registers with high data quality, and are also greatly simplified if common identifiers such as the Swedish personal identity number is used by all registers. Given these circumstances, external validity is high for this thesis.

Another topic related to reuse of the results from this thesis is how to track and improve data quality in a register over time. Instead of repeating costly re-abstraction projects, one promising method could be recurring, e.g., annually, data linkages followed by a combination of feedback to reporting clinics and registration staff, including updates to registration manuals, as well as introduction of logical controls and plausibility checks for variables with low validity. A study of Swedish quality registers summarizes the various data validation activities undertaken. In addition to the methods discussed above, common approaches include sample testing with outlier detection, various types of audits, and the ability to retrospectively revise incorrect data (50). A combination of re-abstraction, data linkages, plausibility checks, and such methods can provide a powerful framework for continuously improvement of data validity in quality registers. There are also multiple ways to improve feedback to reporting units in order to optimize quality assurance. Feedback should be rapid; units with completeness must be notified about this as soon as possible, instead of completeness only being reported in an annual report distributed to the units. Similarly, feedback regarding low validity should be specific and indicate what variable(s) the unit reported incorrectly and preferably also in what way. Recurring revisions of reporting instructions and training of registration staff is an important complement to specific feedback to reporting units.

Another future direction would be to study differences in data quality among health care providers more in-depth to investigate any differences by region, larger versus smaller units, private versus public providers, etc. For example, it would be interesting to investigate if the health care providers with low completeness in NPCR registration (paper I) also had higher proportion of cases with unknown risk category (paper III), although there were no similarities between the subgroups of unregistered men and men with unknown risk category.

Regarding differences in Pca management by SES, quality registers such as NPCR can have an important role in guaranteeing care on equal terms. In addition to highlighting inequalities as done in paper IV, the NPCR registration software could alert the person performing registration and other medical staff whenever cases are registered with characteristics that correspond to subgroups that are known to receive inferior treatment.

Acknowledgements

I am thankful to everyone that has been involved in this PhD work and in particular I want to thank:

My supervisor **Pär Stattin**, for being thorough, challenging, supportive, frank, and kind in excess. It has been a lot of fun. Your genuine interest in research and health care quality improvement inspires me.

My co-supervisor **Mats Lambe**, for providing insightful comments, new perspectives, and encouragement for independent work.

My co-supervisor **Christel Häggström**, for insightful comments and support discussing statistics and all other matters in life, large and small.

Börje Ljungberg, for encouragement, supportive attitude, and a research ambiance.

Radiša Tomić, for great support and enthusiasm, it is always exciting to discuss medicine and research with you.

I had the privilege collaborating with many researchers during this thesis. You have been a great source of knowledge and inspiration. I thank all my co-authors: **Anders Berglund**, **Fredrik Sandin**, **Marcus Westerberg**, and **Eugenio Ventimiglia** for your excellent work with statistical analysis in our papers, making my research even more enjoyable and a real team work. **Hans Garmo**, **David Robinson**, **Marie Hjälm-Eriksson**, **Stefan Carlsson**, and **Anette Wigertz** for additional statistical advice and valuable insights.

Colleagues in other disciplines for your input, in particular **Anders Widmark** for your enthusiasm.

My former fellow PhD students **Sara Wirèn** and **Jon Örn Fridriksson** for sharing your experiences and for your kindness.

Marie Eriksson, **Sofia Wård,** and **Anna-Maria Eriksson** for administrative support and nice company.

The staff at the Urology department, past and present. Always nice to work with you. **Amir Sherif** for your research support.

The staff at the Biobank for a friendly atmosphere and **Robin Myte**, for teaching me that LaTeX code is the real love code.

Regional Cancer Center Uppsala for a friendly and welcoming atmosphere.

Göran Hallmans, **Ulf Gunnarsson**, and **Michael Haney** for constructive feedback during my halfway PhD seminar.

Anders Ödin at ITS for computer support (tea included!).

My friends for sharing your interest in my PhD work.

My family – my **mother Lenče** and **father Radiša** for their constant support, love and great interest in science. My big **brother Kristjan.** Dear **Johan** with our children **Alexander**, **Julia** and **Novak**. I love you!

This work was supported by the Swedish Research Council, Swedish Cancer Society, Västerbotten County Council, and Lion's cancer research fund at Umeå University.

References

1. Dos Santos Silva I. Cancer Epidemiology: Principles and Methods: World Health Organization, International Agency for Research on Cancer; 1999.

2. Forsea AM. Cancer registries in Europe-going forward is the only option. Ecancermedicalscience. 2016;10:641.

3. Tomic K, Sandin F, Wigertz A, Robinson D, Lambe M, Stattin P. Evaluation of data quality in the National Prostate Cancer Register of Sweden. Eur J Cancer. 2015;51(1):101-11.

4. Rosen M. Guldgruvan i hälso- och sjukvården förslag tillgemensam satsning 2011-2015 2010.

5. Gudbjornsdottir S, Cederholm J, Nilsson PM, Eliasson B, Steering Committee of the Swedish National Diabetes R. The National Diabetes Register in Sweden: an implementation of the St. Vincent Declaration for Quality Improvement in Diabetes Care. Diabetes Care. 2003;26(4):1270-6.

6. Karrholm J. The Swedish Hip Arthroplasty Register (http://www.shpr.se/). Acta Orthop. 2010;81(1):3-4.

7. Nationella Kvalitetsregister. Satsningen på Nationella Kvalitetsregister 2016 [Available from: http://kvalitetsregister.se/tjanster/omnationellakvalitetsregister/satsning20122016.2009.html].

8. Nationella Kvalitetsregister. Certifieringsnivåer 2016 [Available from: http://www.kvalitetsregister.se/drivaregister/attredovisaochsokamedel/certifieringsnivaer.1943.html].

9. Vetenskapsrådet. Utilising the Nordic Gold Mines – Infrastructure for Register-based Research in Health and Welfare 2008 [Available from: https://vr.se/download/18.227c330c123c73dc586800013474/1340207480635/StatementhealthandwelfareFINAL.pdf].

10. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur J Epidemiol. 2009;24(11):659-67.

11. Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, et al. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. Int J Cancer. 2015;137(9):2060-71.

12. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer. 2009;45(5):756-64.

13. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 2009;45(5):747-55.

14. Parkin DM CV, Ferlay J, Galceran J, Storm HH, Whelan SL, editors. Comparability and quality control in cancer registration. . Lyon IARC (WHO) and IACR, 1994.

15. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002;9(6):600-11.
16. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. Yale J Biol Med. 2013;86(3):343-58.
17. O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol Ther. 2012;91(3):550-4.
18. Jensen OMP, D.M.; MacLennan, R; Muir, C.S.; Skeet, R.G. . Cancer Registration: Principles and Methods1991.
19. McCart-Snead S, Hillby A. Research Guide to European Data Protection Law. University of California at Berkeley, 2013.
20. QRC Stockholm. Handbok för kvalitetsregister med SLL som huvudman. 2016.
21. Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. Acta Oncol. 2009;48(1):27-33.
22. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11:450.
23. Wallerstedt SM, Wettermark B, Hoffmann M. The First Decade with the Swedish Prescribed Drug Register - A Systematic Review of the Output in the Scientific Literature. Basic Clin Pharmacol Toxicol. 2016;119(5):464-9.
24. Brooke HL, Talback M, Hornblad J, Johansson LA, Ludvigsson JF, Druid H, et al. The Swedish cause of death register. Eur J Epidemiol. 2017;32(9):765-73.
25. Tomic K, Berglund A, Robinson D, Hjalm-Eriksson M, Carlsson S, Lambe M, et al. Capture rate and representativity of The National Prostate Cancer Register of Sweden. Acta Oncol. 2015;54(2):158-63.
26. Van Hemelrijck M, Garmo H, Wigertz A, Nilsson P, Stattin P. Cohort Profile Update: The National Prostate Cancer Register of Sweden and Prostate Cancer data Base--a refined prostate cancer trajectory. Int J Epidemiol. 2016;45(1):73-82.
27. Regionala cancercentrum i samverkan. Om INCA 2017 [Available from: https://www.cancercentrum.se/samverkan/vara-uppdrag/kunskapsstyrning/kvalitetsregister/om-inca/].
28. National Board of Health and Welfare. Swedish Cancer Registry 2016 [Available from: http://www.socialstyrelsen.se/register/halsodataregister/cancerregistret/in english].
29. Swedish Association of Local Authorities and Regions. Appendix 2 - completeness. Quality and efficiency of Swedish Health Care 2011.
30. National Board of Health and Welfare. Information available in the National Patient

Register (NPR) 2016 [Available from: http://www.socialstyrelsen.se/SiteCollectionDocuments/information-in-the-national-patient-register.pdf].

31. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83.

32. Wettermark B, Hammar N, Fored CM, Leimanis A, Otterblad Olausson P, Bergman U, et al. The new Swedish Prescribed Drug Register--opportunities for pharmacoepidemiological research and experience from the first six months. Pharmacoepidemiol Drug Saf. 2007;16(7):726-35.

33. Wallerstedt SM, Wettermark B, Hoffmann M. The First Decade with the Swedish Prescribed Drug Register - A Systematic Review of the Output in the Scientific Literature. Basic Clin Pharmacol Toxicol. 2016.

34. National Board of Health and Welfare. Cause of Death 2017 [Available from: http://www.socialstyrelsen.se/statistics/statisticaldatabase/help/causeofdeath].

35. Fall K, Stromberg F, Rosell J, Andren O, Varenhorst E, South-East Region Prostate Cancer G. Reliability of death certificates in prostate cancer patients. Scand J Urol Nephrol. 2008;42(4):352-7.

36. Godtman R, Holmberg E, Stranne J, Hugosson J. High accuracy of Swedish death certificates in men participating in screening for prostate cancer: a comparative study of official death certificates with a cause of death committee using a standardized algorithm. Scand J Urol Nephrol. 2011;45(4):226-32.

37. Statistics Sweden. Longitudinal integration database for health insurance and labour market studies (LISA) 2017 [Available from: https://www.scb.se/en_/Services/Guidance-for-researchers-and-universities/SCB-Data/Longitudinal-integration-database-for-health-insurance-and-labour-market-studies-LISA-by-Swedish-acronym/].

38. Van Hemelrijck M, Wigertz A, Sandin F, Garmo H, Hellstrom K, Fransson P, et al. Cohort Profile: the National Prostate Cancer Register of Sweden and Prostate Cancer data Base Sweden 2.0. Int J Epidemiol. 2013;42(4):956-67.

39. Mohler J, Bahnson RR, Boston B, Busby JE, D'Amico A, Eastham JA, et al. NCCN clinical practice guidelines in oncology: prostate cancer. J Natl Compr Canc Netw. 2010;8(2):162-200.

40. NPCR. Patientinformation 2016 [Available from: http://npcr.se/hem/undersida-2/].

41. Thygesen LC, Ersboll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. Eur J Epidemiol. 2014;29(8):551-8.

42. Sorensen HT. Regional administrative health registries as a resource in clinical epidemiologyA study of options, strengths, limitations and data quality provided with examples of use. Int J Risk Saf Med. 1997;10(1):1-22.

43. Ray WA. Improving automated database studies. Epidemiology. 2011;22(3):302-4.

44. Olsen J. Register-based research: some methodological considerations. Scand J Public Health. 2011;39(3):225-9.
45. Baron J, Sørensen H, Sox HJ. Clinical epidemiology. In: Olsen J, Saracci R, D. T, editors. Teaching Epidemiology; A Guide for Teachers in Epidemiology, Public Health and Clinical Medicine 4th ed: Oxford University Press; 2015. p. 444-62.
46. Kreuzthaler M, Schulz S, Berghold A. Secondary use of electronic health records for building cohort studies through top-down information extraction. Journal of biomedical informatics. 2015;53:188-95.
47. Martinell M, Stalhammar J, Hallqvist J. Automated data extraction--a feasible way to construct patient registers of primary care utilization. Ups J Med Sci. 2012;117(1):52-6.
48. Ingelsson E, Arnlov J, Sundstrom J, Lind L. The validity of a diagnosis of heart failure in a hospital discharge register. Eur J Heart Fail. 2005;7(5):787-91.
49. Hammar N, Alfredsson L, Rosen M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. Int J Epidemiol. 2001;30 Suppl 1:S30-4.
50. Emilsson L, Lindahl B, Koster M, Lambe M, Ludvigsson JF. Review of 103 Swedish Healthcare Quality Registries. J Intern Med. 2015;277(1):94-136.
51. Jones MP. Indicator and stratification methods for missing explanatory variables in multiple linear regression. Journal of the American Statistical Association. 1996;91(433):222-30.
52. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med. 2010;29(28):2920-31.
53. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
54. Elliott SP, Johnson DP, Jarosek SL, Konety BR, Adejoro OO, Virnig BA. Bias due to missing SEER data in D'Amico risk stratification of prostate cancer. J Urol. 2012;187(6):2026-31.
55. Luo Q, Yu XQ, Cooke-Yarborough C, Smith DP, O'Connell DL. Characteristics of cases with unknown stage prostate cancer in a population-based cancer registry. Cancer Epidemiol. 2013;37(6):813-9.
56. Morisot A, Bessaoud F, Landais P, Rebillard X, Tretarre B, Daures JP. Prostate cancer: net survival and cause-specific survival rates after multiple imputation. BMC Med Res Methodol. 2015;15:54.
57. Adler NE, Ostrove JM. Socioeconomic status and health: what we know and what we don't. Annals of the New York academy of Sciences. 1999;896(1):3-15.
58. Cheng I, Witte JS, McClure LA, Shema SJ, Cockburn MG, John EM, et al. Socioeconomic status and prostate cancer incidence and mortality rates among the diverse population of California. Cancer Causes Control. 2009;20(8):1431-40.

59. Klein J, von dem Knesebeck O. Socioeconomic inequalities in prostate cancer survival: A review of the evidence and explanatory factors. Soc Sci Med. 2015;142:9-18.

Department of Surgical and Perioperative Sciences
Umeå University, SE-901 87 Umeå, Sweden
www.surgsci.umu.se/english

ISBN 978-91-7601-831-6
ISSN 0346-6612
New Series Number 1945

UMEÅ UNIVERSITY