From the DEPARTMENT OF MEDICINE, SOLNA
Karolinska Institutet, Stockholm, Sweden

PROSTATE CANCER PROGNOSTICATION BASED ON CLINICAL AND HISTOPATHOLOGICAL TUMOR FEATURES

Renata Zelić

Stockholm, 2020

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2020

ISBN 978-91-7831-923-7

Prostate cancer prognostication based on clinical and histopathological tumor features

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Renata Zelić

Principal Supervisor:
Associate Professor Andreas Pettersson
Karolinska Institutet
Department of Medicine, Solna
Division of Clinical Epidemiology

Co-supervisors:
Adjunct Professor Olof Akre
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Lorenzo Richiardi
University of Turin
Department of Medical Sciences
Cancer Epidemiology Unit

Opponent:
Professor Monique J. Roobol
Erasmus University Medical Center (MC)
Department of Urology

Examination Board:
Associate Professor Arvid Sjölander
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Professor Per-Uno Malmström
Uppsala University
Department of Surgical Sciences

Professor Martin E. Johansson
University of Gothenburg
Department of Laboratory Medicine

“I did then what I knew how to do. Now that I know better, I do better.”

— Maya Angelou

Abstract

Prostate cancer is the second most common cancer in men worldwide. Of almost 1.3 million newly diagnosed men per year, up to 80% will have localized disease with a characteristically prolonged natural history. Risk stratification and treatment decision-making for these men is currently based on the combination of standard clinical and histopathological predictors, such as the Gleason score, prostate specific antigen (PSA) level and clinical tumor stage at diagnosis. However, these standard predictors are not sufficient to capture the heterogeneity in prognosis for men with localized prostate cancer. As a consequence, these men are often overtreated and may suffer from treatment-related side effects. In this thesis we aimed to improve prognostication for men with localized prostate cancer through validation of existing risk stratification tools based on standard clinical and histopathological factors, and through validation of existing, and identification of novel, prognostic markers.

In Study I, we evaluated if the nested case-control study design is appropriate for estimating relative and absolute risks of dying from prostate cancer in the presence of competing risks. We used a case-control study (ProMort I) nested in the National Prostate Cancer Register of Sweden (NPCR). We found that the relative risks of dying from prostate cancer estimated in ProMort I were comparable to the relative risks estimated in the NPCR. The relative risks of dying from other causes estimated in ProMort I were biased, which led to biased estimates of the absolute risks of dying from prostate cancer. The bias in both the relative and absolute risks was reduced by augmenting competing-risks cases, and especially by augmenting both the competing-risks cases and the controls. Our results indicate that, without the additional extensions to the design, the nested-case control studies are not suitable for the development of models predicting death from prostate cancer in the presence of competing risks.

In Study II, we systematically compared the prognostic performance of the most commonly used pretreatment risk stratification tools in predicting death from prostate cancer using data from the Prostate Cancer data Base of Sweden. The Memorial Sloan Kettering Cancer Center nomogram, Cancer of the Prostate Risk Assessment score and Cambridge Prognostic Groups discriminated death from prostate cancer better than the D’Amico and D’Amico-derived risk grouping systems. The order of performance remained after stratifying by primary treatment and year of diagnosis. Using these tools could improve clinical decision-making.

In Study III, we evaluated if a virtual microscopy system which we developed for central re-review in ProMort I and Study IV can be used interchangeably with standard light microscopy for the histopathological evaluation of prostate cancer. We found good repeatability (i.e., intra-observer agreement) and reproducibility (i.e., inter-observer agreement) for several key prostate cancer histopathological features (i.e., core length, tumor length, primary and secondary Gleason pattern, the Gleason score and the Gleason Grade Groups (GGs)) both within and between light and virtual microscopy. The repeatability and/or reproducibility for some of the rare, or less commonly reported, features and for the percentage of Gleason pattern 4 was poor. The repeatability and/or reproducibility for these

features should be improved before they are used in prognostic models. For all evaluated features, the agreement was similar within and between light and virtual microscopy indicating that light microscopy and our internally developed virtual microscopy system can be used interchangeably for the histopathological evaluation of prostate cancer.

In Study IV, we evaluated if the International Society of Urological Pathology (ISUP) revisions of the Gleason grading systems have improved prostate cancer prognostication. We used a nested case-control study (ProMort II) to compare the prognostic performance of the pre-2005 Gleason score and the ISUP 2014 Gleason score. In our study, the ISUP 2014 Gleason score discriminated death from prostate cancer better than the pre-2005 Gleason score. Our results also indicate that this improvement may be due to classifying all cribriform patterns, rather than poorly formed glands, as Gleason pattern 4. We then evaluated if other histopathological features can further improve the prediction of death from prostate cancer. The number of cores with ≥50% cancer involvement, comedonecrosis and high-grade prostatic intraepithelial neoplasia (HGPIN) predicted death from prostate cancer independently of the GGs. Only comedonecrosis and HGPIN remained independent predictors when added to the model with all the standard predictors (the GGs, age, PSA and clinical tumor stage at diagnosis). Adding these features had minimal impact on the model discrimination.

List of scientific papers

I. Renata Zelić, Daniela Zugna, Matteo Bottai, Ove Andrén, Jonna Fridfeldt, Jessica Carlsson, Sabina Davidsson, Valentina Fiano, Michelangelo Fiorentino, Francesca Giunchi, Chiara Grasso, Luca Lianas, Cecilia Mascia, Luca Molinaro, Gianluigi Zanetti, Lorenzo Richiardi, Andreas Pettersson, and Olof Akre
Estimation of relative and absolute risk in a competing-risk setting using a nested case-control study design: Example from the ProMort study
American Journal of Epidemiology 2019; 188(6):1165-1173

II. Renata Zelić, Hans Garmo, Daniela Zugna, Pär Stattin, Lorenzo Richiardi, Olof Akre, and Andreas Pettersson
Predicting prostate cancer death with different pretreatment risk-stratification tools: a head-to-head comparison in a nationwide cohort study
European Urology 2020;77(2):180-188

III. Renata Zelić, Francesca Giunchi, Luca Lianas, Cecilia Mascia, Gianluigi Zanetti, Ove Andrén, Jonna Fridfeldt, Jessica Carlsson, Sabina Davidsson, Luca Molinaro, Per Henrik Vincent, Lorenzo Richiardi, Olof Akre, Michelangelo Fiorentino, and Andreas Pettersson
Interchangeability of light and virtual microscopy for histopathological evaluation of prostate cancer
Submitted

IV. Renata Zelić, Francesca Giunchi, Jonna Fridfeldt, Jessica Carlsson, Sabina Davidsson, Luca Lianas, Cecilia Mascia, Daniela Zugna, Luca Molinaro, Per Henrik Vincent, Gianluigi Zanetti, Ove Andrén, Lorenzo Richiardi, Olof Akre, Michelangelo Fiorentino, and Andreas Pettersson
Prognostic utility of novel histopathological factors in addition to the Gleason Grade Groups in prostate cancer
Manuscript

Table of contents

1 Introduction 1

1.1 Prostate cancer 1

1.1.1 Incidence 1

1.1.2 Mortality 2

1.1.3 Overdiagnosis 2

1.1.4 Overtreatment 3

1.2 Risk stratification 3

1.3 Prognostic markers 5

1.3.1 Prostate-specific antigen 5

1.3.2 Clinical tumor stage 6

1.3.3 The Gleason grading system 7

1.3.3.1 The evolution of the Gleason grading system 7

1.3.3.2 Impact of the Gleason grading system revisions on prostate cancer prognostication 8

1.3.3.3 Inter-observer reproducibility of the Gleason score 9

1.3.3.4 Validation of Gleason Grade Groups 9

1.3.4 Other histopathological markers 12

1.3.4.1 Measures of tumor extent 12

1.3.4.2 Cribriform pattern and intraductal cancer 13

1.3.4.3 The percentage of Gleason pattern 4 13

1.3.4.4 Perineural invasion 14

1.4 Digital pathology and virtual microscopy 14

2 Aims of the thesis 16

3 Materials and methods 17

3.1 Data sources 17

3.1.1 The National Prostate Cancer Register of Sweden 17

3.1.2 Prostate Cancer data Base Sweden 17

3.2 Study designs, study populations and covariate information 18

3.2.1 Study I 18

3.2.2 Study II 19

3.2.3 Study III and IV 20

3.2.3.1 Slide digitalization and managing 21

3.2.3.2 Histopathological review 23

3.2.3.3 Pre-2005 Gleason grading 25

3.3 Methodological considerations and statistical methods 26

3.3.1 Study I 26

3.3.1.1 Competing risks 26

3.3.1.2 Estimation of absolute risks (CIFs) 27

3.3.1.3 Statistical analyses 28

3.3.2 Study II 28
3.3.2.1 Statistical analyses 29
3.3.3 Study III 29
3.3.4 Study IV 30
3.4 Ethical considerations 30

4 Results 32
4.1 Study I 33
4.2 Study II 34
4.3 Study III 36
4.4 Study IV 38

5 Discussion 41
5.1 Study design for prognostic modelling in localized prostate cancer 41
5.1.1 The nested case-control study design 41
5.1.1.1 Predicting death from prostate cancer 42
5.1.1.2 Competing events 42
5.1.1.3 Non-rare (common) events 43
5.2 Current prognostic models in prostate cancer 43
5.2.1 The best-performing prognostic model 44
5.2.2 The best-performing vs. the “gold-standard” prognostic models 45
5.3 Improvement of the prognostic model performance 45
5.3.1 Reliability and measurement heterogeneity 46
5.3.1.1 The ISUP 2014 Gleason Grading system 46
5.3.1.2 Other histopathological features 47
5.3.2 Model extension with a novel prognostic marker 47
5.3.2.1 Case level vs. core level 48

6 Conclusions 50

7 Future directions 51
7.1 Dynamic prediction models 51
7.1.1 Multiparametric magnetic resonance imaging guided biopsy 52
7.2 The role of treatment in clinical prediction models 52
7.3 Clinical utility 54
7.4 Life expectancy 55

8 Acknowledgements 56

9 References 58

List of abbreviations

AUA	American Urological Association
AUC	Area under the receiver operating curve
BCR	Biochemical recurrence
CAPRA	Cancer of the Prostate Risk Assessment score
CI	Confidence interval
CIF	Cumulative incidence function
C-index	Concordance index
CPG	Cambridge Prognostic Groups
CRS4	The Centre for Advanced Studies, Research and Development in Sardinia
cT	Clinical tumor stage
EAU	European Association of Urology
ERSPC	The European Randomised Study of Screening for Prostate Cancer
GGs	Gleason Grade Groups
GUROC	Genito-Urinary Radiation Oncologists of Canada
HGPIN	High-grade prostatic intraepithelial neoplasia
HR	Hazard ratio
ISUP	International Society of Urological Pathology
mpMRI	Multiparametric magnetic resonance imaging
MSKCC	Memorial Sloan Kettering Cancer Center
NCCN	National Comprehensive Cancer Network
NICE	The National Institute for Health and Care Excellence
NPCR	The National Prostate Cancer register of Sweden
OR	Odds ratios
PAH	Postatrophic hyperplasia
PCBaSe	Prostate Cancer data Base Sweden
PLCO	The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial
PSA	Prostate specific antigen
TNM	Tumor-Node-Metastasis staging system
USPSTF	US Preventive Services Task Force
WHO	World Health Organization
κ	Cohen’s kappa
κ_w	Weighted Cohen’s kappa with linear weights

1 Introduction

1.1 Prostate cancer

1.1.1 Incidence

In 2018, prostate cancer was the second most common cancer in men worldwide with an estimated 1.3 million newly diagnosed cases (1). It was the most frequently diagnosed cancer in 105 countries of the world, most notably in developed regions such as the North America, Northern and Western Europe and Australia (1). In Sweden, where prostate cancer was the most common cancer in 2018, almost 11,000 men were diagnosed with prostate cancer (2, 3).

Prostate cancer incidence has been marked by a slow increase until the early 1990s (and somewhat later in the Nordic countries), followed by a more dramatic increase corresponding to the introduction and adoption of prostate specific antigen (PSA) testing (4, 5) and, finally, a slow decrease in subsequent years (Figure 1.1). Given the low specificity and the high false-positive rate of the PSA test, routine PSA screening led to unnecessary prostate biopsies, overdiagnosis of indolent cancers and, ultimately, to overtreatment (6). For this reason, the US Preventive Services Task Force (USPSTF) has made several changes to the screening recommendations over time (7-9) and the trends in PSA testing, and, possibly, the trends in prostate cancer incidence seem to follow the timing of the changes in the USPSTF recommendations (10-12). After completely discouraging the use of PSA screening tests in 2012, in 2018 the USPSTF recommended PSA testing for men aged 55 to 69 based on individual assessment (9).

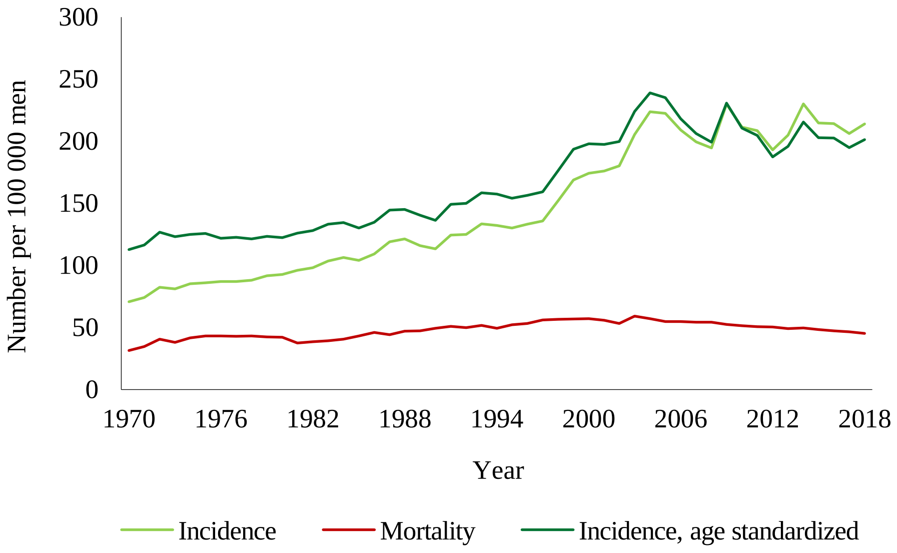


Figure 1.1. Prostate cancer incidence and mortality in Sweden, 1970-2018. Source: The Swedish Cancer Registry and the Cause of Death registry, the National Board of Health and Welfare, Sweden

1.1.2 Mortality

With an estimated 359,000 men dying from prostate cancer in 2018, prostate cancer was the fifth most common cause of cancer-related death in the world (1). In Sweden, 2,500 men died from prostate cancer making it the leading cause of cancer-related death in men in 2018 (3, 13).

Unlike the incidence, prostate cancer mortality has been mostly stable or decreasing over time (Figure 1.1), likely due to the improved treatment and increased detection of early stage disease as a result of PSA screening (5, 14, 15). While PSA screening has undoubtedly led to an increased detection of prostate cancer, especially of localized prostate cancer, the effect on prostate cancer-specific mortality is still a subject of debate (16). Conflicting evidence from two major trials, the prostate arm of the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) (17) and the European Randomised Study of Screening for Prostate Cancer (ERSPC) (18), has been a major driver of this debate (5). The PLCO trial reported no survival benefit due to PSA screening after 13 and 17 years of follow-up (17, 19). On the other hand, the ERSPC reported an overall 21% reduction in cancer-specific mortality in the PSA screened arm at 13 years of follow-up (18) and the results were further confirmed by an updated analysis at 19 years of follow-up (20). A direct comparison of the results is, however, difficult due to the differences in implementation and settings of the two trials. Furthermore, 76% of the men in the control arm of the PLCO trial had at least one PSA test during the study period (21), while in the ERSPC trial there was almost no PSA contamination (20). When these differences were taken into account, analyses indicated compatible results with a 25-31% and 27-32% reduction in prostate cancer mortality due to PSA screening in the ERSPC and PLCO intervention groups, respectively (22, 23). While the ERSCP trial showed a reduction in cancer-specific mortality overall, published results from the Goteborg (24, 25), Spanish (26), Finish (27, 28) and Rotterdam (29, 30) sections of the trial are not uniform. Three sections report no reduction in prostate cancer mortality due to PSA screening (26-28) while the remaining sections report a reduction in prostate cancer mortality (24, 25, 29, 30). In addition, results from a large randomized clinical trial conducted in the United Kingdom (31) also report no reduction in prostate cancer mortality due to the PSA screening.

1.1.3 Overdiagnosis

Screening for prostate cancer aims at identifying high-risk, localized prostate cancer that can be successfully treated. Successful treatment would, in turn, prevent the morbidity and mortality associated with advanced or metastatic prostate cancer. However, PSA screen-detected cancers are mostly asymptomatic cancers that would not cause symptoms and, otherwise, be detected nor contribute to death. This is known as overdiagnosis.

As a part of the informed update of the USPSTF 2012 recommendations (5) the extent of overdiagnosis was evaluated in all the major PSA screening trials (17-20, 24-31). Overdiagnosis was estimated to range from 16.4% (17, 19) to 47.9% (29, 30) of all prostate cancers, and from 20.7% (17, 19) to 58.9% (29, 30) of all screening detected prostate cancers.

Overdiagnosis estimates are highly influenced by differences in study populations and screening practices as well as by the methods used for quantifying overdiagnosis. This is summarized in a review by Leob et al. where the estimates of overdiagnosis ranged from 1.7% to 67% across the range of different methods and underlying populations with differing screening protocols (32).

1.1.4 Overtreatment

Treatment options for prostate cancer include upfront radical treatment with curative intent (i.e., radical prostatectomy or radiation therapy), deferred treatment with curative intent (i.e., active surveillance), deferred treatment without curative intent (i.e., watchful waiting) and upfront androgen deprivation therapy without curative intent. Treatment decision-making is primarily driven by prostate cancer prognosis and life expectancy (33-37). However, in the absence of clear guidelines and strong scientific evidence, treatment of localized cancer is heavily influenced by patients' and clinicians' preferences and beliefs (38).

While men with high-risk prostate cancer are typically offered upfront radical treatment (38, 39), treatment decision is more complex for men with low- or intermediate-risk disease. As, currently, no marker can separate indolent cancers from fast-developing cancers requiring treatment, overdiagnosed men are often overtreated, which, in turn, may lead to unnecessary treatment related side effects such as persistent urinary, sexual and bowel morbidities (5, 40-43). Even with the increased utilization of active surveillance, over 50% of men with low- or intermediate-risk disease are still treated radically (38, 39, 44, 45). The potential harms of diagnosis and treatment should be balanced by improved life expectancy in men with low- and intermediate-risk disease. While upfront radical treatment may reduce the risk of metastatic disease, the long-term effect on prostate cancer mortality is not clear (46-48).

1.2 Risk stratification

Treatment decision-making in prostate cancer is mostly driven by prostate cancer prognosis and life expectancy (33-37, 49). Men with prostate cancer are typically classified into risk groups based on their clinicopathological features, such as PSA level, clinical tumor stage (cT) and Gleason score. In 1998, D'Amico combined these features and grouped men with localized prostate cancer into low-, intermediate- and high-risk groups (49). The D'Amico's risk stratification system quickly became the main standard in clinical practice and the basis for risk stratification in all major prostate cancer guidelines (i.e., the European Association of Urology (EAU) (37, 50), the National Institute for Health and Care Excellence (NICE) (33), the Genito-Urinary Radiation Oncologists of Canada (GUROC) (34), the American Urological Association (AUA) (35), and the National Comprehensive Cancer Network (NCCN) (36)). Incorporation of more granular clinicopathological information (e.g., separating Gleason score 3+4 from 4+3) or introduction of additional clinicopathological parameters (e.g., measures of tumor extent in the diagnostic biopsies) led to further sub-classification of these risk groups into very low- and low-risk group, favorable and unfavorable intermediate-risk group and high- and very high-risk group (36, 51-55). The D’Amico and D’Amico-based risk grouping systems used in all major prostate cancer guidelines are presented in Table 1.1.

Table 1.1. Prostate cancer risk stratification criteria for the most commonly used risk grouping systems

System	Low risk		Intermediate risk		High risk	
	Very low risk	Low risk	Favorable	Unfavorable	High risk	Very high
D'Amico (49)	PSA≤10 and GS≤6 and cT1c-2a		PSA>10-20 or GS=7 or cT2b		PSA>20 or GS=8-10 or cT2c	
EAU (37)	PSA<10 and GS≤6 and cT1c-2a		PSA=10-20 or GS=7 or cT2b		PSA>20 or GS>7 or cT2c	
NICE (33)	PSA<10 and GS≤6 and cT1-2a		PSA=10-20 or GS=7 or cT2b		PSA>20 or GS=8-10 or ≥cT2c	
GUROC (34)	PSA≤10 and GS≤6 and cT1-2a		PSA≤20 and GS≤7 and cT1-2 not otherwise low-risk		PSA>20 or GS=8-10 or ≥cT3a	
AUA (35)	PSA<10 and GG1 and cT1-2a and <34% positive cores and 0 cores with >50% cancer and PSAD<0.15	PSA<10 and GG1 and cT1-2a	PSA=10-<20 or GG2-3 or cT2b-2c		PSA≥20 or GG4-5 or ≥cT3	
AUA_i (35)	PSA<10 and GG1 and cT1-2a and <34% positive cores and no cores with >50% cancer and PSAD<0.15	PSA<10 and GG1 and cT1-2a	GG1 and PSA=10-<20 or GG2 and PSA<10	GG2 and (PSA=10-<20 or cT2b-2c) or GG3 and PSA<20	PSA≥20 or GG4-5 or ≥cT3	
NCCN (36)	PSA<10 and GS≤6 and cT1c and <3 positive cores and ≤50% cancer in each core and PSAD<0.15	PSA<10 and GS≤6 and cT1-2a	PSA=10-20 or GS=3+4 or cT2b-2c and <50% positive cores	PSA=10-20 or GS=3+4/4+3 or cT2b-2c	PSA>20 or GS=4+4/4+5 or cT3a	G1=5 or >4 cores with GS=8-10 or cT3b-4

Abbreviations: EAU, European Association of Urology; NICE, The National Institute for Health and Care Excellence; GUROC, Genito-Urinary Radiation Oncologists of Canada; AUA, American Urological Association; NCCN, National Comprehensive Cancer Network; PSA, Prostate-specific antigen; GS, Gleason score; cT, Clinical tumor stage; GG1-5, Gleason grade groups 1-5; PSAD, Prostate-specific antigen density; G1, primary Gleason pattern

Risk grouping systems are simple to apply in the clinical setting. However, high heterogeneity of patients within risk groups inevitably leads to imprecise outcome prediction. Multivariable model-based risk classification systems circumvent the problem of collapsing patients into broad risk groups by predicting individual risks. Several such models exist in prostate cancer, presented as look-up tables (56-59), risk scores (60) or nomograms (61-64). Even more complicated risk classification models have been developed using artificial neural networks (65-67). In prostate cancer, nomograms have been shown to outperform clinicians and simpler risk stratification tools (68, 69) but also the more complicated tools such as neural networks (70).

Pretreatment risk stratification tools are developed with the aim of assisting clinicians in treatment decision-making for newly diagnosed men. Although they can be developed to predict several relevant clinical outcomes, such as progression-free survival or metastasis-free survival, prostate cancer death and overall survival are most commonly used for treatment decision making. However, most tools have been developed in studies with short follow-up, where biochemical recurrence (BCR), rather than prostate cancer death, was used as the endpoint (49, 52, 56-61, 63, 64, 67). Although BCR is an imperfect surrogate for prostate cancer mortality, only a few of these tools have been validated for prostate cancer death (53, 71-73). Furthermore, most tools have been developed using cohorts of men treated with radical prostatectomy or radiation therapy (49, 52, 56-61, 64, 67), and in selected rather than population-based cohorts (49, 52, 56-59, 61, 64).

Thus, despite the overwhelming number of pretreatment risk stratification tools in prostate cancer (74-77), a considerable proportion of men still remain misclassified, and no single tool is currently recommended for clinical use. Systematic, head-to-head comparison of the most commonly used risk stratification tools with respect to their ability to predict prostate cancer death would clarify which tool performs best and should be used to improve treatment decision-making. Such a tool could also serve as a baseline model or a “gold standard” used to demonstrate independent prognostic value of novel markers.

1.3 Prognostic markers

The standard pretreatment markers of prostate cancer prognosis are PSA, cT and Gleason score at diagnosis. Although the combination of these markers is the basis for most of the risk stratification tools used in clinical practice today, they are not sufficient to capture the heterogeneity in the outcomes of localized prostate cancer. Consequently, a plethora of novel biomarkers and molecular signatures have been evaluated as predictors of prostate cancer prognosis (78-81). Most of these markers are, however, outside of the scope of this thesis and will not be addressed in following chapters.

This thesis focuses primarily on histopathological markers of prostate cancer prognosis. For this reason, PSA, cT, the Gleason score and other commonly evaluated histopathological markers will be described in more detail.

1.3.1 Prostate-specific antigen

PSA is a glycoprotein secreted by prostate epithelial cells that is present in the serum. PSA is organ-specific, not cancer-specific, and the PSA level can be increased as a consequence of non-cancerous conditions, such as benign prostatic hyperplasia or prostatitis. Baseline PSA is a part of risk stratification tools and is commonly used as a marker for monitoring disease progression after curative treatment (35, 37, 49, 53, 60, 80, 82, 83). The role of PSA as a reliable prognostic marker is not without controversies. PSA levels are subject to large analytical and biological variation (84) and the differentiation between indolent and aggressive cancer is sub-optimal as many men with low PSA levels seem to have aggressive disease (85).

1.3.2 Clinical tumor stage

The Tumor-Node-Metastasis (TNM) system was jointly developed by the American Joint Commission on Cancer and the Union for International Cancer Control and is used globally as the benchmark for cancer staging (86). The TNM system is a measure of the extent of the primary tumor (T stage), spread to lymph nodes (N stage) and distant metastases (M stage) (Table 1.2). The cT is based on the digital rectal examination of the prostate (86) and is thus quite a subjective measure of tumor extent (Table 1.2). Although magnetic resonance imaging is expected to improve the accuracy of cT staging, it appears to have high specificity, but poor and heterogeneous sensitivity (87), and is, for now, not recommended as a replacement for the digital rectal examination (86).

Table 1.2. The Tumor-Node-Metastasis (TNM) staging system according to the 8th edition of the American Joint Committee on Cancer staging of prostate cancer.

Stage	Description
Clinical tumor stage	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Clinically unapparent tumor neither palpable nor visible by imaging
T1a	Tumor incidental histologic finding in 5% or less of tissue resected
T1b	Tumor incidental histologic finding in more than 5% of tissue resected
T1c	Tumor identified by needle biopsy (for example, because of elevated PSA)
T2	Tumor confined within prostate[1]
T2a	Tumor involves one-half of one lobe or less
T2b	Tumor involves more than one-half of one lobe but not both lobes
T2c	Tumor involves both lobes
T3	Tumor extends through the prostate capsule[2]
T3a	Extracapsular extension (unilateral or bilateral)
T3b	Tumor invades seminal vesicle(s)
T4	Tumor is fixed or invades adjacent structures other than seminal vesicles, such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall
Regional lymph nodes	
NX	Regional lymph nodes were not assessed
N0	No regional lymph node metastasis
N1	Metastasis in regional lymph node(s)
Distant metastasis	
M0	No distant metastasis
M1	Distant metastasis

[1] Tumor found in one or both lobes by needle biopsy, but not palpable or reliably visible by imaging, is classified as T1c

[2] Invasion into the prostatic apex or into (but not beyond) the prostatic capsule is classified as T2, not T3

1.3.3 The Gleason grading system

In 1966, Dr. Gleason created a 5-point grading system based on the histological patterns of prostate cancer (Figure 1.2A) (88). The Gleason score, ranging from 2-10, was defined as a sum of the first and the second most common pattern, and it was demonstrated that the probability of prostate cancer-specific mortality progressively increased with the increasing Gleason score (88, 89). Since then, the Gleason score has been considered one of the most powerful prognostic factors in prostate cancer.

At the time the Gleason grading system was introduced, the Mostofi system, also known as the World Health Organization (WHO) grading system, was frequently used for grading prostate cancer (90). The WHO grading system was based on cellular anaplasia and the degree of glandular differentiation and classified prostate cancers into well differentiated, moderately differentiated and poorly differentiated. This system has, however, been entirely abandoned in favor of the Gleason grading system.

1.3.3.1 The evolution of the Gleason grading system

The Gleason score has undergone a series of changes over time, most notably two major revisions by the International Society of Urological Pathology (ISUP) in 2005 (91) and in 2014 (92).

The ISUP 2005 revision addressed both the interpretation of the morphological patterns and the reporting methods. The most notable changes were recommendations against assigning Gleason pattern 1 and 2, narrowing down the definition of Gleason pattern 3 by including most of the cribriform glands and poorly formed glands in the definition of Gleason pattern 4, and, finally, defining Gleason score as the sum of the most common and the highest Gleason pattern (Figure 1.2B). As a consequence, pathologists more often assigned Gleason score 7 and tumors with Gleason score 6 had better prognosis. This tendency of assigning a higher Gleason score over time is known as a grade inflation (93, 94).

In 2014, the ISUP further modified the definition of the morphological patterns by including all cribriform glands and glomeruloid glands in the definition of Gleason pattern 4 (Figure 1.2C). Furthermore, ISUP endorsed the five-tiered Gleason Grade Groups (GGs) where GG1 is defined as Gleason score ≤6, GG2 as Gleason score 3+4, GG3 as Gleason score 4+3, GG4 as Gleason score 4+4, 3+5, 5+3 and GG5 as Gleason score 9-10. The endorsement of the GGs was based on the claims that:

1. The GGs stratify prostate cancer better that the current system (where the current system was the three-tiered Gleason score (≤6, 7, 8-10)),
2. The number of grading categories was reduced from 2-10 to 1-5,
3. Having GG1 instead of Gleason score 6 as the lowest grade could reduce overtreatment of indolent cancers, and

4. The GGs are based on the modified Gleason grading system which bears little resemblance to the original Gleason grading system (92, 95, 96), which should justify it being introduced as a new classification system.

The ISUP 2014 revision was subsequently adopted by the 2016 WHO classification of tumors of the urinary system and male genital organs (97).

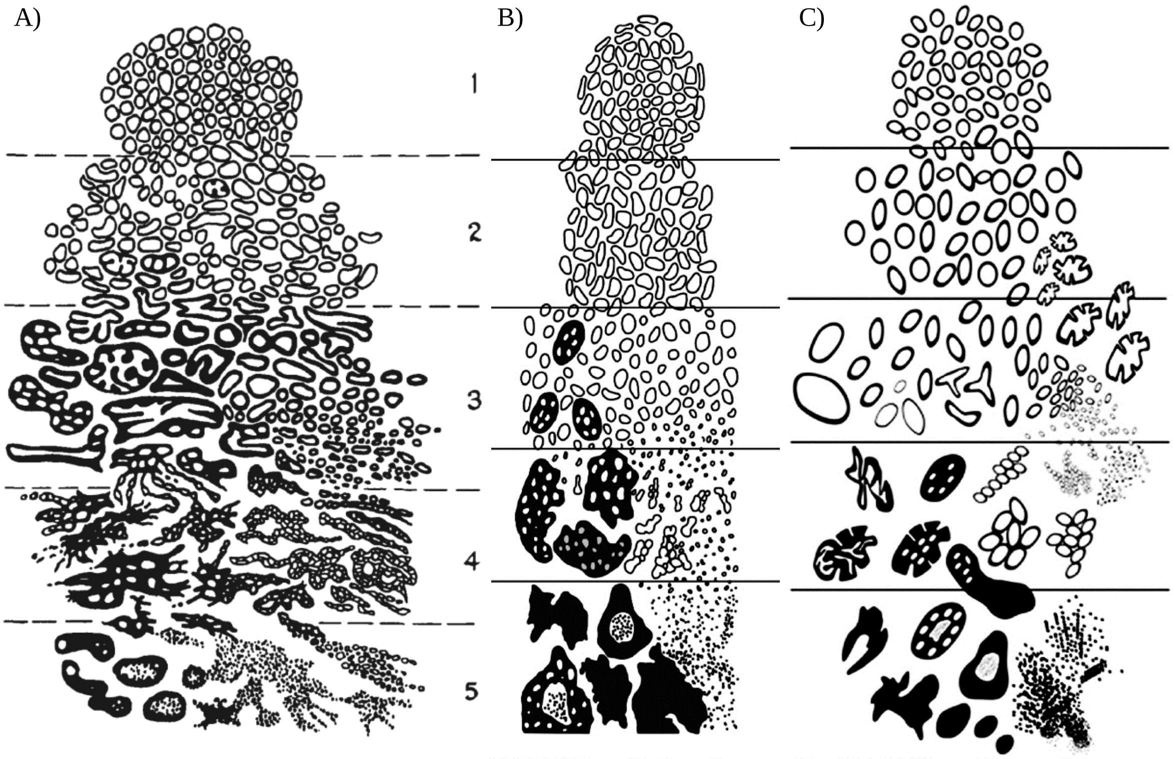


Figure 1.2. The evolution of the Gleason grading system. Original (pre-2005) Gleason (A), International Society of Urological Pathology (ISUP) 2005 Gleason (B) and ISUP 2014 Gleason grading system (C).

A) Reprinted from Gleason DF. Histologic grading of prostate cancer: a perspective. Hum Pathol. 1992;23:273–279, with permission from Elsevier; B) and C) Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Contemporary Approach to Gleason Grading of Prostate Cancer. In: Prostate Biopsy Interpretation by Shah R.B., Zhou M. © 2019.

1.3.3.2 Impact of the Gleason grading system revisions on prostate cancer prognostication

The Gleason grading system revisions aimed at improving inter-pathologist agreement and, ultimately, prostate cancer prognostication (91, 92, 95). To evaluate improvements in prognostication, a direct comparison of the Gleason scores assigned according to the different Gleason grading systems is necessary (98). However, only a few studies have compared the ISUP 2005 Gleason score to the pre-2005 Gleason score in predicting BCR and found either a small (99, 100) or no improvement (101). None of the studies validating the GGs as a predictor of adverse outcomes in prostate cancer (see more information under heading 1.3.3.4.) have also performed a re-review of the same samples according to both the ISUP

2014 Gleason grading criteria and the ISUP 2005 or pre-2005 Gleason grading criteria. Only one study has compared the GGs to previous Gleason grading revisions by modelling the GGs and the diagnostic Gleason score in predicting BCR, but they did not compare the prognostic performance of the two models (102). Thus, even though a plethora of studies have unsurprisingly confirmed that the GGs are a prognostic factor in prostate cancer, it is still not clear if the changes in the Gleason grading system introduced in 2005 and 2014 have improved prostate cancer prognostication.

1.3.3.3 Inter-observer reproducibility of the Gleason score

Prior to the ISUP 2005 revision, the agreement for the Gleason score ranged from 0.16 to 0.70 among uropathologist (103-106) and 0.00 to 0.88 among general urologists (106, 107) (Table 1.3). The general opinion was that the ISUP 2005 revision led to an improvement in the inter-observer reproducibility (108). This improvement was ascribed to either the more specific definition of patterns 3 and 4, or to Gleason score 2-5 no longer being used (94, 99). However, the agreement for the ISUP 2005 Gleason score ranged from 0.48-0.68 among uropathologists (109-113) and from -0.13-0.68 among general pathologists (110, 113-115), indicating no obvious improvement in agreement (Table 1.3). Similarly, the agreement for the ISUP 2014 Gleason score ranged from 0.43-0.75 among general pathologists (116, 117), while for the GGs it ranged from 0.39-0.75 among general pathologists (116, 117) and 0.48-0.89 among uropathologists (118) (Table 1.3).

The lack of obvious improvement in the inter-observer agreement with the ISUP 2005 and 2014 revisions could be due to the slow adoption of the new grading criteria, different interpretations of the guidelines or to differences in study design and methods used to quantify the agreement. However, several studies have demonstrated that the agreement can be improved by additional training (110), use of reference images (118-120) or various techniques for improving reproducibility, such as web-based education or the use of interactive digital slides with heat maps (110, 121-123).

1.3.3.4 Validation of Gleason Grade Groups

Since the ISUP endorsed the GGs, a plethora of studies have evaluated the ability of the GGs to predict BCR (96, 124-132) and/or death from prostate cancer (133-140) (Table 1.4). However, most of these validation studies were based on selected samples of treated men with a short follow-up. Furthermore, in most of the studies there was no central re-review of the diagnostic biopsies according to the ISUP 2014/WHO 2016 criteria. In fact, of four studies that had access to centrally re-reviewed diagnostic biopsies (102, 134, 136, 140), only one study has done so according to the ISUP 2014/WHO 2016 criteria (102) (Table 1.4).

Given that none of the validation studies re-reviewed the same samples according to different Gleason grading criteria, the claims of better prognostic accuracy of the GGs have been based on comparisons with different groupings of the Gleason score, most commonly the three-tiered Gleason score (≤6, 7, ≥8), and on minimal changes in model discrimination (e.g., change in the second/third decimal of the area under the receiver operating curve (AUC)/concordance index (C-index)) (Table 1.4) (98). However, it has also been shown that using the GGs seems to result in less upgrading on prostatectomy specimens (102, 141), which indicates better identification of potentially aggressive tumors, and, in turn, has great clinical implications.

Table 1.3. Inter-observer reproducibility of the Gleason score

Author	Year	Sample size	Gleason score	Pathologists		Inter-observer agreement	Kappa type
				No.	Type		
McLean (103)	1997	71	Pre-2005	3	Urologic	Range: 0.16-0.29	Weighted
Allsbrook (104)	2001	46	Pre-2005	10	Urologic	Range: 0.56-0.70	Weighted
Allsbrook (107)	2001	38	Pre-2005	41	General	0.44 (Range: 0.00-0.88)	Simple
Glaessgen (105)	2004	279[1]	Pre-2005	4	Urologic	Range: 0.48-0.55	Weighted
Oyama (106)	2005	37	Pre-2005	8	General	0.49	Simple
			Pre-2005	6	Urologic	0.68	
Melia (111)	2006	81	ISUP 2005	9	Urologic	0.54 (range: 0.49-0.61)	Fleiss
Griffiths (110)	2006	20	ISUP 2005[7]	24	General	0.33	Simple
			ISUP 2005[8]		General	0.41	
			ISUP 2005		Urologic	0.62	
Veloso (142)	2007	110[2]	ISUP 2005[9]	3	Mixed[12]	Range: 0.32-0.44	Weighted
			ISUP 2005[10]			Range: 0.31-0.44	
			ISUP 2005[11]			Range: 0.39-0.50	
Mulay (114)	2008	40	ISUP 2005[7]	4	General	0.46 (Range: 0.36-0.65)	NR
			ISUP 2005[8]			0.54 (Range: 0.46-0.68)	
Singh (115)	2011	20[3]	ISUP 2005	21	General	Range: -0.13 to 0.55	Simple
Rodriguez-Urrego (109)	2011	50	ISUP 2005	4	Urologic	0.54	Simple
Harnden (113)	2011	20	ISUP 2005	5	Urologic	0.57	Simple
				19	General	0.61	
				27	Mixed	0.60	
Goodman (112)	2012	1,905[4,5]	ISUP 2005	2	Urologic	0.56 (95% CI: 0.48-0.63)	Weighted
Abdollahi (143)	2012	101	Pre-2005	5	NR	0.29	NR
Abdollahi (122)	2013	150	ISUP 2005[7]	3	NR	0.25 (Range: 0.14-0.39)	NR
			ISUP 2005[8]			0.52 (Range: 0.39-0.65)	
Ozkan (116)	2016	197[6]	ISUP 2014	2	General	0.43 (95%CI: 0.42-0.48)	Simple
			GGs	2		0.39 (95%CI: 0.34-0.47)	
Qureshi (144)	2016	47	NR	7	General	0.5	Simple
Al Nemer (117)	2017	126	ISUP 2014	4	General	0.75 (95%CI: 0.71-0.79)	Fleiss
			GGs			0.75 (95%CI: 0.71-0.79)	
Egevad (118)	2018	90[4]	GGs	23	Expert	Range: 0.48 - 0.89	Weighted

Abbreviations: ISUP, International Society of Urological Pathology; GGs, Gleason Grade Groups; NR, not reported

[1] 69 patients with 279 slides with cancer

[2] Number of reviewed cores per pathologist not equal

[3] 10 biopsy samples, 8 transurethral resection of the prostate (TURP) samples, 2 radical prostatectomy samples

[4] Reviewed using digital microscopy

[5] 268 patients with 1,905 slides

[6] 407 cores belonging to 34 patients. Tumor was detected in 197 slides (cores) by both pathologists

[7] Before intervention

[8] After intervention

[9] Gleason score calculated as a sum of primary and secondary Gleason pattern

[10] Gleason score calculated as a sum of primary and tertiary (when present) Gleason pattern

[11] The highest core level Gleason score

[12] Two pathologists were experienced in urological pathology and one was less experienced

Table 1.4. Gleason Grade Groups validation studies

Author	Population			Central review	Outcome[1]	Comparison	Model performance
	Period	Sample size	Treatment				
Leapman (138)	1995-2014	10,529	mixed	no	PCSM	GGs vs. GS (extended)[2]	-
Beckmann (133)	2006-2013	4,268	mixed	no	PCSM	GGs	-
Berney (134)	1990-2003	988	WW/early hormones	yes[3]	PCSM	Overall GGs vs. Worst GGs	0.756 vs. 0.752
Bondarenko (124)	2006-2016	621	RP - robot assisted	no	BCR	GGs vs. GS (6,7,8-10) vs. GS (6,3+4,4+3,8-10) vs. GS (6,7,8,9-10)	0.724 vs. 0.740 vs. 0.730 vs. 0.745
Chen (135)	2010	13,798	WW/early hormones	no	PCSM	GGs vs. GS (6,7,8-10)	0.908 vs. 0.907
Dell'Oglio (125)	2005-2014	1,624	RP	no	BCR	GGs vs. GS (6,7,8-10) vs. GS (6,3+4,4+3,8-10) vs. GS (6,7,8,9-10)	0.660 vs. 0.653 vs. 0.656 vs. 0.657
Epstein (126)	2005-2014	16,172	RP	no	BCR	GGs vs. GS (6,7,8-10) vs. GS (6,3+4,4+3,8-10) vs. GS (6,7,8,9-10)	0.813 vs. 0.805 vs. 0.811 vs. 0.806
He (137)	2006-2012	331,320	mixed	no	PSCM	GGs	-
Kirmiz (127)	2012-2017	8,052	RP	no	BCR	GGs vs. GS (6,7,8-10)	0.76 vs. 0.75
Loeb (128)	2005-2007	5,880 RP: 4,325 RT: 1,555	RP/RT	no	BCR	GGs vs. GS (6,7,8-10) vs. GS (6,7,8,9-10)	RP: 0.659 vs. 0.658 vs. 0.658 RT: 0.727 vs. 0.738 vs. 0.730
Mathieu (129)	2005-2014	27,122	RP	no	BCR	GGs vs. GS (6,7,8-10)	0.743 vs. 0.740
Offerman (102)	2002-2015	339	RP	yes[4]	BCR	GGs vs. diagnostic GS (6,3+4,4+3,8,9-10)	-
Pierorazio (96)	2004-2011	7,850	RP	no	BCR	GGs	-
Pompe (139)	2004-2009	242,531 RP: 91,565 RT: 38,184 EBRT: 52,926 NLT: 59,856	mixed	no	PCSM	GGs vs. GS (6,7,8-10)	RP: 0.813 vs. 0.804 RT: 0.731 vs. 0.727 EBRT: 0.759 vs. 0.750 NLT: 0.817 vs. 0.810
Shulman (130)	2005-2015	2,509	RP	no	BCR	GGs	-
Spratt (131)	1994-2013	3,694	RP	no	BCR	GGs vs. GS (6,7,8-10)	0.67 vs. 0.65
Spratt (140)	1990-2013	847	EBRT	yes	PCSM	GGs vs. GS (6,7,8-10)	0.752 vs. 0.733
Yeong (132)	2005-2014	638	RP	partial[5]	BCR	GGs vs. GS (6,7,8-10)	0.687 vs. 0.647
Delahunt (136)	2003-2007	496	RT[6]	yes[7]	PCSM	GGs vs. GS (6,7,8,9,10)	0.782 vs. 0.750

Abbreviations: PCSM, Prostate cancer-specific death; GGs, Gleason Grade Groups; GS, Gleason score; RP, Radical prostatectomy; WW, Watchful waiting; BCR, Biochemical recurrence; RT, Radiation therapy; EBRT, External beam radiation therapy; NL, No local therapy

[1] When available, results for prostate cancer specific mortality are reported. Otherwise, results for biochemical recurrence are reported.

[2] ≤3 + 3, 3 + 4, 4 + 3, 4 + 4, 4 + 5, 5 + 4, 5 + 5.

[3] Central review according to the Gleason scoring system (Epstein, 2010)

[4] Central review according to the ISUP 2014 criteria

[5] Central review according to the ISUP 2005 criteria of 44 men diagnosed in 2005

[6] Men treated with androgen suppression 6 months prior to radiation therapy and men treated with 12 months of androgen suppression after radiotherapy

[7] Central review according to the ISUP 2005 criteria, Gleason score then recoded to GGs

1.3.4 Other histopathological markers

Most men diagnosed with prostate cancer have their diagnosis made on a needle biopsy. Prostate cancer biopsy tissue contains a vast amount of information, some of which is routinely recorded by pathologists in a pathology report. In addition to the Gleason score, pathologists are required to report different measures of tumor extent (i.e., the number of positive cores/total number of cores and length of tissue involved by carcinoma in mm or the linear extent of prostatic tissue involved by carcinoma as a percentage (%)) and the presence/absence of extraprostatic extension (145). Reporting of additional information, such as the percentage of Gleason pattern 4 or 4/5, the presence of perineural invasion, and intraductal carcinoma, is only a recommendation (145). At ISUP 2014 revision, it was additionally recommended to report the GGs for individual cores, the percentage of Gleason pattern 4 for cores with Gleason score 7 and the presence of intraductal carcinoma (95).

Given the need for novel markers that can separate indolent from aggressive cancers, many of these routinely reported histopathological features, as well as some features which are not routinely reported, have been studied as potential predictors of prostate cancer prognosis (146). The most commonly studied histopathological features are described in more detail below.

1.3.4.1 Measures of tumor extent

Currently, there is no consensus on how to best quantify tumor extent in prostate biopsies. Different measures of tumor extent, such as the number or percentage of cores with cancer or measures of linear extent of cancer (i.e., total length and percentage of cancer in mm), have been evaluated as potential prognostic factors. While the results for the number of cores with cancer and total length of cancer are not conclusive (147-151), the results for the percentage of cores with cancer and the total percentage of cancer are more consistent. The percentage of cores with cancer is an established predictor of BCR (52, 148, 150, 152-155) and has also been shown to predict death from prostate cancer (154). Similarly, the total percentage of cancer has been repeatedly identified as a predictor of BCR (147, 149, 150, 155, 156) and of death from prostate cancer (157, 158). These measures are highly correlated (151, 156) and there seems to be no gain in modelling them together (153). Which of the two is a better measure of tumor extent is not clear (150, 155, 156).

When modelled together with other established predictors of prostate cancer prognosis, the contribution of both the percentage of cores with cancer and the total percentage of cancer to discrimination seems to be minimal (148, 157), which brings their clinical utility into question. Nevertheless, some of these measures are already incorporated in risk stratification tools and used for clinical decision making. The percentage of cores with cancer and the number of cores with >50% cancer involvement separate very low- and low-risk cancer in the AUA guidelines (35). The percentage of cores with cancer is also used for pretreatment risk stratification using the Cancer of the Prostate Risk Assessment score (CAPRA) score (60), and together with the number of cores with cancer and the number of cores with ≤50% cancer involvement in the NCCN guidelines (36). The percentage of cores with cancer, the number of cores with cancer, and the number of cores with ≤50% cancer involvement are also a part of the criteria for active surveillance (82, 159-161).

1.3.4.2 Cribriform pattern and intraductal cancer

Gleason pattern 4 is characterized by four distinct growth patterns: poorly-formed, fused, glomeruloid and cribriform pattern (92). Cribriform pattern has been associated with an unfavorable biologic behavior, and has often been studied as a potential predictor of prostate cancer prognosis (162). The presence of cribriform pattern in radical prostatectomy samples has been associated with BCR (162-166) as well as with metastasis after radical prostatectomy (164). Cribriform pattern in biopsy samples of men with Gleason score 3+4 has been shown to predict upstaging (167) and BCR after radical prostatectomy (168).

Intraductal cancer in radical prostatectomy or biopsy samples has been identified as a predictor of BCR (168, 169), clinical progression-free survival (170) and death from prostate cancer (169). Given the microscopic similarity of intraductal cancer and cribriform pattern, several studies have evaluated the predictive value of the presence of cribriform pattern and/or intraductal carcinoma. The presence of cribriform pattern and/or intraductal carcinoma in radical prostatectomy samples predicted BCR independently of the Gleason score (171). Furthermore, the presence of cribriform pattern and/or intraductal carcinoma on biopsy samples predicted death from prostate cancer (172), and incorporating these two patterns into the GGs has been shown to somewhat improve discrimination of death from prostate cancer compared to the standard GGs (C-index: 0.79 vs. 0.76) (173).

1.3.4.3 The percentage of Gleason pattern 4

Both the ISUP 2014 revision and the WHO recommend reporting percentage of Gleason pattern 4 for Gleason score 7 prostate cancer in needle biopsies and RP samples (95, 97), however, the method for quantification is left optional (95). Different methods for quantifying the percentage of Gleason pattern 4 have been studied as potential predictors of prostate cancer prognosis, such as the overall percentage of Gleason pattern 4 (total length (in mm) of Gleason pattern 4/total length (in mm) of cancer), maximum percentage of Gleason pattern 4 in one core, total length (in mm) of Gleason pattern 4 etc. Another potential issue regards the cases for which it is recommended to record the percentage of Gleason pattern 4. The ISUP 2014 authors stated that they do not record it if any other core has GG5 since treatment decision is more straightforward for men with GG5 and the percentage of Gleason pattern 4 has little, if any, clinical relevance (95).

The overall percentage of Gleason grade 4 on both radical prostatectomy and biopsy samples has been shown to be an independent predictor of adverse pathology at RP (174-177), BCR (166, 174, 176-178) and prostate cancer death among men with GG2 and/or GG3 (179). The overall percentage of Gleason grade 4 has been shown to outperform the maximum percentage of Gleason pattern 4 (174, 180). However, when modelled with PSA, cT and the percentage of cores with cancer, the total length of the Gleason pattern 4 outperformed both the maximum and the overall percent of Gleason pattern 4 in predicting adverse pathology at radical prostatectomy (177).

1.3.4.4 Perineural invasion

Perineural invasion is a well-known mechanism for the extraprostatic spread of prostate cancer (181) and as such has long been recognized as a potential prognostic factor. A plethora of studies have evaluated the association between perineural invasion on biopsy or radical prostatectomy specimens and BCR, with conflicting results. At least four systematic reviews (182-185), of which three include a meta-analysis (182, 184, 185), have confirmed the association of perineural invasion with BCR after radical prostatectomy or radiation therapy. Nevertheless, the authors still remained cautious when discussing their summary findings due to substantial heterogeneity across the evaluated studies and because of the presence of selection and publication bias.

Overall, relatively few studies focus on the prognostic significance of perineural invasion on biopsies. A recent study evaluated perineural invasion on biopsies as a predictor of BCR and found a 50% increase in the rate of BCR, although with wide confidence intervals (CIs) (hazard ratio (HR): 1.55; 95% CIs: 0.98-2.45) (186). When the authors pooled their results with the results from the three largest published studies with similar design, methods and research question (187-189), the combined estimate supported perineural invasion on prostate biopsy specimens as a strong independent predictor of BCR after radical prostatectomy (186). In addition, perineural invasion on prostate biopsy seems to also be a predictor of death from prostate cancer (190, 191). These results indicate that perineural invasion should be a required component of histopathologic review and it may be relevant for clinical decision-making in prostate cancer.

1.4 Digital pathology and virtual microscopy

Advancements in whole slide imaging technology and software development have led to the development of digital pathology and virtual microscopy (192). In digital pathology, glass slides are digitalized using a scanner, stored, and viewed locally or transmitted over a network for remote viewing on a computer or other electronic devices using a virtual microscopy software interface that emulates the light microscopy experience (192-194). Digital pathology has mostly been used for education, quality assurance, research, image analysis, collaborations and seeking a specialist second opinion (195-197). However, owing to recent approvals by the US Food and Drug Administration, digital pathology solutions are also starting to be used in clinic practice.

Several studies have assessed the interchangeability of standard light and virtual microscopy in prostate cancer by evaluating the inter-method, intra-observer agreement and/or the intra-method, inter-observer agreement (109, 113, 198, 199) for several histopathological features, including the Gleason score. The inter-method, intra-observer agreement for the Gleason score ranged from 0.49 to 0.77 (109, 198, 199). The inter-observer agreement on light microscopy (range: 0.54-0.61) (109, 113) was overall similar to the inter-observer agreement on virtual microscopy (range: 0.45-0.62) (109, 112, 113, 199) indicating interchangeability of the two methods. One of the well-known downsides of virtual microscopy, which could potentially limit the use in clinical practice, is the longer review time (200). However, review time will probably be shortened with improvements in software design, and by automating several of the most time-consuming parts of slide annotation, such as circling different regions of interest. This opens the door for many exciting possibilities, such as using machine learning methods for automation.

2 Aims of the thesis

The overarching aim of this thesis is to improve prognostication for men with localized prostate cancer through validation of the existing risk stratification tools based on standard clinical and histopathological factors, and through validation of the existing, and identification of novel, prognostic markers.

The study specific research aims were:

Study I. To evaluate if the relative and absolute risks of dying from prostate cancer estimated in the competing risk setting using the nested case-control study (ProMort I) are comparable to the relative and absolute risks estimated in the underlying cohort, and to quantify the bias in the risk estimates.

To explore alternative approaches for estimating relative and absolute risks in the competing-risks setting using the nested case-control study design.

Study II. To compare the prognostic performance of the most commonly used pretreatment risk prediction tools in predicting death from prostate cancer, overall and stratified by primary treatment (active surveillance/watchful waiting, radical prostatectomy/radiation therapy and androgen deprivation therapy) and by year of diagnosis (1998-2002, 2003-2006, 2007-2016).

Study III. To evaluate if the standard light microscopy and a virtual microscopy system which we developed for the central re-review in ProMort I and Study IV can be used interchangeably for the histopathological evaluation of prostate cancer.

To evaluate the repeatability (i.e., intra-method, intra-observer agreement) and the reproducibility (i.e., intra-method, inter-observer agreement) for different key histopathological features in prostate cancer, including the ISUP 2014 Gleason grading system for both light and virtual microscopy.

Study IV. To evaluate if the Gleason grading system revisions have improved prostate cancer prognostication by comparing the prognostic performance of the pre-2005 Gleason score and the ISUP 2014 Gleason score in predicting death from prostate cancer.

To evaluate if additional histopathological features (e.g., specific tumor features) can further improve the ability to predict death from prostate cancer.

3 Materials and methods

3.1 Data sources

All the studies in this thesis are based on data from the National Prostate Cancer Register of Sweden (NPCR) and Prostate Cancer data Base Sweden (PCBaSe), a research database constructed by linking the NPCR to other national registers and demographic databases.

3.1.1 The National Prostate Cancer Register of Sweden

The NPCR is a cancer quality register including virtually all incident cases of prostate cancer in Sweden since 1998 (201). Compared to the Swedish National Cancer Register, to which reporting is mandatory and regulated by law, the NPCR has a 98% coverage (202).

Data is registered in the NPCR using four registration forms: a diagnostic form, a form for subsequent work-up and primary treatment, as well as separate forms for radiation therapy (since 2007) and radical prostatectomy (since 2015). The NPCR contains detailed information on:

1. Diagnostic workup (e.g., date and hospital of diagnosis, cause for diagnostic workup leading to cancer diagnosis (PSA-screening, lower urinary tract symptoms, other symptoms)),
2. Tumor features (e.g., clinical TNM classification, biopsy tumor differentiation (Gleason score or WHO grade), serum PSA level at diagnosis), and
3. Planned primary treatment (i.e., active surveillance, watchful waiting, radical prostatectomy, radiation therapy or primary androgen deprivation therapy) within 6 months of diagnosis.

In 2007, the NPCR started registering more detailed information on the biopsy procedure (i.e., indicators of tumor extent such as the number of cores taken at biopsy, the number of cores with cancer, the total length of all biopsy cores and the total length of cancer in all cores), prostate volume and radical prostatectomy and radiation therapy.

Vital status in the NPCR is updated yearly by linkage to the Swedish Population Register. Date and cause of death, coded according to the 10th revision of the International Classification of Diseases, are obtained through linkage to the Swedish Cause of Death Register. Prostate cancer specific death is defined as death where prostate cancer was coded as the underlying cause of death. For more information on the registers used in this thesis, recorded information, their coverage and validity, please see Table 3.1.

3.1.2 Prostate Cancer data Base Sweden

In 2008, the Swedish personal identity number was used to link NPCR to a number of national population-based health-care registers and demographic databases, and construct a research database named PCBaSe (201, 203). The NPCR was first linked to the Swedish Cancer Register, the Total Population Register and the Cause of Death Register, and, subsequently, to the National Patient Register, the Prescribed Drug Register, the Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA) (Table 3.1) as well as several other national (quality) registers (203). The PCBaSe linkages are updated every three years.

Table 3.1. Overview of the registers and databases providing data for PCBaSe. Only the registers from which information was used in this thesis are presented.

Registry	Recorded information	Coverage	Update
The Swedish Cancer Register (204)	Personal information, medical data (e.g., date and bases for diagnosis, tumor site, histological type, stage) and follow-up data (date and cause of death, migration date)	96%[1] (205)	Annual
The Total Population Register (206, 207)	Personal information, birth-related data (e.g., date and country), address data, income, citizenship, country of immigration/emigration, and dates of death and immigration/emigration	0.25-0.5%[2] (207)	Daily
The Cause of Death Register (13, 208)	Personal information, birth-related data, date of death, underlying and contributing cause(s) of death, information on autopsy and surgery within 4 weeks prior to death	for PCa[3]: 86-96% (209, 210)	Annual
The National Patient Register (204)	Personal patient information, geographical data, administrative data (e.g., inpatient (IP) and outpatient (OP) date of admission and discharge), medical data (e.g., main and secondary diagnosis, procedures)	IP: 100% (211) OP: 80%	Monthly
LISA (212, 213)	Personal and family-related demographic data, civil status, birth-related data, data on immigration/emigration, highest level of education, data on occupation, employment status and income, and socioeconomic indices	NA	Annual

Abbreviations: LISA, Longitudinal Integration Database for Health Insurance and Labour Market Studies; NA, Not available

[1] The estimated capture rate of all cancers to the Swedish Cancer Registry compared with the National Patient Register

[2] The estimated over-coverage

[3] 86-96% refers to agreement with cause of death determined by a medical record review

3.2 Study designs, study populations and covariate information

3.2.1 Study I

In Study I, we used a case-control sample from the NPCR, called ProMort I. ProMort I is an ongoing study which aims to identify tissue-based, molecular biomarkers of death from prostate cancer for men with low- or intermediate-risk prostate cancer. As this thesis focuses on clinical and histopathological markers of prostate cancer, ProMort I was used only to evaluate if the nested case-control design can be used to evaluate absolute risks of dying from prostate cancer in the competing risks setting. Furthermore, as a part of ProMort I, we developed a virtual microscopy system which was used in Study III and Study IV (for more details see the section 3.4.2.1).

ProMort I was nested among all men in the NPCR diagnosed with low- or intermediate-risk prostate cancer between January 1, 1998 and December 31, 2011. We defined low- or intermediate-risk prostate cancer as cT1-2, Gleason score ≤7 (or WHO grade 1 when information on the Gleason score was missing), serum PSA<20 ng/mL, and no signs or non-assessed status of lymph node (N0 or Nx) or distant (M0 or Mx) metastases. Of approximately 130,000 men in the NPCR, 57,952 men fulfilled these criteria. Follow-up was available until December 31, 2012. All men who died from prostate cancer during follow-up (n=1,735) were selected as cases. For each case, we randomly selected one control, matched on year and hospital of diagnosis. The control had to be alive at the date of death of the respective case. Cases without an eligible control within the matching stratum (n=25) were excluded from the study. The final ProMort I data set included 1,710 cases and 1,710 controls.

Information on age, cT, Gleason score/WHO grade and PSA at diagnosis, as well as vital status and cause of death, was abstracted from the NPCR. We assigned Gleason score ≤6 to 140 cases and 103 controls with WHO differentiation grade 1 and no information on the Gleason score.

3.2.2 Study II

Study II is a cohort study including all men in PCBaSe 4.0 (the fourth update of PCBaSe), who were diagnosed with non-metastatic (i.e., not M1 or N1) prostate cancer between January 1, 1998 and December 31, 2016 (n=154,811). Follow-up was available until December 31, 2016. Prostate cancer death was used as the main outcome.

Information on age, PSA, clinical TNM stage, primary and secondary Gleason pattern, Gleason score and tumor extent at diagnosis was abstracted from PCBaSe, and used to define the risk stratification tools assessed in this study:

1. Risk group systems (D’Amico (49), NICE (33), GUROC (34), AUA (35), EAU (37), NCCN (214) and the Cambridge Prognostic Groups (CPG) (53)) (see also Table 1.1),
2. Risk scores (CAPRA score (60)), and
3. Nomograms (pre-operative Memorial Sloan Kettering Cancer Center (MSKCC) nomogram (83)).

We also abstracted information on planned primary treatment, year of diagnosis, Charlson Comorbidity Index, marital status and education level from PCBaSe.

Missing values for the variables included in the risk stratification tools were imputed using multivariate imputation by chained equation (215, 216). Information on cT2-3 sub-stage (i.e., cT2a, cT2b, cT2c, cT3a, cT3b) is not recorded in PCBaSe and could not be imputed. Instead, we used a cohort of men diagnosed with prostate cancer between 1995 and 2015 who were treated with proton-boost radiation therapy at the Uppsala University Hospital, Uppsala, Sweden (217) to develop regression models predicting the probability of cT2 and cT3 sub-stage. These models were then used to predict the probability of cT2 and cT3 sub-stage for each study subject in PCBaSe with known cT2 or cT3 stage, respectively. Each patient was assigned the cT2 or cT3 sub-stage category with the highest predicted probability.

3.2.3 Study III and IV

Study III is a reliability/measurement study conducted in a subsample (N=60) of Study IV with the aim of validating an internally developed virtual microscopy system used for central re-review in ProMort I and Study IV. Study IV is a case-control study nested in the NPCR, named ProMort II.

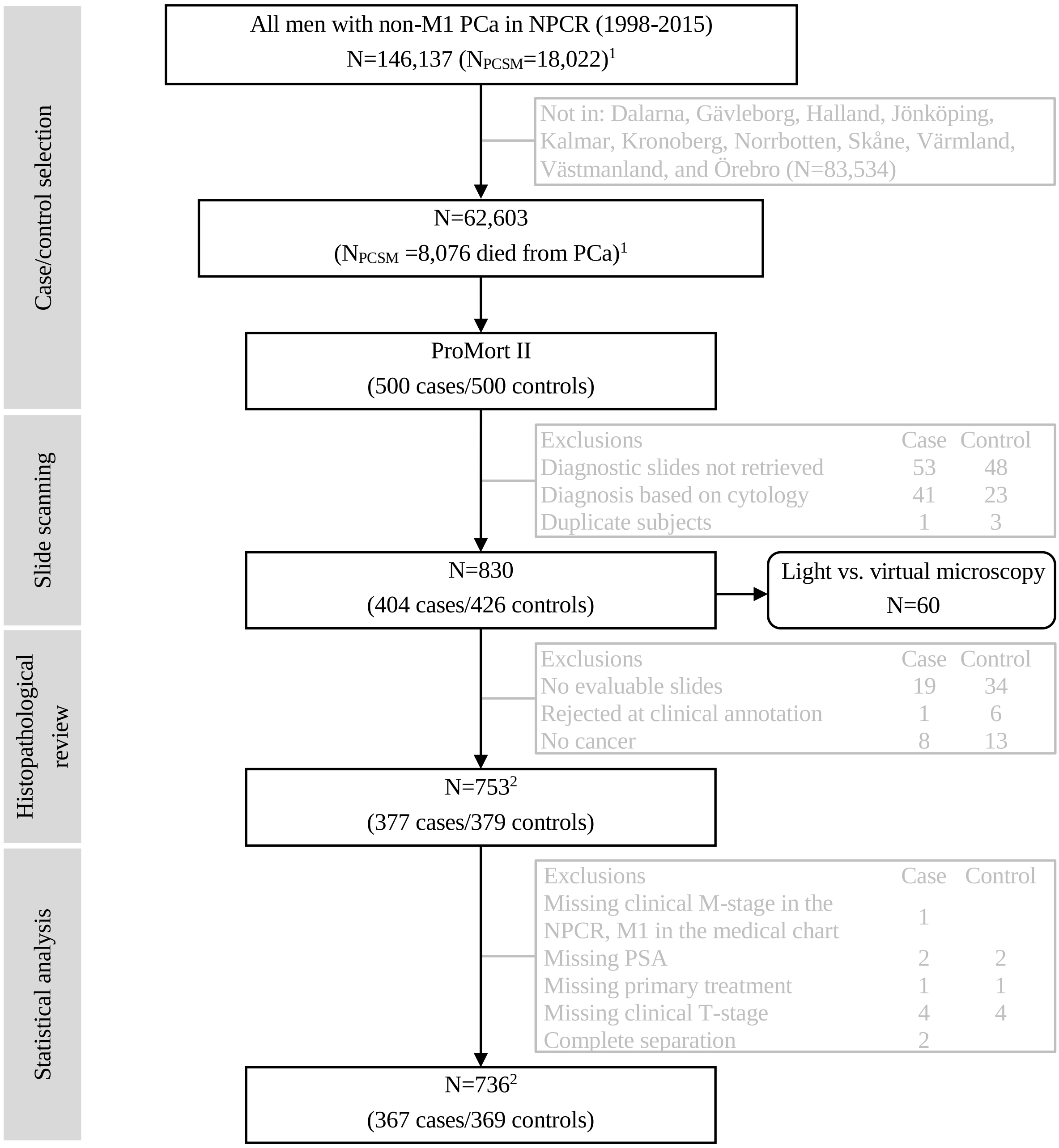

Case/control selection
Slide scanning
Histopathological review
Statistical analysis
All men with non-M1 PCa in NPCR (1998-2015)
N=146,137 (N_PCSM=18,022)1
Not in: Dalarna, Gävleborg, Halland, Jönköping, Kalmar, Kronoberg, Norrbotten, Skåne, Värmland, Västmanland, and Örebro (N=83,534)
N=62,603
(N_PCSM =8,076 died from PCa)1
ProMort II
(500 cases/500 controls)
Exclusions Case Control
Diagnostic slides not retrieved 53 48
Diagnosis based on cytology 41 23
Duplicate subjects 1 3
N=830
(404 cases/426 controls)
Light vs. virtual microscopy
N=60
Exclusions Case Control
No evaluable slides 19 34
Rejected at clinical annotation 1 6
No cancer 8 13
N=7532
(377 cases/379 controls)
Exclusions Case Control
Missing clinical M-stage in the NPCR, M1 in the medical chart 1
Missing PSA 2 2
Missing primary treatment 1 1
Missing clinical T-stage 4 4
Complete separation 2
N=7362
(367 cases/369 controls)

Figure 3.1. Flow chart of the selection of cases and controls for Study III and IV

Abbreviations: PCa, Prostate cancer; N_{PCSM}, The number of men who had died from prostate cancer; NPCR, the National Prostate Cancer Register of Sweden

[1] Based on the data extracted from NPCR June 5, 2020, but restricted to match conditions at April 11, 2017, when ProMort II was sampled

[2] Includes duplicate subjects (cases selected as controls (n=1) and controls selected more than once (n=3))

ProMort II cases and controls were selected from all men in the NPCR who were diagnosed with non-metastatic prostate cancer (i.e., not M1) between January 1, 1998 and December 31, 2015. Given that all diagnostic slides belonging to the selected cases and controls were to be centrally re-reviewed by the study pathologists, cases and controls were selected from 11 (out of 21) counties in the NPCR deemed most likely to respond to our request for slides (Dalarna, Gävleborg, Halland, Jönköping, Kalmar, Kronoberg, Norrbotten, Skåne, Värmland, Västmanland, and Örebro). Follow-up was available until December 31, 2015. Of all men who had died from prostate cancer by the end of the follow-up we randomly selected 500 cases and matched them to 500 men who had not died from prostate cancer (controls) by year and county of diagnosis. Controls had to be alive at the date of death of the respective case. The complete selection process is described in Figure 3.1.

Information on age, PSA, clinical TNM stage, primary and secondary Gleason pattern, and the Gleason score at diagnosis, as well as planned primary treatment, was abstracted from the NPCR. In addition, for the 404 cases and 426 controls for whom we had successfully scanned the diagnostic slides, we performed a medical chart review to extract detailed information on the clinicopathological features at diagnosis, biopsy procedure, primary treatment, pathology after radical prostatectomy, BCR, castration resistance, metastasis and death. Medical charts were successfully reviewed for 282 cases (69.8%) and 297 controls (69.7%). The extracted information was used when information abstracted from the NPCR was missing.

3.2.3.1 Slide digitalization and managing

We first retrieved the diagnostic biopsy slides from the Pathology wards across Sweden. Out of the 1,000 sampled men, the diagnostic slides belonging to 830 men (83%), 404 cases and 426 controls, were retrieved (Figure 3.1). The slides were then scanned at the Örebro University Hospital, Örebro, Sweden, using a Pannoramic 250 Flash II digital slide scanner (3DHistech Ltd., Budapest, Hungary) with a 40x objective. Scanned images had a resolution of 0.19 microns/pixel. In total, we scanned 5,536 slides.

After the scanning, the images were uploaded to a virtual microscopy system developed by the Centre for Advanced Studies, Research and Development in Sardinia (CRS4), Pula, Italy as a part of the ProMort I study (218). The virtual microscopy system is composed of two integrated components (Figure 3.2):

1. Ome_seadragon (https://github.com/crs4/ome_seadragon), a plugin for the Open Microscopy Environment Remote Objects (OMERO) platform (219) which enables viewing, handling and annotation of the 3DHistech images. The image management is based on the OMERO.server which supports over 140 different image formats and allows for storing of meta-information (e.g., classification TAGs or Regions of Interests (ROIs)). The ome_seadragon simplifies the integration of the images stored within OMERO into external web systems (220), adds Deep Zoom Image format support to OMERO and, through OpenSlide libraries (https://openslide.org), increases the number of supported image formats. The user side of ome_seadragon is a

specialized viewer developed from the open source viewer OpenSeadragon (https://openseadragon.github.io). Real time annotation tools which are based on paper.js libraries (http://paperjs.org) enable navigation through Whole Slide Imaging (WSI) and annotation by drawing different 2D shapes, as well as taking precise measures (e.g., ROI length or area).

2. The ProMort Image Management System (https://github.com/crs4/ProMort), a clinical annotation platform which manages the review worklist and clinical annotation process (i.e., definition and clinical annotation of ROIs). This software embeds the ome_seadragon plugin and allows users to navigate and annotate digital slides while acquiring the ROIs. Clinical annotations are performed via a dedicated user interface which has been designed specifically for ProMort.

Both components are web-based applications developed to run on all modern browsers and require no specific hardware or operative system. The pathologists involved in the study (Michelangelo Fiorentino and Francesca Giunchi) used either a desktop PC, with a 22 inch Olivetti OLISCREEN22 display, running the Google Chrome browser or a 2018 iPad Pro, with a 12.9 inch display, running the Safari browser.

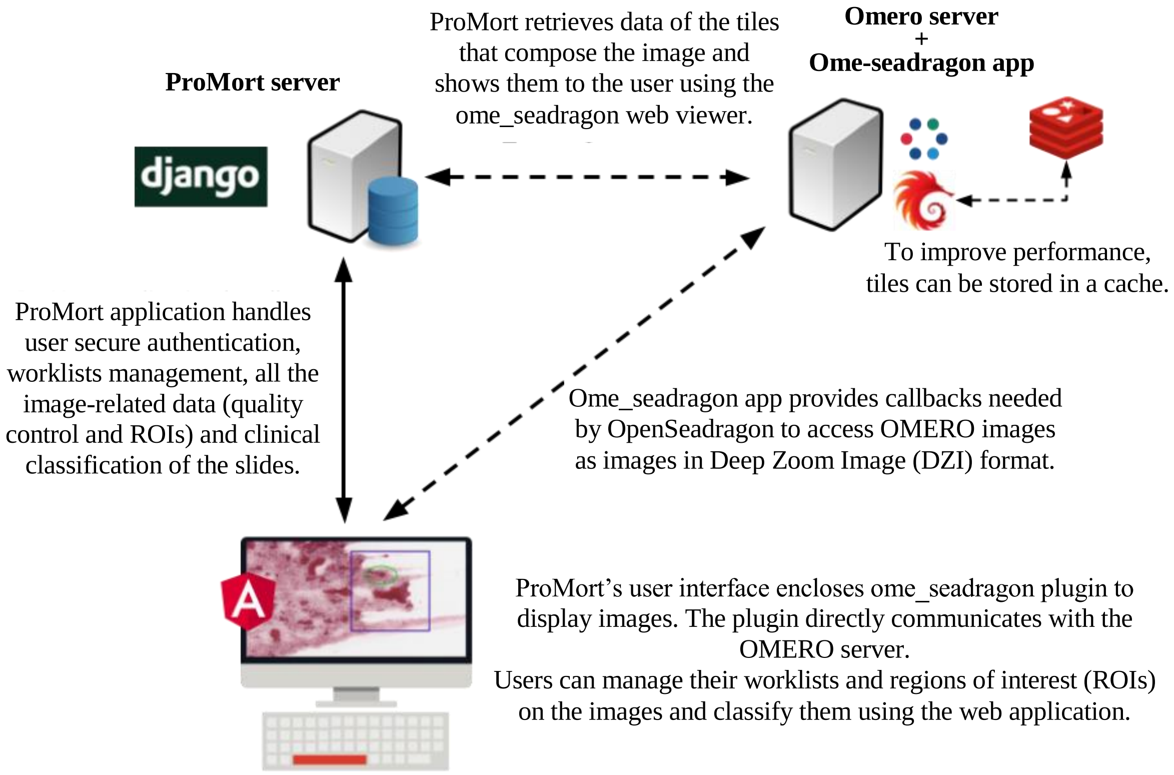

ProMort server
django
ProMort retrieves data of the tiles that compose the image and shows them to the user using the ome_seadragon web viewer.
Omero server
+
Ome-seadragon app
To improve performance, tiles can be stored in a cache.
ProMort application handles user secure authentication, worklists management, all the image-related data (quality control and ROIs) and clinical classification of the slides.
Ome_seadragon app provides callbacks needed by OpenSeadragon to access OMERO images as images in Deep Zoom Image (DZI) format.
ProMort's user interface encloses ome_seadragon plugin to display images. The plugin directly communicates with the OMERO server.
Users can manage their worklists and regions of interest (ROIs) on the images and classify them using the web application.

Figure 3.2. A simplified schematic representation of the virtual microscopy system developed by the Centre for Advanced Studies, Research and Development in Sardinia (CRS4), Pula, Italy, which was used for the central re-review in ProMort I and ProMort II

3.2.3.2 Histopathological review

The two study pathologists, with 6 and 13 years of experience respectively as dedicated genitourinary pathologists, performed the re-review of all scanned images according to the 2016 WHO classification of tumors of urinary system and male genital organs (97). The pathologists were blinded to the case-control status and to the original clinical and histopathological information of all slides.

We first selected 60 random cases and controls out of the 830 men in ProMort II whose diagnostic slides had been successfully scanned. The selected men were diagnosed in Örebro county (n=25) and Värmland county (n=35) (Figure 3.1). Slides belonging to these 60 men were used in Study III to evaluate the interchangeability of standard light microscopy and the above-described virtual microscopy system. The study pathologists reviewed all cores belonging to the 60 selected men using both light and virtual microscopy according to a pre-specified protocol (Figure 3.3). Using this protocol allowed us to estimate the intra- and inter-observer agreement for both light and virtual microscopy.

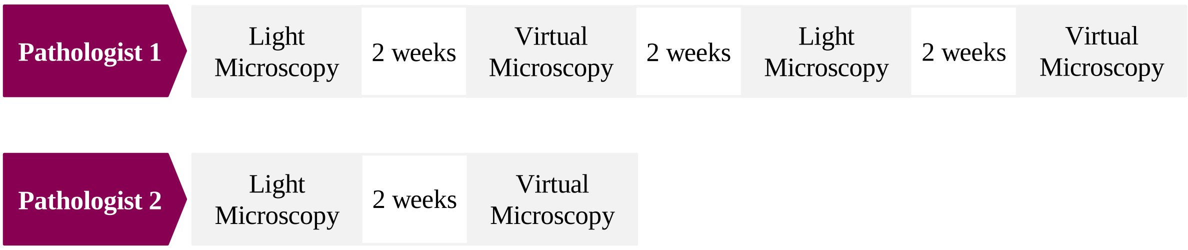
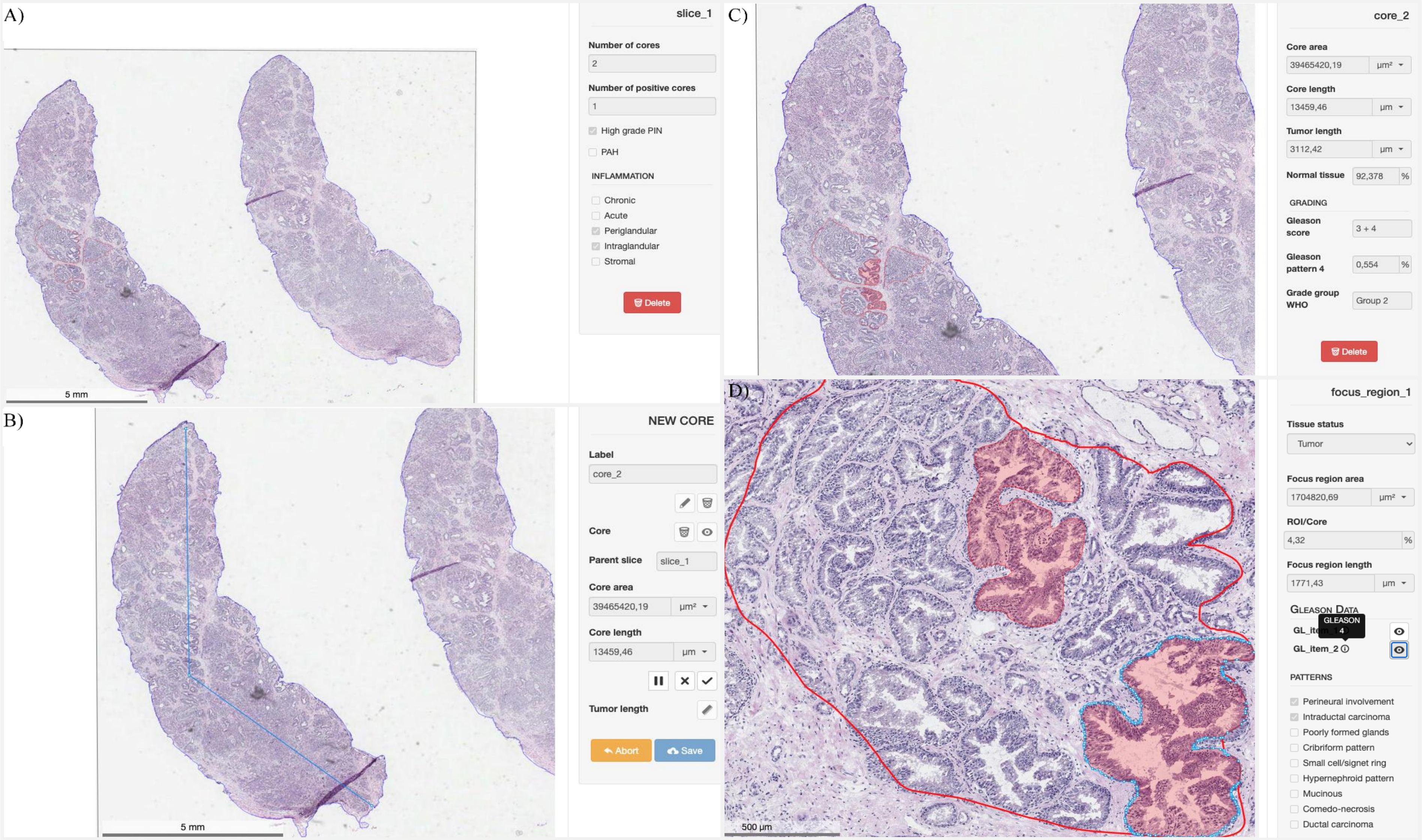


Figure 3.3. Pre-specified review protocol for evaluation of interchangeability of light and virtual microscopy

Slides belonging to the remaining 770 subjects were subsequently reviewed only by one pathologist. In total, 8,982 cores belonging to 770 subjects were reviewed, of which 3,713 cores belonging to 749 subjects contained cancer (Figure 3.1). A mock-up example of the review process is presented in the Figure 3.4. The features recorded during the re-review are presented in Table 3.2. Case level summaries were calculated as the sum across all cores for continuous features, and as presence or absence in at least one of the cores for binary features. The highest core-level GGs/Gleason score was used as an overall GGs/Gleason score for a case.



slice_1

Number of cores
2

Number of positive cores
1

☑ High grade PIN
☐ PAH

INFLAMMATION
☐ Chronic
☐ Acute
☑ Periglandular
☑ Intraglandular
☐ Stromal

Delete

NEW CORE

Label
core_2

Core

Parent slice
slice_1

Core area
39465420,19 µm²

Core length
13459,46 µm

Tumor length

Abort
Save

core_2

Core area
39465420,19 µm²

Core length
13459,46 µm

Tumor length
3112,42 µm

Normal tissue
92,378 %

GRADING

Gleason score
3 + 4

Gleason pattern 4
0,554 %

Grade group WHO
Group 2

Delete

focus_region_1

Tissue status
Tumor

Focus region area
1704820,69 µm²

ROI/Core
4,32 %

Focus region length
1771,43 µm

GLEASON DATA
GL_item_1
GLEASON 4
GL_item_2

PATTERNS
☑ Perineural involvement
☑ Intraductal carcinoma
☐ Poorly formed glands
☐ Cribriform pattern
☐ Small cell/signet ring
☐ Hypernephroid pattern
☐ Mucinous
☐ Comedo-necrosis
☐ Ductal carcinoma

Figure 3.4. A mock-up example of the histopathological review using the virtual microscopy system for a slide (A), core (B), core with cancer (C) and core with Gleason score 7 (D). The presented annotations serve only to demonstrate how the virtual microscopy system can be used. The annotated areas may not correspond to the actual cancer or Gleason score 7.

Table 3.2. Histopathological features recorded by the pathologists during the central re-review

Recorded for:	Feature	Unit
Core	Length	microns
	Area	squared microns
	Cancer	yes/no
Core with cancer	Length	microns
	Area	squared microns
	Primary Gleason pattern	3-5
	Secondary Gleason pattern	3-5
	ISUP 2014 Gleason score	6, 3+4/4+3, 8, 9, 10
	Gleason Grade Groups	1-5
	Poorly formed glands	yes/no
	Cribriform pattern	yes/no
	Hypernephroid pattern	yes/no
	Comedonecrosis	yes/no
	Small-cell/signet ring cell-like cancer	yes/no
	Perineural invasion	yes/no
	Intraductal carcinoma	yes/no
	Ductal carcinoma	yes/no
	Mucinous carcinoma	yes/no
Core with Gleason score 7	Area of Gleason pattern 4[1,2]	squared microns
Slide	Acute inflammation	yes/no
	Chronic inflammation	yes/no
	Periglandular inflammation	yes/no
	Intraglandular inflammation	yes/no
	Stromal inflammation	yes/no
	High-grade prostatic intraepithelial neoplasia (HGPIN)	yes/no
	Post-atrophic hyperplasia (PAH)	yes/no

Abbreviations: ISUP, International Society of Urological Pathology

[1] The percentage of Gleason pattern 4 was calculated as (area of Gleason pattern 4/tumor area)*100

[2] On light microscopy, the percentage of Gleason pattern 4 was assessed by "eye-balling" and categorized as <10, 10-19%, 20-29% etc.)

3.2.3.3 Pre-2005 Gleason grading

To approximate the pre-2005 Gleason grading we used information on cribriform pattern, poorly formed glands and hypernephroid pattern. Gleason pattern 4 (primary or secondary) was downgraded to Gleason pattern 3 whenever Gleason pattern 4 was assigned based on the:

1. Cribriform pattern only,
2. Poorly formed glands only and
3. Cribriform pattern and/or poorly formed glands.

In all three definitions hypernephroid pattern was graded as Gleason pattern 4. The pre-2005 Gleason score for each core was then calculated as the sum of the back-transformed primary and secondary Gleason pattern.

As a secondary approach, we used the diagnostic Gleason score recorded in the NPCR and in medical charts as the pre-2005 Gleason score. This approach was restricted to men diagnosed with prostate cancer before 2006.

3.3 Methodological considerations and statistical methods

Study I is a method application study that deals with some of the challenges related to the nested case-control study design. Given the non-standard methodology, the methods section for this study will be described in more detail. In Study II-IV we used standard methodology, and only a short summary of the statistical methods will be given here.

3.3.1 Study I

Given the difficulty in separating indolent from aggressive prostate cancer, especially for men with low- and intermediate-risk disease, we intended to use ProMort I to not only identify novel molecular markers of prostate cancer prognosis, but also to build a new prognostic model, or update an existing one. To be clinically useful, a prognostic model needs to be able to predict the absolute risk of the outcome of interest given the different combinations of the predictor values. Relative risks (e.g., odds ratios (ORs), risk ratios or HRs) are not directly interpretable and are used only to obtain absolute risks of the outcome.

To use ProMort I for prognostic modelling, we needed to deal with two issues:

1. Competing risks – men diagnosed with prostate cancer are on average old and, given the prolonged natural history of prostate cancer, especially among men with low- and intermediate-risk disease, they are more likely to die with, rather than from, prostate cancer.
2. Study design – the best design for prognostic modelling is a prospective cohort study. However, ProMort I is a nested case-control study.

Nested case-control studies are typically used for the estimation of relative risks. However, if adequate methods are used, nested-case control studies can be used to obtain unbiased estimates of absolute risks (221-226). These methods have also been extended to a setting where secondary outcomes are of interest (227, 228), and to the competing risk setting (229, 230).

3.3.1.1 Competing risks

In Study I, we focused on the cause-specific hazards approach for dealing with competing risks (231, 232) as the way controls were selected in ProMort I precluded the use of other approaches, such as the subdistribution hazards approach (233, 234).

The presence of competing risks implies that a subject is at risk of having K different events. In this setting, the cause-specific hazard function, $\lambda_k(t)$, represents the instantaneous risk of dying from the event k given that the subject is still alive at time t:

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, K = k | T \geq t)}{\Delta t}$$

The cumulative incidence function (CIF) for the event of interest k (i.e., prostate cancer death), $I_k(t)$, is a function of the cause-specific hazard for both the event of interest and the competing event(s) (i.e., death from other causes). $I_k(t)$ is defined as the probability of dying from the event k at the time t given that the subject can die from other causes:

$$I_k(t) = \int_0^t \lambda_k(u) \prod_{k=1}^{K} S_k(u) du$$

3.3.1.2 Estimation of absolute risks (CIFs)

To estimate the absolute risks using ProMort I, we used the inverse probability weighting method proposed by Samuelson (225). This method has been described in the context of the partial likelihood which is used to estimate parameters in the Cox proportional hazards model (235). In the partial likelihood, the baseline hazard function is not specified, and in order to estimate it we would need to use additional estimators, such as the Breslow estimator (235). Since we were interested in estimating both the HRs and the CIFs in Study I, we decided to use flexible parametric survival models instead of the Cox proportional hazards model (236, 237). In flexible parametric survival models, the baseline hazard function is fully specified and estimated by maximizing the full likelihood (236, 237).

Thus, to estimate CIFs in ProMort I, we used the weighted full likelihood method where the weights for cases and controls equal the inverse of their selection probability. In nested case-control studies, typically, all cases are sampled and their weight equals one. Given that the proportion of eligible cases which were not included in ProMort I was very small (1.5%), all cases were assigned with a weight of one. The selection probability for a control i, p_i, was calculated using an extension of the method proposed by Samuelson (225) which accounts for the presence of ties in failure times and for additional matching (221, 227):

$$p_i = 1 - \prod_{j: a_i \leq T_j \leq T_i} \left(1 - min\left(1, \frac{mb_{ji}}{n_{ji} - b_{ji}}\right)\right)$$

At each event time T_j, a subject i who entered the study at time a_i ($a_i \leq T_j$), was censored or failed at time T_i ($T_i \geq T_j$), and who satisfied the matching criteria, could be sampled as a control. m is the number of controls selected per case at each event time T_j. n_{ji} is a risk set at time T_j which satisfied the matching criteria and b_{ji} is the number of tied subjects that failed at the time T_j who satisfied the matching criteria.

The weights, ω_i, are defined as:

$$\omega_i = \frac{1}{p_i}$$

Of note, in this type of analysis, matching is broken and all unique individuals are pooled for the analysis (224). For controls who were selected more than once, we kept only one control record. For a control who later became a case, we kept only the case record.

3.3.1.3 Statistical analyses

We first estimated HRs and CIFs in ProMort I using the above-described inverse probability weighting approach. The flexible parametric model was fitted as described by Hinchliffe et al. (236). The cause-specific HRs and the corresponding 95% CIs of death from prostate cancer and death from other causes were estimated simultaneously (237, 238), and the CIFs were obtained by combining the cause-specific HR estimates (239). The HRs and CIFs estimated in ProMort I were compared to the HRs and CIFs estimated in the NPCR.

Then we used two extensions of the inverse probability weighting approach to the setting with more than one endpoint, including competing risks, where:

1. Both the competing risk cases and the competing risk controls were augmented ("Method 1") (240), and
2. Only the competing risk cases were augmented ("Method 2") (229).

The main idea behind the two methods is to reuse the controls and the cases selected for one endpoint as controls in the analysis of another endpoint, with or without a new control selection. The HRs and CIFs estimated using these two alternative approaches were also compared to the estimates from the NPCR.

Finally, we evaluated the bias in the HRs and CIFs estimated in ProMort I using the inverse probability weighting approach. To do so, we used the same selection criteria as for ProMort I to draw 1,500 random nested case-control subsamples from the NPCR. The bias in the log(HRs) was calculated as the absolute difference in the log(HRs) estimated in the 1,500 subsamples and the log(HRs) estimated in the NPCR. We also calculated the bias in CIFs of dying from prostate cancer at 5, 10 and 15 years of follow-up, as well as the coverage probability of their 95% CIs. The bias in CIFs was defined as the absolute difference in CIFs estimated in the 1,500 subsamples and CIFs estimated in the NPCR.

3.3.2 Study II

In Study II, we initially planned to use PCBaSe 4.0 to formally externally validate the most commonly used prostate cancer risk stratification tools. External validation and comparable information on each tool's ability to predict prostate cancer death in untreated patients are key for informed decision-making in clinical practice.

Formal external validation is not possible without having information on the intercept (i.e., baseline survival function in models analyzing time-to-event) and the linear predictor from the original prognostic model (241, 242). However, for most of the risk stratification tools evaluated in this study, information on the intercept and/or the linear predictor from the original models has not been published. Furthermore, most of the risk stratification tools have been developed or validated to predict BCR and not prostate cancer-specific death. We were thus not able to perform a formal external validation. Instead, we re-estimated the linear predictor for each risk stratification tool in PCBaSe and performed a head-to-head comparison of their prognostic performance in predicting prostate cancer death.

3.3.2.1 Statistical analyses

The prognostic performance of the different risk stratification tools was evaluated using a split-sample approach. The linear predictor for each risk stratification tool was re-estimated in the training dataset and the models were internally validated in the testing dataset.

We used the cause-specific hazards approach to account for the presence of competing events. The cause-specific HRs and 95% CIs for prostate cancer death and death from other causes were estimated using the Cox proportional hazards model (235). Time at risk was calculated from the date of diagnosis until the date of death, emigration or end of follow-up (December 31, 2016), whichever came first. The models predicting death from prostate cancer included only the risk grouping system, while the models predicting death from other causes also included age and year of diagnosis, the Charlson Comorbidity Index, marital status, education level and primary treatment. The cause-specific hazards for prostate cancer death and death from other causes were then combined to obtain the CIFs for prostate cancer death (243).

Model performance was evaluated in terms of discrimination and calibration. Discrimination was evaluated by the C-index adapted for competing risks (244, 245) in the full training datset, and stratified by primary treatment (active surveillance/watchful waiting, radical prostatectomy/radiation therapy and androgen deprivation therapy) and by year of diagnosis (1998-2002, 2003-2006, 2007-2016). The C-index was estimated by truncating the maximum follow-up time in the testing datasets at 1-19 years of follow-up. Calibration was evaluated by comparing the non-parametric CIFs (243) with the mean predicted CIFs at 5, 10 and 15 years of follow-up.

Of note, since multiple imputation was used to deal with the missing covariate information, the HRs, CIFs and C-indices for each risk stratification tool were combined across the imputed datasets (246).

3.3.3 Study III

In Study III, we assessed the repeatability (i.e., intra-observer agreement) and reproducibility (i.e., inter-observer agreement) of the ISUP 2014/WHO 2016 Gleason grading system evaluated on light microscopy and the virtual microscopy system which was developed for central re-review in ProMort I and ProMort II. The intra- and inter-observer agreement were evaluated within and between the two microscopy methods.

The agreement was evaluated using Cohen's kappa (κ) for binary variables (247), weighted Cohen's kappa (κ_w) with linear weights for ordinal variables (248), and Bland and Altman's limits of agreement for continuous variables (249). For descriptive purposes, $\kappa/\kappa_w<0$ was considered as no agreement, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 as almost perfect agreement (250).

3.3.4 Study IV

In Study IV, we conducted two separate analyses. Both analyses were performed using all subjects with complete information on all covariates (Figure 3.1).

We first evaluated if the Gleason grading system revisions have improved prostate cancer prognostication by comparing the prognostic performance of the pre-2005 Gleason score and the ISUP 2014 Gleason score in predicting death from prostate cancer. The pre-2005 Gleason score was approximated using two approaches. In the first approach, we back-transformed the ISUP 2014 Gleason score into the pre-2005 Gleason score as described in the section 3.2.3.3. In the second approach we used the diagnostic Gleason score restricted to all men diagnosed before 2006. ORs and 95% CIs of the association between the pre-2005 Gleason score and the ISUP 2014 Gleason score and death from prostate cancer were estimated using unconditional logistic regression adjusted for the matching variables (year and county of diagnosis, and follow-up time) and primary treatment. Multivariable models included untransformed age, PSA (transformed using restricted cubic splines with knots at 4.5, 16 and 105.65 ng/ml) and cT at diagnosis.

In the second analysis, we evaluated if additional histopathological features are independent predictors of death from prostate cancer. As above, ORs and 95% CIs of the association between each histopathological feature and death from prostate cancer were estimated using unconditional logistic regression adjusted for the matching variables and primary treatment. We then evaluated if each histopathological feature predicts death from prostate cancer independently of GGs, and, finally, independently of GGs, age, PSA and cT at diagnosis. We also explored the presence of statistical interaction with the GGs.

The prognostic performance for different models was evaluated by calculating the AUC (251). As a sensitivity analysis, we repeated all the analyses using conditional logistic regression.

3.4 Ethical considerations

In this thesis, we used information from registries, medical charts and diagnostic slides, all of which contain sensitive personal information. In case of release of sensitive data for research, all participants should be re-contacted. However, in large-scale research, re-contacting might not always be practicable, feasible or even possible. Many of the involved patients may no longer be alive and non-response could threaten study validity. For all the studies in this thesis, the Research Ethics Committee concluded that the potential benefits for the community outweigh the potential risks and the requirement for consent was waived. Since potential risks included violation of individual patient privacy, handling of sensitive data needed to be given the highest consideration. Below I will describe precautionary measures taken to minimize the risk of violation of individual patient integrity.

For the purpose of Study I and Study II, access to the registry data (i.e., the NPCR and PCBaSe) was possible only through the NPCR server in Uppsala, Sweden. The NPCR server has a very strict import/export policy and the risk for violation of individual patient privacy was very low. All men in ProMort I and ProMort II datasets, which were used in Study I, III and IV, were assigned a study-specific identification number by the NPCR upon the release of the data. Pseudonymized ProMort I and ProMort II data was then stored on a secure server at the Clinical Epidemiology Division at Karolinska Institutet, Stockholm, Sweden and handled according to the institution's guidelines for information security.

For ProMort I and ProMort II we performed a centralized re-review of the diagnostic slides, and for ProMort II additional information was extracted from the medical charts. For this reason, the key between a study-specific identification number and the personal identification number could be accessed by selected collaborators. For each man, the diagnostic slides and the medical charts were obtained from the diagnostic hospital/pathology ward and sent to Örebro University Hospital, Örebro, Sweden. After the diagnostic slides were scanned and the information was abstracted from the medical charts, personal identification numbers were replaced by the study-specific identification numbers and the slides and chart were returned to the respective institutions. The de-identified images were sent to the CRS4, Pula, Italy, and securely stored on the CRS4 server. The images were assigned with another random identification number at the time of histopathological review. Data extracted from the histopathological review and the data extracted from the medical charts were kept on the server at the Clinical Epidemiology Division at Karolinska Institutet and handled according to the institution's guidelines for information security.

4 Results

Brief summary of the results:

Study I. The relative risks of dying from prostate cancer estimated using the nested case-control study design were, as expected, comparable to the estimates from the underlying cohort. The estimates of the relative risks of dying from other causes were, however, biased, which introduced bias in the estimates of the absolute risks of dying from prostate cancer in the competing-risks setting.

Study II. The pretreatment risk stratification tools that performed best in predicting death from prostate cancer were the MSKCC nomogram, CAPRA score and CPG system. These tools discriminated best regardless of the primary treatment and year of diagnosis.

Study III. The repeatibility and reproducibility of the ISUP 2014 Gleason grading system within and between light and virtual microscopy was good. The repeatability and/or reproducibility for some of the rare, or rarely reported, features (e.g., intraductal cancer, inflammation, HGPIN and PAH), as well as for the percentage of Gleason pattern 4, was poor.

For all evaluated features, the agreement was similar within and between light and virtual microscopy which indicates interchangeability of light microscopy and our internally developed virtual microscopy system for the histopathological evaluation of prostate cancer.

Study IV. The ISUP 2014 Gleason score discriminated death from prostate cancer better than the pre-2005 Gleason score, likely due to classifying all cribriform patterns, rather than poorly formed glands, as Gleason pattern 4. In addition, comedonecrosis and HGPIN predicted death from prostate cancer independently of the GGs, age, PSA and cT at diagnosis.

4.1 Study I

In Study I, we evaluated if the nested case-control study design (ProMort I) can be used to estimate the relative and absolute risks of dying from prostate cancer in the competing risks setting.

When we compared the relative risks of dying from prostate cancer in ProMort I to those in the NPCR, the point estimates were overall similar (Figure 4.1A) and the mean absolute bias was generally close to zero for all covariates. The point estimates for death from other causes were, however, generally biased for ProMort I (Figure 4.1B) with the largest mean absolute bias for age (−3.813, −0.118, and 0.118 for ages ≤55.0 years, 65.1–75.0 years, and >75.0 years, respectively). Only a few subjects in the age ≤55.0 category had died from other causes and were sampled in ProMort I, leading to extreme and unreliable estimates.

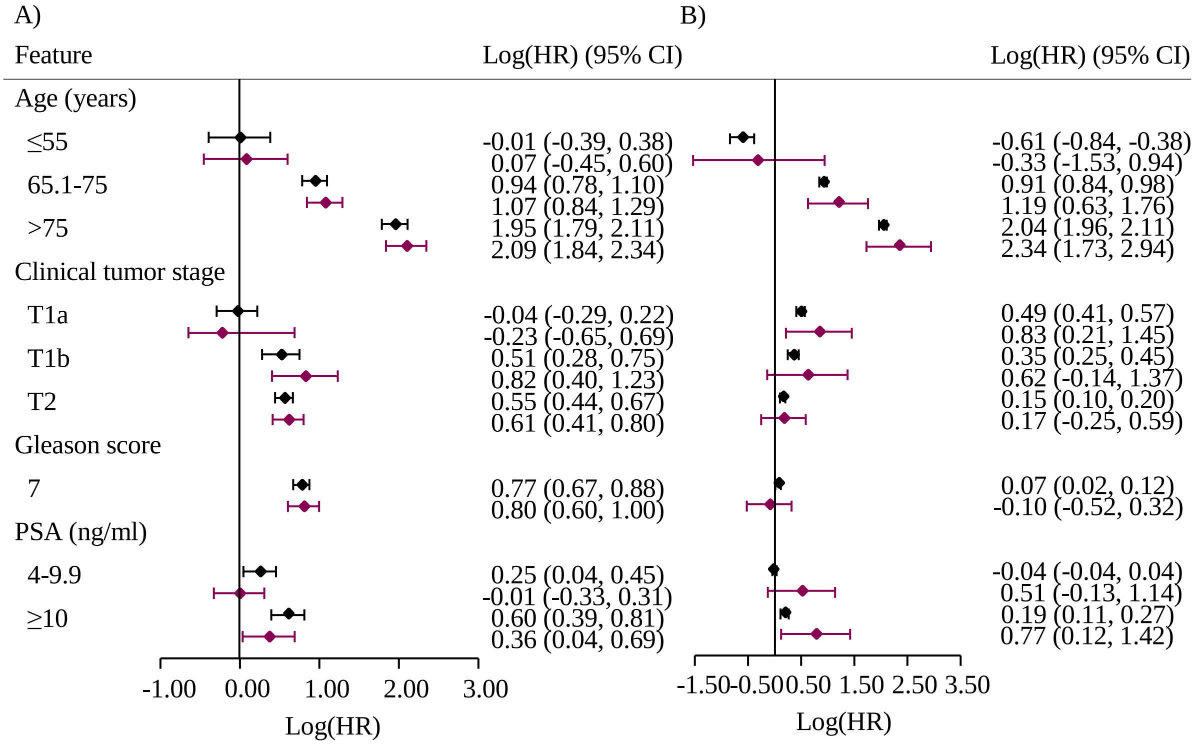


Figure 4.1. Cause-specific log hazard ratios (HR) for the risks of dying from prostate cancer (A) and other causes (B) in the NPCR (black) and ProMort (plum)

The 5-, 10-, and 15-year CIFs of death from prostate cancer were, overall, similar in ProMort I and the NPCR. However, the bias in the ProMort I estimates increased with age, and was the largest in the age >75.0 years category (Figure 4.2), where we also saw the largest mean absolute bias (0.011, 0.025, and 0.025 at 5, 10, and 15 years of follow-up, respectively).

Augmenting competing-risks cases (Method 2), and especially augmenting both the competing-risks cases and the controls (Method 1), reduced the bias in the estimates of the relative risks of dying from other causes and thus also the bias in the estimates of the absolute risks of dying from prostate cancer in the competing-risks setting (results presented in the supplementary material for Study I).

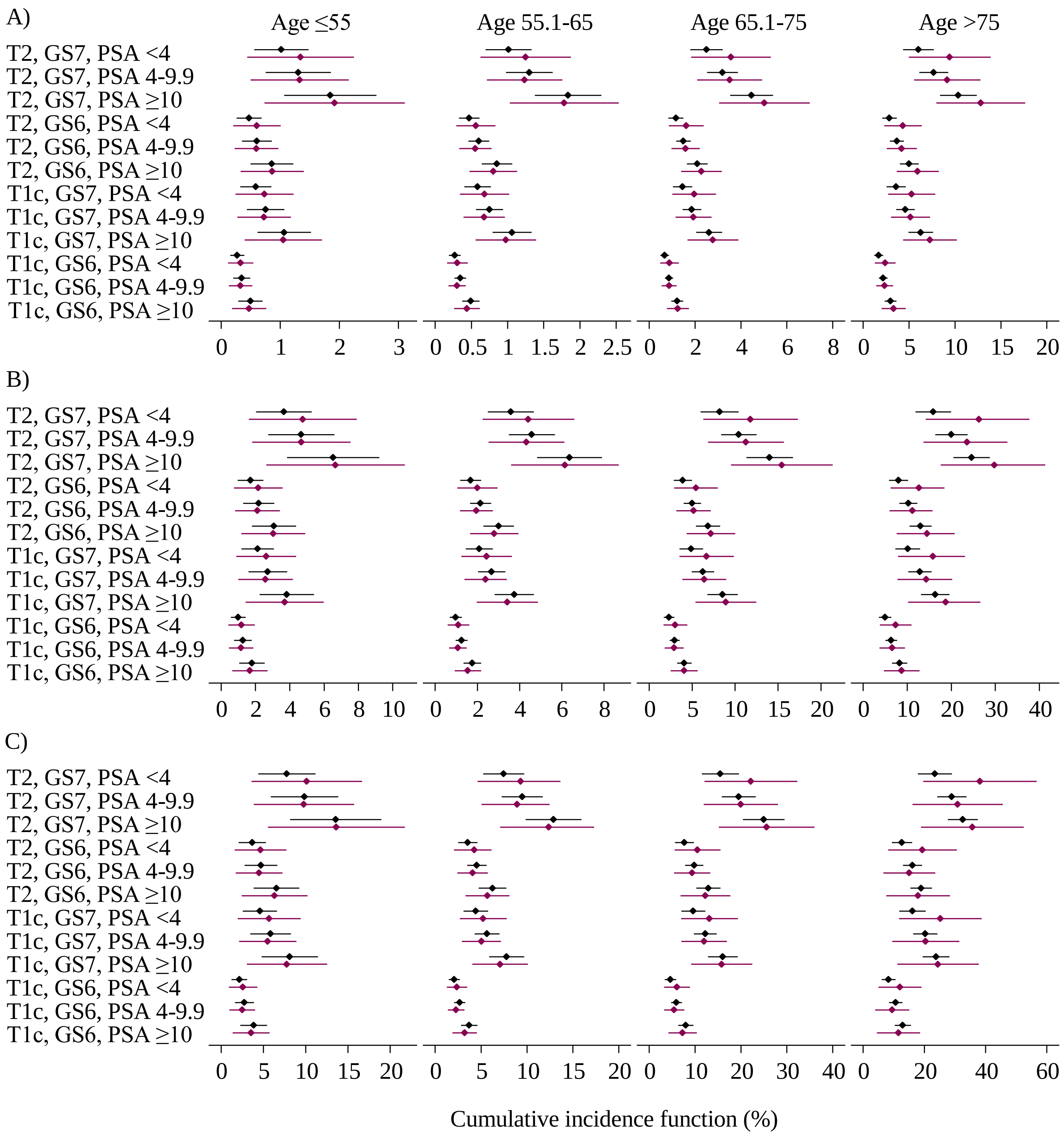


Figure 4.2. Cumulative incidence function and 95% confidence intervals of dying from prostate cancer at 5 (A), 10 (B) and 15 (C) years of follow-up in the NPCR (black) and ProMort I (plum)

4.2 Study II

In Study II, we used 139,515 men diagnosed with prostate cancer, of whom 15,961 (11.4%) died from prostate cancer, to systematically compare how well the most commonly used pretreatment risk stratification tools predict death from prostate cancer.

Overall, all tools discriminated death from prostate cancer well, and the C-index ranged from 0.73 (95% CI: 0.72-0.73) to 0.80 (95% CI: 0.80-0.81) at 10 years. As expected, the discrimination generally improved with the increasing granularity of the risk stratification tool and was lowest for the three-tiered D’Amico risk group system and highest for the MSKCC nomograms (Figure 4.3).

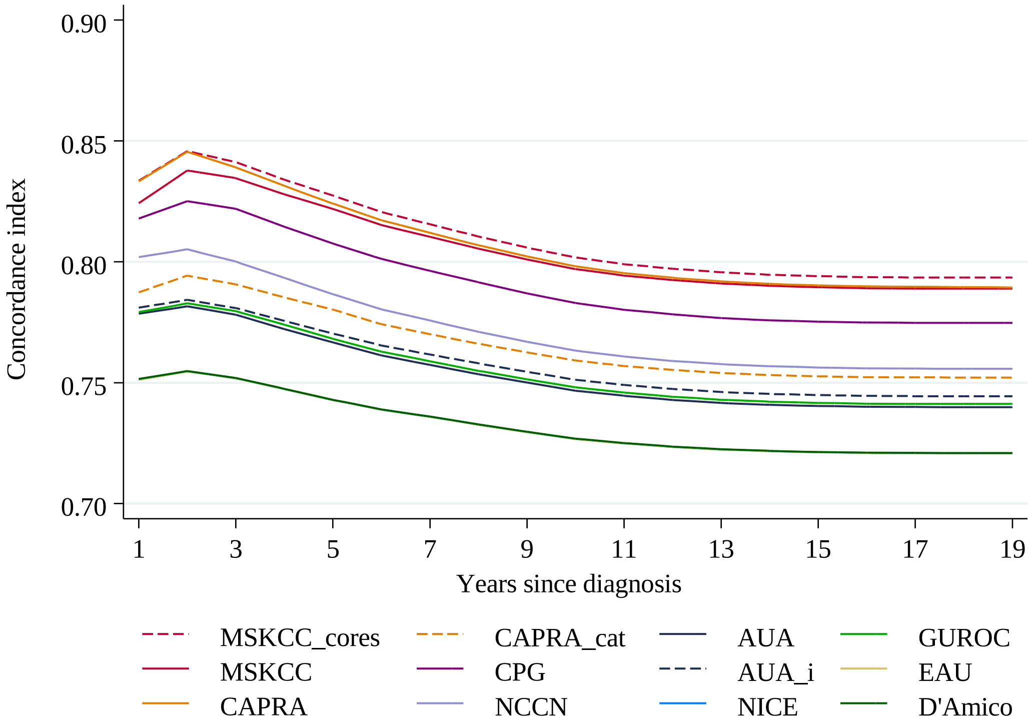


Figure 4.3. Pooled concordance index for prostate cancer death

The probabilities of dying from prostate cancer predicted using the D’Amico risk group system and the MSKCC nomogram are presented in Figure 4.4. The probability of dying from prostate cancer 15 years after diagnosis was 3.1%, 8.3% and 29.5% for men diagnosed with the low-, intermediate- and high-risk cancer according to the D’Amico criteria (Figure 4.4C). The individual probabilities of dying from prostate cancer predicted using the MSKCC nomogram varied widely within each D’Amico risk group. The predicted probabilities ranged from 1.6-20.6%, 1.6-40.5% and 1.6-49.4% within the low-, intermediate- and high- D’Amico risk groups, respectively (Figure 4.4C).

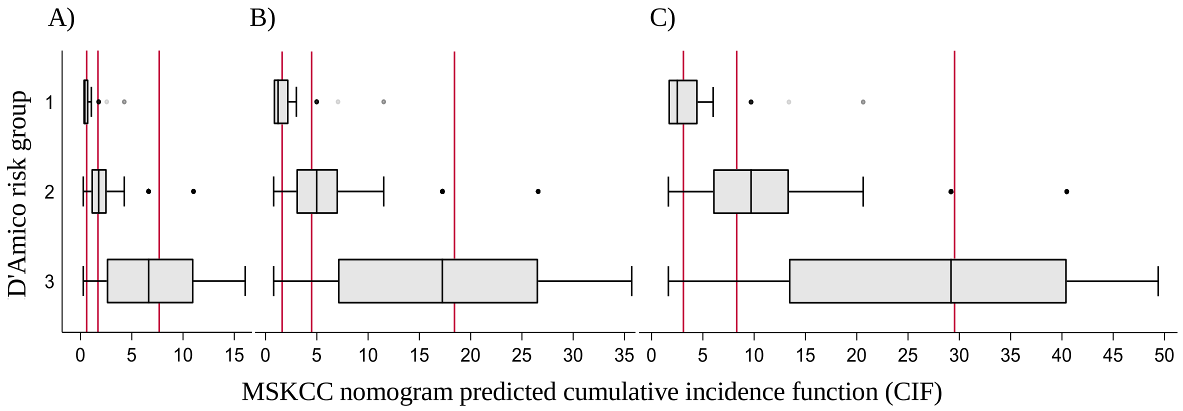


Figure 4.4. Range of the MSKCC nomogram predicted probabilities of dying from prostate cancer within the D’Amico risk groups. We used the average predicted cumulative incidences in the deciles of the MSKCC nomogram. Vertical red lines indicate the average predicted probability in each D’Amico risk group at 5 (A), 10 (B) and 15 (C) years of follow-up

When evaluated in different treatment groups, the discrimination was similar among men in the active surveillance/watchful waiting group and men in the radical prostatectomy/radiation therapy group. However, among men treated with primary androgen deprivation therapy, the discrimination was substantially poorer ranging from 0.56 (95% CI: 0.55-0.56) to 0.64 (95% CI: 0.63-0.65). For all risk stratification tools, the discrimination improved in more recently diagnosed cohorts. Among men diagnosed before 2003, the discrimination ranged from 0.66 (95% CI: 0.65-0.67) to 0.73 (95% CI: 0.72-0.75) compared to 0.77 (95% CI: 0.76–0.78) to 0.85 (95% CI: 0.84-0.86) among men diagnosed 2007-2016.

For all risk stratification tools, the observed and predicted probabilities of dying from prostate cancer were generally similar. However, the predicted probabilities were generally underestimated, especially at 5 years of follow-up, in the highest-risk category of the NCCN system, CAPRA score, and deciles of the MSKCC linear predictor.

4.3 Study III

In Study III, we evaluated if light and virtual microscopy can be used interchangeably for the histopathological evaluation of prostate cancer by examining the repeatability (i.e., intra-observer agreement) and reproducibility (i.e., inter-observer agreement) of the ISUP 2014/WHO 2016 Gleason grading system using both microscopy methods.

The intra-observer agreement for most of the features evaluated on the core level was similar for the two methods (Figure 4.5A), indicating good repeatability regardless of the method used. For the Gleason related features, the agreement ranged from substantial to almost perfect (primary Gleason pattern: κ_{wLM}=0.80 vs. κ_{wVM}=0.84; secondary Gleason pattern: κ_{wLM}=0.67 vs. κ_{wVM}=0.66; GGs: κ_{wLM}=0.85 vs. κ_{wVM}=0.84). For features which were rare, or for which reporting is not obligatory in clinical practice, such as intraductal cancer, the agreement was somewhat lower, but better when virtual microscopy was used (Figure 4.5A). The intra-observer agreement for the percentage of Gleason pattern 4 was overall poor, but, again, somewhat better on virtual microscopy.

The inter-observer agreement for the Gleason-related features was similar for the two methods (Figure 4.5B), ranging from moderate/substantial to almost perfect (primary Gleason pattern: κ_{wLM}=0.72-0.90 vs. κ_{wVM}=0.78-0.80; secondary Gleason pattern: κ_{wLM}=0.58-0.75 vs. κ_{wVM}=0.67-0.68; GGs: κ_{wLM}=0.80-0.89 vs. κ_{wVM}=0.83) indicating good reproducibility regardless of the method used. For the remaining features, the agreement was somewhat lower, but similar for the two methods, except for mucinous carcinoma, perineural invasion, small-cell signet ring cell-like carcinoma, HGPIN and chronic inflammation, where it was better for light than virtual microscopy (Figure 4.5B). The inter-observer agreement for the percentage of Gleason pattern 4 was overall poor.

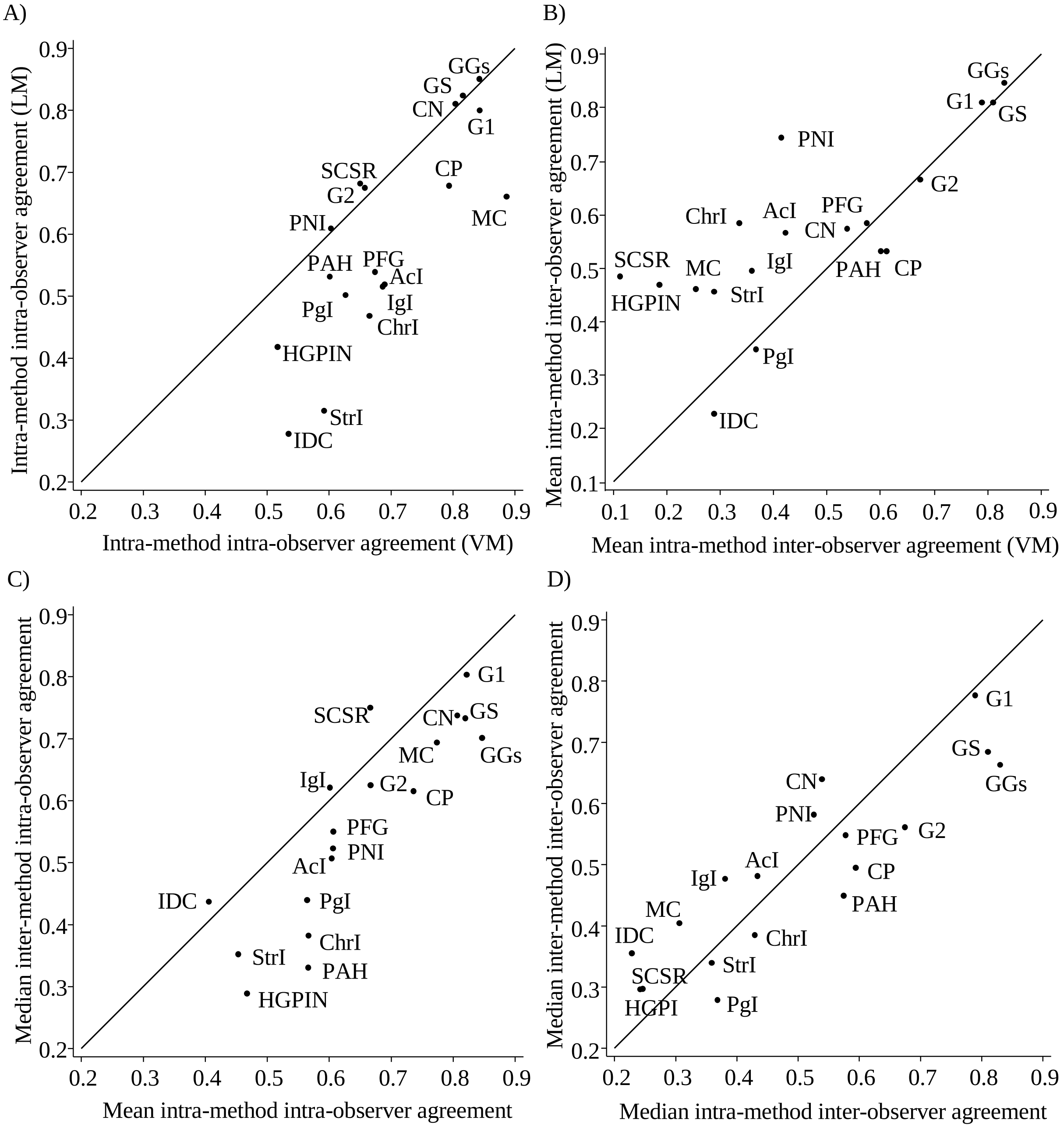


Figure 4.5. Repeatability (A), reproducibility (B) and interchangeability (C, D) plot for all characteristics evaluated on the core and slide level

Abbreviations: LM, Light microscopy; VM, Virtual microscopy; GGs, Gleason Grade Groups; GS, Gleason score; CN, Comedonecrosis; G1, Primary Gleason pattern; SCSR, Small-cell signet ring cell-like cancer; G2, Secondary Gleason pattern; CP, Cribriform pattern; MC, Mucinous cancer; PNI, Perineural invasion; PFG, Poorly formed glands; PAH, Postatrophic hyperplasia; AcI, acute inflammation. IgI, Intraglandular inflammation; PgI, Periglandular inflammation; ChrI, chronic inflammation; HGPIN, High-grade prostatic intraepithelial neoplasia; StrI, Stromal cancer; IDC, Intraductal cancer

Finally, the median agreement between light and virtual microscopy was similar to the average/median agreement within the two methods, both when it was evaluated intra-observer (Figure 4.5C) and inter-observer (Figure 4.5D), indicating interchangeability of light and virtual microscopy for most of the evaluated features. However, for most of the features evaluated on the slide level, median inter-method intra-observer agreement was lower than the average intra-method intra-observer agreement (Figure 4.5C), probably due to the higher intra-observer agreement on virtual microscopy. The absolute difference between the percentage of Gleason pattern 4 measured using light vs. virtual microscopy was up to 22 percentage points larger for both the intra- and inter-observer comparisons. These results indicate overestimation of the percentage of Gleason pattern 4 when light microscopy is used.

4.4 Study IV

In Study IV, we first evaluated if the ISUP 2005 and 2014 Gleason grading revisions have improved prediction of death from prostate cancer. Then we investigated if any additional histopathological feature predicts death from prostate cancers independently from the GGs, as well as independently from GGs, age, PSA and cT at diagnosis.

The GGs and ISUP 2014 Gleason score performed equally and better than the pre-2005 Gleason score back-transformed using only cribriform pattern or both cribriform and poorly formed glands in discriminating death from prostate cancer in univariable (p=0.003 and p=0.005, respectively) and multivariable models (p=0.066 and p=0.097, respectively). There was, however, no difference in discrimination between the ISUP 2014 Gleason score and the pre-2005 Gleason score back-transformed using only poorly formed glands (p=0.296 and p=0.830 in univariable and multivariable models) (Table 4.1). These results indicate that the small improvement in discrimination of the ISUP 2014 Gleason score vs. pre-2005 Gleason score could be due to classifying all cribriform patterns, rather than poorly formed glands, as Gleason pattern 4.

Table 4.1. Prognostic performance of univariable and multivariable models with different Gleason grading system revisions in predicting death from prostate cancer

	Univariable analysis[1]			Multivariable analysis[2]		
	AUC	95% CIs		AUC	95% CIs	
Pre-2005 Gleason score[3]	0.820	0.790	0.850	0.845	0.818	0.873
Pre-2005 Gleason score[4]	0.832	0.803	0.861	0.853	0.826	0.880
Pre-2005 Gleason score[5,6]	0.819	0.789	0.849	0.844	0.816	0.872
ISUP 2014 Gleason score	0.840	0.811	0.868	0.854	0.828	0.881
Gleason Grade Groups	0.839	0.811	0.868	0.854	0.827	0.881

Abbreviations: AUC, Area under the receiver operating curve; ISUP, International Society of Urological Pathology

[1] Adjusted for the matching variables (year and county of diagnosis, and follow-up time) and primary treatment

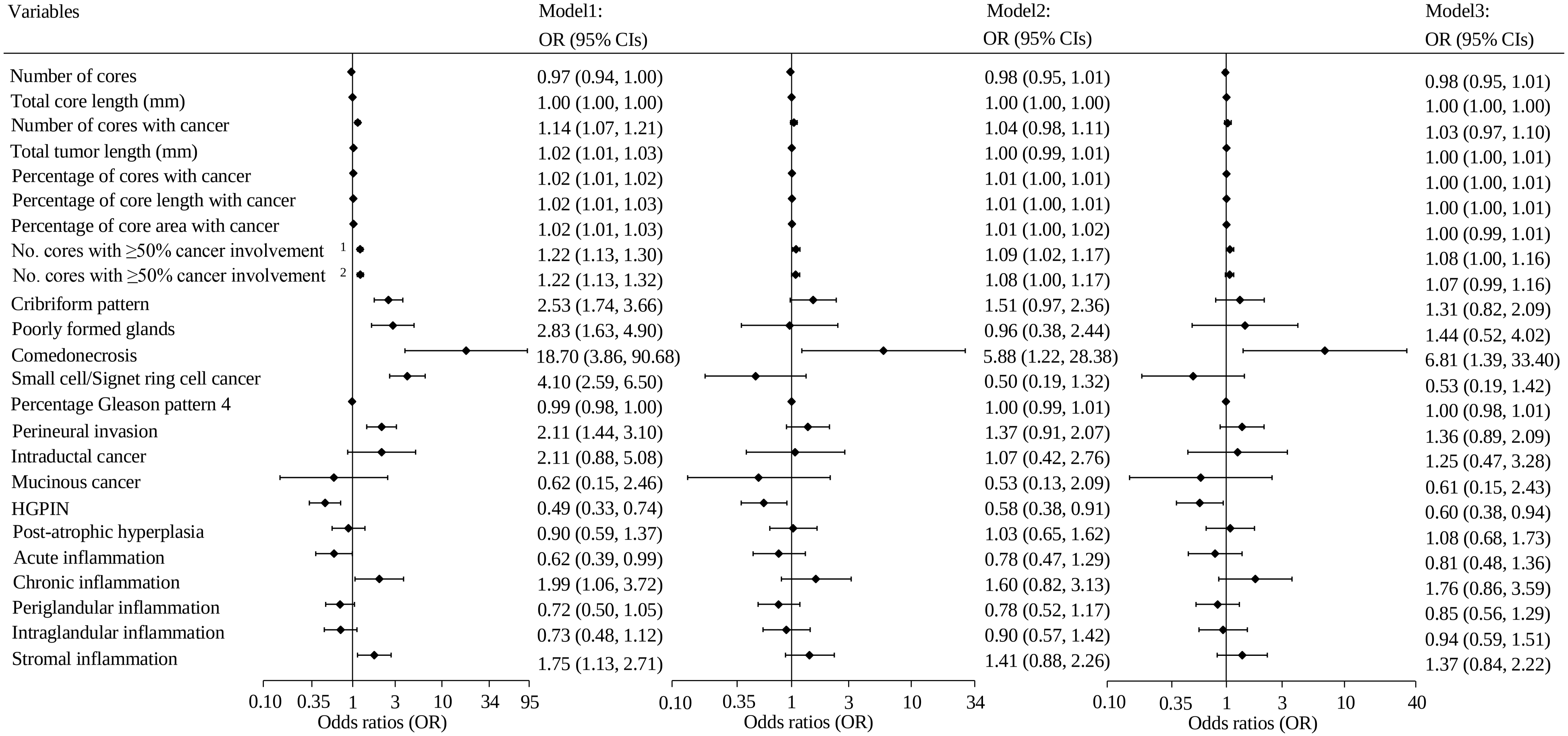
[2] Adjusted for the matching variables, clinical tumor stage, age, PSA level and primary treatment

[3] Back-transformation using the cribriform pattern only

[4] Back-transformation using the poorly formed glands only

[5] Back-transformation using the cribriform pattern and/or poorly formed glands

[6] 362 cases and 360 controls used in the analysis due to the complete separation



Variables	Model1: OR (95% CIs)	Model2: OR (95% CIs)	Model3: OR (95% CIs)
Number of cores	0.97 (0.94, 1.00)	0.98 (0.95, 1.01)	0.98 (0.95, 1.01)
Total core length (mm)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Number of cores with cancer	1.14 (1.07, 1.21)	1.04 (0.98, 1.11)	1.03 (0.97, 1.10)
Total tumor length (mm)	1.02 (1.01, 1.03)	1.00 (0.99, 1.01)	1.00 (1.00, 1.01)
Percentage of cores with cancer	1.02 (1.01, 1.02)	1.01 (1.00, 1.01)	1.00 (1.00, 1.01)
Percentage of core length with cancer	1.02 (1.01, 1.03)	1.01 (1.00, 1.01)	1.00 (1.00, 1.01)
Percentage of core area with cancer	1.02 (1.01, 1.03)	1.01 (1.00, 1.02)	1.00 (0.99, 1.01)
No. cores with ≥50% cancer involvement [1]	1.22 (1.13, 1.30)	1.09 (1.02, 1.17)	1.08 (1.00, 1.16)
No. cores with ≥50% cancer involvement [2]	1.22 (1.13, 1.32)	1.08 (1.00, 1.17)	1.07 (0.99, 1.16)
Cribriform pattern	2.53 (1.74, 3.66)	1.51 (0.97, 2.36)	1.31 (0.82, 2.09)
Poorly formed glands	2.83 (1.63, 4.90)	0.96 (0.38, 2.44)	1.44 (0.52, 4.02)
Comedonecrosis	18.70 (3.86, 90.68)	5.88 (1.22, 28.38)	6.81 (1.39, 33.40)
Small cell/Signet ring cell cancer	4.10 (2.59, 6.50)	0.50 (0.19, 1.32)	0.53 (0.19, 1.42)
Percentage Gleason pattern 4	0.99 (0.98, 1.00)	1.00 (0.99, 1.01)	1.00 (0.98, 1.01)
Perineural invasion	2.11 (1.44, 3.10)	1.37 (0.91, 2.07)	1.36 (0.89, 2.09)
Intraductal cancer	2.11 (0.88, 5.08)	1.07 (0.42, 2.76)	1.25 (0.47, 3.28)
Mucinous cancer	0.62 (0.15, 2.46)	0.53 (0.13, 2.09)	0.61 (0.15, 2.43)
HGPIN	0.49 (0.33, 0.74)	0.58 (0.38, 0.91)	0.60 (0.38, 0.94)
Post-atrophic hyperplasia	0.90 (0.59, 1.37)	1.03 (0.65, 1.62)	1.08 (0.68, 1.73)
Acute inflammation	0.62 (0.39, 0.99)	0.78 (0.47, 1.29)	0.81 (0.48, 1.36)
Chronic inflammation	1.99 (1.06, 3.72)	1.60 (0.82, 3.13)	1.76 (0.86, 3.59)
Periglandular inflammation	0.72 (0.50, 1.05)	0.78 (0.52, 1.17)	0.85 (0.56, 1.29)
Intraglandular inflammation	0.73 (0.48, 1.12)	0.90 (0.57, 1.42)	0.94 (0.59, 1.51)
Stromal inflammation	1.75 (1.13, 2.71)	1.41 (0.88, 2.26)	1.37 (0.84, 2.22)

Figure 4.6. Odds ratios (OR) and 95% confidence intervals (CIs) of the association between different histopathological characteristics and death from prostate cancer

Model 1, Adjusted for the matching variables (year and county of diagnosis and follow-up time) and primary treatment; Model 2, Adjusted for the matching variables, Gleason Grade Groups and primary treatment; Model 3, Adjusted for the matching variables, Gleason Grade Groups, clinical tumor stage, age and PSA at diagnosis and primary treatment

[1] Calculated as a number of cores with (tumor length/core length)*100≥50

[2] Calculated as a number of cores with (tumor area/core area)*100≥50

Almost all evaluated histopathological features were predictors of death from prostate cancer in the univariable analysis. However, only comedonecrosis, HGPIN and the number of cores with ≥50% cancer involvement predicted death from prostate cancer independently of the GGs (Figure 4.6). After additional adjustment for age, cT and PSA at diagnosis, comedonecrosis (OR: 6.8, 95% CIs: 1.4-33.4) and HGPIN (OR: 0.6, 95% CIs: 0.4-0.9) remained individual predictors (Figure 4.6), however with minimal impact on the discrimination (AUC: 0.86 vs 0.85 for both features). We also evaluated if there were any statistical interactions between the histopathological factors and the GGs. Adding an interaction term improved the model only for the number of cores (likelihood ratio test: p=0.002) and percentage of Gleason grade 4 (likelihood ratio test: p=0.033). The percentage of Gleason grade 4 was associated with death from prostate cancer only among men with GG3 (OR: 1.05, 95% CIs: 1.01-1.09). The impact of the added interaction term on discrimination was, however, minimal.

5 Discussion

This thesis focuses on improving prognostication for men with localized prostate cancer. Although it does not include a comprehensive evaluation of all aspects of prognostic modelling in prostate cancer, it addresses several very important issues related to study design, model development and model updating. Rather than discussing the study results one by one as is done within each study manuscript, in the following text I will discuss our findings in the context of the aforementioned issues.

5.1 Study design for prognostic modelling in localized prostate cancer

Localized prostate cancer, and especially low- and intermediate-risk prostate cancer, has a prolonged natural history, and long-term cancer-specific mortality in these patients is quite low. In a relatively recent study by Klotz et al. (252), the cancer-specific survival for men on active surveillance was 98.1% and 94.3% at 10 and 15 years of follow-up, respectively. Similarly, in PCBaSe, the observed 15-year cancer-specific mortality for men with D'Amico low- and intermediate-risk cancer who were treated with mixed modalities was 5.1% and 12.2%, respectively (253). Thus, even 15 years after diagnosis, only a small proportion will have died from prostate cancer, which makes prostate cancer death a rare outcome in men with low- and intermediate-risk disease. This also means that to study such a rare outcome, we need large cohorts of men followed-up for a very long time. Collecting additional information from such large cohorts (e.g., through a central histopathological re-review) makes cohort study design unfeasible. This issue is not an uncommon issue and is typically addressed by using well-known cost-effective cohort sub-sampling designs, such as case-cohort or nested case-control designs (254, 255).

5.1.1 The nested case-control study design

ProMort I and ProMort II were sampled with the above-described reasoning in mind. ProMort I is an ambitious ongoing project which aims to identify novel tissue-based molecular prognostic markers for men with low- and intermediate-risk prostate cancer. ProMort II, on the other hand, aims at identifying novel histopathological markers for men with localized (i.e., low-, intermediate- or high-risk) prostate cancer. For both ProMort I and ProMort II, the nested case-control design is an appropriate study design for identifying novel prognostic markers as it gives unbiased relative risk estimates, which is also what we observed in Study I. However, to understand if a novel marker actually improves prediction, it should be evaluated in addition to established prognostic factors, i.e., the "gold standard" prognostic model (256). This brings us to two very important questions:

1. Is the nested case-control study design appropriate for the development of models to predict death from prostate cancer?
2. What is the "gold standard" prognostic model in prostate cancer?

5.1.1.1 Predicting death from prostate cancer

For a prognostic model to be clinically useful, estimates of absolute risks are essential. It has been shown that the nested case-control design can be used to obtain unbiased estimates of absolute risks if appropriately analyzed (221, 222, 224-226, 230). Appropriate analysis here implies the use of weights which are based on the inverse of the probability of being selected into the study (221, 225, 226). The inverse probability weighting methods are, however, underutilized in epidemiological practice, mostly because the weights are difficult to obtain without access to the underlying cohort. When the underlying cohort is available, the calculation of weights can be implemented in R using the multileNCC package (257), or in Stata, using the code we published as a part of Study I (218). Once the weights are correctly estimated, these methods deliver virtually identical information compared to the analysis of the entire cohort with the advantage of reduced costs and reduced computational burden.

5.1.1.2 Competing events

In addition to the estimates of prostate cancer-specific survival, Klotz et al. reported that, in the same cohort of men on active surveillance described above, overall survival was 80% and 62% at 10 and 15 years of follow-up, respectively (252). This means that 10 and 15 years after a prostate cancer diagnosis, most men will be alive, and of those who have died, most will have died from causes other than prostate cancer (258-260). In the setting where a man with prostate cancer is more likely to die from a competing event, using the nested case control-study design to predict death from prostate cancer becomes more complicated.

Nested case-control studies are selected on the outcome. In ProMort I, the cases were all men who had died from prostate cancer and the controls were selected from the men who were still alive and at risk of dying from prostate cancer at the time the corresponding case died. It has been shown that the inverse probability weighting methods can still be used to make valid inferences for secondary, nonexclusive, outcomes (227, 228). However, to make inference on competing risks, these methods need to be extended further. To use nested case-control studies for competing risks analysis, we need to either modify the control sampling (233), treat the nested case-control study as a missing-data problem (240) or sample an additional set of competing risk cases (229) or both competing risk cases and controls (230). In Study I, we showed that simply using the inverse probability weighting method leads to biased estimates of the relative risks of dying from other causes, and that this translates into biased estimates of the absolute risks of dying from prostate cancer. However, adding competing risk cases to the analysis, or even better, adding both competing risk cases and controls, minimized these biases in Study I. This shows that nested-case control study design can indeed be used for the development of models predicting death from prostate cancer in the competing risk setting, but only with additional extensions to the design.

It is important to note that when the nested case-control sampling has already been done and the underlying cohort is not available, we can no longer modify the control sampling (233) nor can we treat the case-control study as a missing-data problem (240). If the nested case-control study design is to be used for more than just identification of novel prognostic markers, the above-described issues should be considered at the study design phase. Rather than trying to fix the issues by augmenting competing risks controls and/or cases (229, 230), or by resampling controls for the event of interest by using a different sampling strategy (233), both of which carry additional cost, we should instead consider if alternative designs, such as the case-cohort design, are more feasible.

5.1.1.3 Non-rare (common) events

As mentioned before, both ProMort I and ProMort II are nested case-control studies. However, there is an important difference in their sampling. In ProMort II, where we sampled cases and controls among all men with non-metastatic prostate cancer, death from prostate cancer was not as rare as it was among the men with low- and intermediate-risk cancer who were sampled in ProMort I. In fact, in the 11 counties in the NPCR from which ProMort II cases and controls were sampled, 8,076 men (out of 62,603) had died from prostate cancer by the end of follow-up (Figure 3.1). In this population, prostate cancer death is a non-rare, or even a common, event. Sampling all cases, as is typically done in nested case-control studies, would thus be unfeasible. Instead, in ProMort II, we used a modified nested case-control sampling design to select a sub-sample of all cases (n=500) and their corresponding controls (n=500). Such a nonrepresentative sampling of cases was outlined in a paper by Langholz and Borgan in 1995 (261), but, to the best of my knowledge, there were no examples of its application. A recent doctoral thesis, however, described how the inverse probability weighting methods can be extended to modified nested case-control and case-cohort studies (262). By using modified inverse probability weights we could, in theory, obtain estimates of the individual absolute risks of dying from prostate cancer in ProMort II. Whether we could also account for the competing events using the above-described methods, remains, for now, unclear.

5.2 Current prognostic models in prostate cancer

Pretreatment prognostic models, or risk stratification tools, are critical not only for the appropriate treatment decision-making at the time of diagnosis, but also for benchmarking the utility of novel prognostic markers. The risk stratification tools used in all major prostate cancer guidelines (EAU (37, 50), NICE (33), GUROC (34), AUA (35), and NCCN (36)) are based on the three-tiered D’Amico risk group system (49). It is becoming increasingly clear that, even with the additional sub-stratification of the D’Amico low-, intermediate- and high-risk groups, these tools are just too crude to be used as the “gold standard” (263). Indeed, in Study II, we showed that subdividing the NCCN low-risk group into very low and low, the NCCN/AUA intermediate-risk group into favorable and unfavorable, and, finally, the NCCN high-risk group into the high and very high has a minimal impact on discrimination. Which of the currently used risk stratification tools is then best at predicting death from prostate cancer?

5.2.1 The best-performing prognostic model

Of all the compared risk stratification tools in Study II, the more complex, model-based tools such as the MSKCC nomogram (83), CAPRA score (60) and CPG risk groups (53), discriminated death from prostate cancer better than D'Amico and D'Amico-derived risk groups. That discrimination improves when more complex risk stratification tools are used is an expected consequence of finer risk stratification and use of continuous information. However, it is important to note that the C-index is not a function of the actual predicted probabilities (264). The probability of the correct ranking of risks in pairs of men with and without the outcome is not a relevant measure of clinical utility. While in Study II we used the C-index as a convenient measure to rank risk stratification tools according to how well they discriminate death from prostate cancer, it is not obvious if a higher C-index translates into improved prediction of individual probabilities of dying from prostate cancer.

When deciding on the appropriate treatment, a clinician is primarily interested in the individual probability of death from prostate cancer. The best performing tool should thus predict this probability as accurately as possible. All risk stratification tools evaluated in Study II had similar observed and predicted probabilities of death from prostate cancer with some underestimation in the highest-risk categories for the MSKCC nomogram, CAPRA score and NCCN risk groups. However, the compared observed and predicted probabilities are population averages. For the risk stratification tools which do not finely stratify men with higher risk prostate cancer, the average predicted probabilities are influenced by the larger number of men with lower risk prostate cancer within the same group. To demonstrate how lumping together a large group of men with different risks of dying from prostate using the D'Amico risk groups may influence clinical decision-making, we plotted a distribution of the individual risks predicted using the best performing tool, the MSKCC nomogram, within each D'Amico risk group (Figure 4.4). Although perhaps not sufficiently emphasized in the published paper, this joint distribution plot is striking. At 15 years, within the D'Amico intermediate-risk group, 25% of the men had a MSKCC predicted probability of dying from prostate cancer between 1.6% and 6%. This range corresponds to the range of MSKCC predicted values within the D'Amico low-risk group. Furthermore, over 50% of the men had a MSKCC predicted probability higher than the D'Amico intermediate-risk group probability (8.3%), and of them, 25% had a predicted probability of 13.4-20.6%, and some extreme cases had a predicted probability of 49.4% (Figure 4.4). A similar wide range of MSKCC predicted probabilities was present also within the NCCN risk groups (data not shown), which, again, demonstrates that simply sub-stratifying low-, intermediate- and high-risk groups is not sufficient. This finer risk stratification could surely facilitate treatment decision for some of the men in the D'Amico intermediate risk group. Of note, we performed no formal quantification of the clinical usefulness/net benefit of the prediction models, such as decision curve analysis, in Study II.

5.2.2 The best-performing vs. the "gold-standard" prognostic models

There is no formal definition of the "gold standard" prognostic model in prostate cancer. Intuitively, such a model should be parsimonious and contain relevant, readily available baseline features which are established and strong predictors of death from prostate cancer. Special attention should be given to the functional form of continuous variables and presence of interaction between predictors (265). Of note, this is a gross simplification of the prognostic model development process, and there is an extensive literature focusing on the technical and practical aspects of the optimal model development and validation process (266). As previously mentioned, prognostic models in prostate cancer typically include the Gleason score, PSA and cT. Although age at diagnosis is a predictor of death from prostate cancer, of all risk stratification tool we evaluated in Study II, age was included only in the CAPRA score.

The best-performing tools we identified in Study II (MSKCC nomogram, CAPRA score, CPG risk groups) improve prediction of death from prostate cancer when compared to the D'Amico and D'Amico-derived risk grouping systems. However, these tools are still sub-optimal. For the MSKCC nomogram, at the time Study II was conducted, age was not included in the model and primary and secondary Gleason pattern were dichotomized. Both the CAPRA score and the CPG risk grouping system categorize PSA. The CAPRA score also dichotomizes primary and secondary Gleason pattern as well as age at diagnosis, which is reduced to the categories <50 and ≥50 years of age. Given the plethora of available prognostic models in prostate cancer (68, 74, 75, 77), development of novel models based on the standard clinical variables is difficult to justify (263). However, using one, or several, standard methods to update (267, 268) the existing best performing models could get them closer to the "gold standard" ideal. Until then, consistent adoption of one or a few of the best performing tools in both clinical practice and research will allow for more personalized treatment decisions, facilitate the introduction of novel biomarkers and improve comparability across studies.

Of note, the pretreatment MSKCC nomogram was updated in 2020 to include age, which is modelled as a continuous variable, and to replace the dichotomized Gleason patterns with the full range of the GGs (269). These changes seem to have had a minimal impact on the MSKCC internally validated C-index (0.79 in 2019 vs. 0.79 in 2020). We cannot, however, exclude the possibility that even minimal improvements in discrimination could translate into a more correct risk stratification and more appropriate management for some patients. Thus, it remains to be seen how this updated model performs in predicting individual probabilities of death from prostate cancer, in the competing risks setting.

5.3 Improvement of the prognostic model performance

In the preceding text I only briefly hinted at the technical and practical complexities of prognostic model development (266). The performance of prognostic models can be improved through, for example, careful selection of predictors, choosing the correct functional form of continuous predictors, reduction of predictor misclassification, the inclusion of interactions between predictors as well as through model-updating methods (265, 266). Discussing all of these methods is beyond the scope of this thesis. Here I will focus only on improvement through reduction of predictor misclassification and on one of the model-updating methods: the addition of new prognostic markers to an existing model. These issues were addressed in Study III and IV.

5.3.1 Reliability and measurement heterogeneity

Reliability of the predictor variables and measurement heterogeneity are important issues in prognostic model development (270). Reliability refers to the repeatability (i.e., intra-observer agreement) and/or reproducibility (i.e., inter-observer agreement). Measurement heterogeneity refers to differences in the procedure and/or instruments used to measure the predictors. It has been shown that prognostic models including unreliable/misclassified predictors perform suboptimally on internal (271, 272) and, especially, external validation (273-275). In prostate cancer, all standard variables used for developing prognostic models are known to be unreliably measured and/or are subject to measurement heterogeneity. As described in the Introduction of this thesis, PSA levels are subject to large analytical and biological variation (84), cT is based on subjective digital rectal examination (86), and the Gleason score is notorious for its low inter-observer agreement, which has been only minimally improved by the ISUP 2005 and 2014 revisions (Table 1.3).

While the misclassification of PSA and cT was not directly addressed in this thesis, in Study II, we imputed cT2 and cT3 sub-stage to reduce misclassification of men in risk groups and to improve comparability across the different risk stratification tools. In Study III we evaluated the reliability (i.e., intra-method agreement) and the measurement heterogeneity (i.e., inter-method agreement) of the ISUP 2014 Gleason grading system, including the Gleason score and the GGs as well as many other histopathological features (Table 3.2). Furthermore, in Study IV we evaluated the potential improvement in prostate cancer prognostication due to the revisions of the Gleason grading system.

5.3.1.1 The ISUP 2014 Gleason Grading system

The main aim of Study III was to demonstrate interchangeability of light and virtual microscopy for the purpose of using virtual microscopy for the central re-review of ProMort I and II. To achieve this aim we evaluated the repeatability and reproducibility of different histopathological features on both light and virtual microscopy.

Overall, we found better repeatability and reproducibility of the ISUP 2014 Gleason grading system compared to previous studies which evaluated the pre-2005 Gleason score (103-107, 143), ISUP 2005 Gleason score (109-115, 122, 142), ISUP 2014 Gleason score (116, 117) and GGs (116, 117). The agreement between the pathologists in Study III was better than the agreement in the studies evaluating the pre-2005 and ISUP 2005 Gleason score regardless of the experience of the pathologists (general or uropathologists). For the ISUP 2014 Gleason score and GGs our pathologist agreed better than the general pathologists in previous studies (116, 117). Only one previous study has evaluated the agreement between uropathologists for the GGs and their results were similar to ours (118). Of note, the sample size in Study III was relatively small (n=60) and we evaluated the agreement only between two uropathologists who have been working together for more than 7 years. Our findings on agreement, thus, might not reflect the agreement between unrelated uropathologists, and they quite likely do not reflect the agreement between general pathologists.

The interpretation of these results is not straightforward. It is possible that the changes introduced by the ISUP 2014 revision have improved the inter-observer agreement for the Gleason score. Improved agreement would indicate an improved reliability of the Gleason score, or GGs, and potentially improved prediction of death from prostate cancer. Indeed, in Study IV, we found that the ISUP 2014 Gleason score discriminates death from prostate cancer better than the pre-2005 Gleason score. However, it is unclear if the improved discrimination is explained by increased reliability of the ISUP 2014 Gleason score or by better classification of tumor aggressiveness. Our results indicate that this improvement could be due to classifying all cribriform patterns, rather than poorly formed glands, as Gleason pattern 4. As cribriform pattern seems to have an unfavorable biologic behavior (see also section 1.3.4.2), interpreting all cribriform patterns as Gleason pattern 4 could lead to better classification of men with aggressive disease.

Furthermore, given the similar agreement within and between light and virtual microscopy in Study III, we also confirmed the interchangeability of virtual and light microscopy for the ISUP 2014 Gleason grading system (109, 112, 113, 199). These results indicate no measurement heterogeneity when the histopathological review is performed on light vs. virtual microscopy.

5.3.1.2 Other histopathological features

Compared to previous studies, in Study III we found better repeatability and reproducibility for cribriform pattern (276, 277), poorly formed glands (276, 277) and comedonecrosis (278), but not for the percentage of Gleason pattern 4 (277). For some of the features which are rare, or are rarely reported (e.g., intraductal cancer, inflammation, HGPIN and PAH), we did not find good repeatability and/or reproducibility. Unreliable measurements for these features, and for the percentage of Gleason pattern 4, caution against using them in prognostic models until repeatability and reproducibility are improved. For all these features, however, our findings were similar on light and virtual microscopy, which indicates interchangeability and little to no measurement heterogeneity of light vs. virtual microscopy.

5.3.2 Model extension with a novel prognostic marker

Finally, in Study IV, we evaluated some of the previously proposed or established histopathological markers as well as other histopathological features as predictors of death from prostate cancer. Critical evaluation of novel prognostic markers can be summarized into several phases, including: proof of concept, prospective validation, evaluation of incremental value, and assessment of clinical utility and cost effectiveness (279). In study IV, we were interested in validating existing and identifying novel histopathological markers. We also evaluated whether these markers could improve a "gold standard" prediction model.

While almost all evaluated histopathological features in Study IV predicted death from prostate cancer in univariable analyses, only the number of cores with ≥50% cancer involvement, comedonecrosis and HGPIN remained predictive independently of GGs. When these markers were added to a "gold standard" prognostic model, only comedonecrosis and HGPIN remained independent predictors of death from prostate cancer, although with minimal impact on discrimination (C-index: 0.86 vs 0.85). Notably, we did not use the best-performing prognostic model identified in Study II as the "gold standard" model. Instead we used a somewhat optimized version including the whole range of the GGs, continuous age, PSA modeled using restricted cubic splines and cT. This model is actually quite similar to the update of the MSKCC nomogram in 2020 (269). In addition, by adding primary treatment in all the models, we also accounted for the potential variability in the outcome due to primary treatment. Our study is the first to show that the presence of comedonecrosis is prognostic even after adjustment for the GGs and other standard clinicopathological factors. While interesting, this finding may have limited clinical application, as all men with comedonecrosis are assigned GG5 and are thus typically recommended treatment. Finally, the inverse association of HGPIN with death from prostate cancer is opposite to the only study describing HGPIN as a predictor of lethal prostate cancer (280), and this result should be interpreted with care until confirmed or disputed by additional research.

Given that ProMort II was sampled using a modified nested case-control sampling scheme, we evaluated the improvement in prediction only in terms of discrimination. By applying the modified inverse probability weights in the analysis of ProMort II we could, as discussed above, try to estimate the individual probabilities of dying from prostate cancer. This could help us understand if the identified histopathological markers actually improve risk stratification compared to the "gold standard" model, regardless of the minimal impact on the discrimination.

5.3.2.1 Case level vs. core level

Prostate cancer diagnosis is based on needle biopsy findings, where several biopsy cores are sampled per patient. However, this core-level information is usually not accounted for in the prognostic models. Instead, only case-level summaries are used. For example, the highest Gleason score on a single core or the global Gleason score are typically used to assign a case-level Gleason score. While there seems to be no difference between the highest and the global Gleason score or GGs in predicting BCR (281), the global GGs seem to have slightly higher agreement with the GGs on radical prostatectomy (282). However, neither of these summary case-level measurements truly takes into account the different GGs in all the sampled cores. The recommendation to record the percentage of Gleason pattern 4/5 was an attempt to quantify the extent of high grade cancer in the biopsy core(s). There are several proposed ways of including this information in the prognostic models and the overall percentage of Gleason pattern 4 seems to outperform the highest percentage of Gleason pattern 4 (174, 180). How the information on the Gleason score/GGs and the percentage of Gleason pattern 4/5 is to be best combined is also not clear. Furthermore, there seems to be no consensus on the best way to quantify tumor extent. Many different, highly correlated measures have been proposed (see also the section 1.3.4.1), with no clear recommendations on the "best" measure. Finally, for binary histopathological features, case-level summaries typically refer to presence or absence in any of the sampled cores. For example, a man with perineural invasion in one core is treated the same as a man with perineural invasion in all of the cores. By using a single-summary case-level measure, we are losing a lot of potentially prognostic information. One of the next steps in improving prognostic models in prostate cancer should thus be finding optimal ways of using the core-level information for each man.

6 Conclusions

Study I. The nested case-control study design can be used to obtain unbiased estimates of the relative risks of dying from prostate cancer. However, in the competing risks setting, nested case-control studies with augmented competing-risks cases and controls provide more valid absolute risk estimates. Thus, without additional extensions to the design, nested-case control studies are not suitable for the development of models predicting death from prostate cancer in the competing risks setting.

Study II. The MSKCC nomogram, CAPRA score and CPG system discriminate death from prostate cancer better than the D'Amico and D'Amico-based risk grouping systems. Using these tools leads to finer individual risk prediction and using them in clinical practice could improve treatment decisions. Furthermore, these tools should be used to benchmark novel biomarkers and using them consistently in research could improve comparability across studies.

Study III. Our findings indicate that light microscopy and our internally developed virtual microscopy system can be used interchangeably. We found good repeatability and reproducibility for key histopathological features, such as the ISUP 2014 Gleason score, GGs, perineural invasion, and cribriform pattern. The repeatability and/or reproducibility for some of the rare, or rarely reported, features, and for the percentage of Gleason pattern 4, was poor and should be improved before they are used in clinical practice.

Study IV. The ISUP 2014 Gleason score discriminates death from prostate cancer better than the pre-2005 Gleason score. This improvement is likely due to classifying all cribriform patterns, rather than poorly formed glands, as Gleason pattern 4.

Comedonecrosis and HGPIN predict death from prostate cancer independently of the GGs, age, PSA and cT at diagnosis, however, their impact on model discrimination is minimal. Future studies should confirm our results and evaluate if adding comedonecrosis and/or HGPIN to the "gold standard" model improves risk stratification.

7 Future directions

Several important issues regarding the development of prognostic models, such as sampling designs and handling of competing events, have been discussed in the Methods and Discussion sections and will not be further expanded here. In this section I will focus on several topics which have not been covered in this thesis, but should, nevertheless, be thought of when developing prognostic models. These topics deal with the improvement of current practices of prognostic model development for prostate cancer, and in general.

7.1 Dynamic prediction models

The prognostic models discussed in this thesis are built using baseline features and are sometimes referred to as static models. Static models do not account for temporal changes in the population or in the features used for model development and they could become outdated over time. Furthermore, the implementation of prognostic models into clinical practice will change clinical decision-making and, ultimately, the outcome of interest. Consequently, risk predictions obtained using these models may be inaccurate which, in turn, leads to inappropriate treatment decision-making. Prognostic models, therefore, tend to become "victims of their own success" (283, 284). Models which account for the changes occurring over time are called dynamic models. In the current literature, dynamic prognostic modelling refers to either continuous model updating or to incorporation of time-dependent covariates.

The performance of static prognostic models tends to deteriorate over time. This phenomenon is also known as calibration drift (285, 286). Calibration drift is a consequence of differences between the population used for model development and the population to which the model is applied. These differences may refer to shifts in the outcome rate, patient case mix, or associations between predictors and outcomes (287, 288). In this setting, dynamic modelling refers to the application of discrete or continuous model updating methods (289, 290). Discrete model updating methods use new data over time and apply one, or several, standard model updating methods, such as simple intercept correction, adjustment of coefficients, or inclusion of novel predictors in the model (287, 289, 290). Continuous, or Bayesian, model updating methods combine the information obtained from the past data with the new data to obtain updated estimates (289, 290). Such updating methods seem to have little impact on discrimination of the model. However, they often lead to improved calibration (290).

In time-to-event data, dynamic prediction modelling refers to the incorporation of time-dependent covariate information into the prognostic model (283). The time-dependent covariate information refers to the longitudinal covariate data collected during treatment or follow-up which is often stored in electronic health records. The two most common approaches to dynamic prediction modelling are joint modelling and landmark analysis (283, 291), but other methods have also been described (292). Joint models simultaneously estimate the model for the longitudinal process and the model for the time-to-event data, while the landmark analysis consists of a series of Cox proportional hazards models estimated at predefined landmark times. Furthermore, several different methods have been proposed to extend dynamic prediction modelling to the competing risks setting, such as extension of the dynamic landmark models (293), combination of the pseudo-observations with the landmark analysis (294) and extension of the landmark models to the Fine and Gray model, i.e., the landmark subdistribution hazards model and supermodel (295).

Although rarely applied in practice, probably due to a lack of access to the longitudinal data, no available guidelines or implementation difficulties, dynamic prognostic models seem to improve the accuracy of the predicted individual risks (296, 297) and they should be implemented in future prognostic models (291).

7.1.1 Multiparametric magnetic resonance imaging guided biopsy

A very timely example of a potential cause of calibration drift is the implementation of multiparametric magnetic resonance imaging (mpMRI) in clinical practice. mpMRI before biopsy as a triage test has been shown to reduce the number of unnecessary biopsies (298) and it seems to be a cost-effective strategy for diagnosing clinically significant prostate cancer in biopsy-naïve men (299). mpMRI-guided biopsy procedure outperforms systematic biopsy procedure in the detection of significant cancer in the repeat-biopsy setting, while in biopsy-naïve men, a combination of mpMRI-guided biopsy and systematic biopsy performs the best (300, 301). These findings have already led to changes in prostate cancer guidelines. For example, in the 2019 edition of the EAU guidelines for prostate cancer, mpMRI is recommended prior to biopsy in patients with suspected clinically significant prostate cancer, both targeted and systematic biopsy are recommended for biopsy-naïve men, and for men with a prior negative biopsy only targeted biopsy is recommended, albeit weakly (50). Although mpMRI has moderate inter-observer reproducibility and the optimal number of targeted cores per mpMRI ROI is still not determined (302), once these are optimized, it is likely that mpMRI-guided biopsy will completely replace systematic biopsies.

It is expected that the implementation of mpMRI-guided biopsies in clinical practice will lead to calibration drift of prognostic models in prostate cancer. Whether the performance of these models can be improved by model updating methods, such as including mpMRI-related information, or by accounting for the potential differences in biopsy cores sampled using mpMRI-guided biopsy vs. systematic biopsy procedure, remains to be seen.

7.2 The role of treatment in clinical prediction models

To guide treatment decision-making in men diagnosed with prostate cancer, individual probabilities of dying from prostate cancer predicted using prognostic models should reflect the probability of dying from prostate cancer in the absence of treatment. Most prognostic models for prostate cancer have been developed using selected populations of radically treated men or using mixed populations of treated and untreated men. Treatment lowers the risk of dying from prostate cancer, yet treatment is usually ignored when prognostic models are developed. Not accounting for treatment when developing a prognostic model leads to underestimation of the probability of the outcome in untreated men (303), and to a biased model performance when validated in differently treated populations (304).

When developing a prognostic model, treatment can be modeled either as a time-invariant variable or as a time-dependent variable. Groenwold et al. compared several methods of accounting for time-invariant treatment when developing a prognostic model for a binary outcome (303). They found that ignoring an effective treatment leads to incorrect predicted probabilities of the outcome, that restricting analysis to untreated individuals is suitable only when treatment allocation is random, and that, when treatment allocation is not random, including treatment as a covariate in the model results in better predictive performance compared to other methods (303). Sperrin et al. extended this work to time-dependent treatment (305). They proposed a counterfactual framework for the development of prognostic models and they showed that using marginal structural models resulted in unbiased predicted probabilities of a binary outcome (305). In a recent study, Pajouheshnia et al. evaluated seven strategies of accounting for time-dependent treatment when developing a prognostic model for time-to-event outcomes (306). They compared models where the treatment was ignored to models developed by excluding treated patients, censoring treated patients at the time of treatment, using inverse probability of treatment weighting, modelling treatment as a binary covariate, modelling treatment as a time-varying covariate and, finally, using marginal structural models with time-varying inverse probability of treatment weights. They confirmed that ignoring the treatment when developing a prognostic model is theoretically inferior. However, when compared to other methods, ignoring the treatment and modelling time-dependent treatment as a binary covariate led to only a small overestimation of the predicted probabilities of the outcome and model performance varied minimally when different approaches were used (306).

Van Geloven et al. proposed a somewhat different approach (307), which was inspired by the European Medicines Agency framework for dealing with additional treatments started after baseline, and other post-baseline but pre-outcome events, in clinical trials (308). In this approach, the choice of strategy for accounting for time-dependent treatment should be based on the question the researcher wants to address using the prediction model (307). An overview of the four proposed strategies is presented in Table 6.1. Depending on the strategy, risk predictions may be very different. Thus, the question of interest and choice of the strategy should ideally be predefined. For example, pretreatment risk stratification tools in prostate cancer are used to guide treatment decision-making for newly diagnosed men. Thus clinicians are interested in the risk of dying from prostate cancer in the absence of treatment. According to the proposed framework, prognostic models addressing this question should be developed using the hypothetical strategy.

Table 6.1. Overview of four strategies for dealing with treatment initiation after baseline in prognostic models

Strategy	Estimand	Example	Estimators	Key assumptions
Ignore treatment	Risk of the event, regardless of treatment	Risk of cardiovascular events where some patients will initiate statins according to routine-care prescriptions	Survival model for T, do not censor at V	Treatment assignment policy in application setting similar to development data
Composite	Risk of the event or treatment initiation	Risk of a composite of cardiovascular death, myocardial infarction and treatment with PCI or CABG	Survival model for min(T, V)	Treatment assignment policy in application setting similar to development data
While untreated	Risk of the event occurring before treatment is started	Risk of dying while on the waiting list for a liver transplant	Competing risks methods	Treatment assignment policy in application setting similar to development data
Hypothetical	Risk of the event if treatment was never started	Risk of a natural pregnancy without IVF treatment	Survival model for T, censor at V or include treatment as time-dependent covariate in the model and set to 0 when predicting	Exchangeability, consistency and positivity

Abbreviations: Estimand, the target quantity that we aim to estimate; T, time to event of interest; V, time to start of treatment; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting; IVF, in vitro fertilization

Reprinted from Nan van Geloven et al. Prediction meets causal inference: the role of treatment in clinical prediction models. Eur J Epidemiol. 2020;35:619–630, licensed under Creative Commons CC BY. © 2020, Springer Nature

7.3 Clinical utility

Although there is a plethora of risk stratification tools in prostate cancer, the vast majority of them is not used by clinicians. To be implemented into standard clinical practice, it should be clear how these tools are intended to be used to guide treatment decision. However, risk stratification tools, or prognostic models in general, do not recommend decisions to clinicians. Instead, clinicians are provided with predicted probabilities of the outcome of interest without being told what to do with that information (309). It is also often not clear if treatment decisions based on the prognostic model will improve patient outcome(s) compared to treatment decisions based on the current standard of care. The gold standard for evaluating clinical utility is a randomized clinical trial where prediction-based decision-making is compared to standard of care decision-making. The implementation of such a trial is, however, both practically and ethically challenging. Sachs at al. proposed to use observational data to optimize a prediction-based decision rule and to evaluate its clinical utility by emulating a randomized clinical trial (310). The proposed framework consists of three separate steps:

1. Development of a prognostic model,

2. Development and optimization of a proposed decision rule based on the prognostic model, and
3. Evaluation of the clinical utility of the proposed decision rule.

A step-by-step example of the proposed framework was made available in the appendix to the paper (310). These studies could then be used to motivate better-informed confirmatory randomized clinical trials.

7.4 Life expectancy

Given the risk of overtreatment, evaluation of life expectancy can help balance the potential for life gained against the potential for harm caused by treatment and is currently recommended for prostate cancer clinical decision-making in several guidelines (35, 36, 50). However, there are no clear recommendations regarding the best method to predict life expectancy and clinicians seem to either not consider it when making treatment decision, or are poor judges of it, and are prone to both under- and over-estimation, regardless of their experience (311-313). Despite the fact that there are several available tools and online calculators predicting life expectancy (312), in practice, simple age-adjusted life tables seem to still be the most commonly recommended (50) and used by clinicians. Predicting life expectancy using age alone might be satisfactory only for men without any additional comorbidities. However, age and comorbidity are independent predictors of other-cause mortality in men with prostate cancer (314, 315) and they should be both considered when making treatment decision (316, 317). Adding additional predictors of other-cause mortality would lead to more individual life expectancy predictions, improved decision-making and potentially decreased overtreatment. Given that most of the current tools predicting death from other causes in prostate cancer patients seem to either be inappropriate for clinical use, or to provide questionable estimates (318), developing and validating a novel life expectancy calculator (using population-based samples of treated and untreated men (316)) in addition to a novel dynamic model predicting death from prostate cancer in the competing risks setting should be the next step toward informed treatment decision-making.

8 Acknowledgements

I would like to express my sincerest gratitude to all those without whom this PhD would not have been possible. Thank you.

To my supervisors:

- I will forever be grateful to Lorenzo for changing the course of my (professional) life. Your supervision, both here and in Italy, has been invaluable. Your knowledge is something I aspire to.
- Olof, for giving me this opportunity, and for finding the time in your hectic schedule to be there when it matters.
- Andreas, for being just the kind of supervisor I needed you to be. Thank you for making this PhD run smoothly, for being practical, decisive, honest, and always present. I never felt alone on this journey.

To all my KEP hikers:

- Andrei, for always taking the time, for literally drawing things for me when words would not suffice, for listening and helping and being a true friend.
- Peter, for selflessly sharing your time and knowledge, for all the hiking, hammocking and the fanciest fika Nackareservatet has ever seen.
- Marios, for knowing everything, every single thing.
- Huiling, for being so enthusiastic in all you do, so open, down-to-earth and so informed.
- Kelsi, for being the most capable and proactive person ever.

I have missed you all in these last few months, but this thesis was much easier to write without being distracted by your greatness.

To the Cancer epi group, and especially Karin Ekström Smedby, for taking me in and making me feel like a part of your group.

To Clinical Epidemiology Division, for being my home for the past four years.

To all the ProMort collaborators, for being an invaluable part of this project. Your work has laid the foundation for this thesis.

The collection of data in the NPCR of Sweden was made possible by the continuous work of the NPCR steering group: Pär Stattin (chairman), Anders Widmark, Camilla Thellenberg, Ove Andrén, Ann-Sofi Fransson, Magnus Törnblom, Stefan Carlsson, Marie Hjälm-Eriksson, David Robinson, Mats Andén, Jonas Hugosson, Ingela Franck Lissbrant, Maria Nyberg, Ola Bratt, Lars Egevad, Olof Ståhl, Calle Walller, Olof Akre, Per Fransson, Eva Johansson, Fredrik Sandin, and Karin Hellström.

A special thanks to Dr. Silvia Johansson for providing us with the data necessary to impute cT sub-stage in Study II.

To my friends in Sweden, for making Sweden feel like home and for showing me that there is more to life than just work. This year would have been much more challenging without you. Thank you for being there.

To my family and my extended family (yes, Olga, I mean you) back home, I am who I am because of you.

9 References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68(6):394-424.
2. National Board of Health and Welfare, Centre for Epidemiology. The Swedish Cancer Register. Stockholm. (https://www.socialstyrelsen.se/en/statistics-and-data/registers/register-information/swedish-cancer-register/). (Accessed 17.06.2020.).
3. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer* 2018;103:356-87.
4. Kvale R, Auvinen A, Adami HO, et al. Interpreting trends in prostate cancer incidence and mortality in the five Nordic countries. *J Natl Cancer Inst* 2007;99(24):1881-7.
5. Fenton JJ, Weyrich MS, Durbin S, et al. Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2018;319(18):1914-31.
6. Wolf AM, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin* 2010;60(2):70-98.
7. U. S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2008;149(3):185-91.
8. Moyer VA, Force USPST. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012;157(2):120-34.
9. U. S. Preventive Services Task Force, Grossman DC, Curry SJ, et al. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* 2018;319(18):1901-13.
10. Jemal A, Fedewa SA, Ma J, et al. Prostate Cancer Incidence and PSA Testing Patterns in Relation to USPSTF Screening Recommendations. *JAMA* 2015;314(19):2054-61.
11. Sammon JD, Abdollah F, Choueiri TK, et al. Prostate-Specific Antigen Screening After 2012 US Preventive Services Task Force Recommendations. *JAMA* 2015;314(19):2077-9.
12. Negoita S, Feuer EJ, Mariotto A, et al. Annual Report to the Nation on the Status of Cancer, part II: Recent changes in prostate cancer trends and disease characteristics. *Cancer* 2018;124(13):2801-14.
13. National Board of Health and Welfare, Centre for Epidemiology. The Cause of Death Register. (https://www.socialstyrelsen.se/statistik-och-data/register/alla-register/dodsorsaksregistret/). (Accessed 16.06.2020.).
14. Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. *Eur Urol* 2012;61(6):1079-92.
15. Wong MC, Goggins WB, Wang HH, et al. Global Incidence and Mortality for Prostate Cancer: Analysis of Temporal Patterns and Trends in 36 Countries. *Eur Urol* 2016;70(5):862-74.
16. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: a review of current evidence. *JAMA* 2014;311(11):1143-9.
17. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst* 2012;104(2):125-32.
18. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 2014;384(9959):2027-35.

19. Pinsky PF, Miller E, Prorok P, et al. Extended follow-up for prostate cancer incidence and mortality among participants in the Prostate, Lung, Colorectal and Ovarian randomized cancer screening trial. *BJU Int* 2019;123(5):854-60.

20. Osses DF, Remmers S, Schroder FH, et al. Results of Prostate Cancer Screening in a Unique Cohort at 19yr of Follow-up. *Eur Urol* 2019;75(3):374-7.

21. Pinsky PF, Blacka A, Kramer BS, et al. Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *Clin Trials* 2010;7(4):303-11.

22. de Koning HJ, Gulati R, Moss SM, et al. The efficacy of prostate-specific antigen screening: Impact of key components in the ERSPC and PLCO trials. *Cancer* 2018;124(6):1197-206.

23. Tsodikov A, Gulati R, Heijnsdijk EAM, et al. Reconciling the Effects of Screening on Prostate Cancer Mortality in the ERSPC and PLCO Trials. *Ann Intern Med* 2017;167(7):449-55.

24. Arnsrud Godtman R, Holmberg E, Lilja H, et al. Opportunistic testing versus organized prostate-specific antigen screening: outcome after 18 years in the Goteborg randomized population-based prostate cancer screening trial. *Eur Urol* 2015;68(3):354-60.

25. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. *Lancet Oncol* 2010;11(8):725-32.

26. Lujan M, Paez A, Angulo JC, et al. Prostate cancer incidence and mortality in the Spanish section of the European Randomized Study of Screening for Prostate Cancer (ERSPC). *Prostate Cancer Prostatic Dis* 2014;17(2):187-91.

27. Kilpelainen TP, Pogodin-Hannolainen D, Kemppainen K, et al. Estimate of Opportunistic Prostate Specific Antigen Testing in the Finnish Randomized Study of Screening for Prostate Cancer. *J Urol* 2017;198(1):50-7.

28. Kilpelainen TP, Tammela TL, Malila N, et al. Prostate cancer mortality in the Finnish randomized screening trial. *J Natl Cancer Inst* 2013;105(10):719-25.

29. Bokhorst LP, Bangma CH, van Leenders GJ, et al. Prostate-specific antigen-based prostate cancer screening: reduction of prostate cancer mortality after correction for nonattendance and contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. *Eur Urol* 2014;65(2):329-36.

30. Roobol MJ, Kranse R, Bangma CH, et al. Screening for prostate cancer: results of the Rotterdam section of the European randomized study of screening for prostate cancer. *Eur Urol* 2013;64(4):530-9.

31. Martin RM, Donovan JL, Turner EL, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial. *JAMA* 2018;319(9):883-95.

32. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol* 2014;65(6):1046-55.

33. Graham J, Kirkbride P, Cann K, et al. Prostate cancer: summary of updated NICE guidance. *BMJ* 2014;348:f7524.

34. Lukka H, Warde P, Pickles T, et al. Controversies in prostate cancer radiotherapy: consensus development. *Can J Urol* 2001;8(4):1314-22.

35. Sanda MG, Cadeddu JA, Kirkby E, et al. Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. Part I: Risk Stratification, Shared Decision Making, and Care Options. *J Urol* 2018;199(3):683-90.

36. Mohler JL, Antonarakis ES, Armstrong AJ, et al. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2019;17(5):479-505.

37. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* 2017;71(4):618-29.
38. Cooperberg MR, Broering JM, Carroll PR. Time trends and local variation in primary treatment of localized prostate cancer. *J Clin Oncol* 2010;28(7):1117-23.
39. Cooperberg MR, Carroll PR. Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. *JAMA* 2015;314(1):80-2.
40. Donovan JL, Hamdy FC, Lane JA, et al. Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *N Engl J Med* 2016;375(15):1425-37.
41. Wei JT, Dunn RL, Sandler HM, et al. Comprehensive comparison of health-related quality of life after contemporary therapies for localized prostate cancer. *J Clin Oncol* 2002;20(2):557-66.
42. Korfage IJ, Essink-Bot ML, Borsboom GJ, et al. Five-year follow-up of health-related quality of life after primary treatment of localized prostate cancer. *Int J Cancer* 2005;116(2):291-6.
43. Lardas M, Liew M, van den Bergh RC, et al. Quality of Life Outcomes after Primary Treatment for Clinically Localised Prostate Cancer: A Systematic Review. *Eur Urol* 2017;72(6):869-85.
44. Mahal BA, Butler S, Franco I, et al. Use of Active Surveillance or Watchful Waiting for Low-Risk Prostate Cancer and Management Trends Across Risk Groups in the United States, 2010-2015. *JAMA* 2019.
45. Loeb S, Berglund A, Stattin P. Population based study of use and determinants of active surveillance and watchful waiting for low and intermediate risk prostate cancer. *J Urol* 2013;190(5):1742-9.
46. Hamdy FC, Donovan JL, Lane JA, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *N Engl J Med* 2016;375(15):1415-24.
47. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical Prostatectomy or Watchful Waiting in Prostate Cancer - 29-Year Follow-up. *N Engl J Med* 2018;379(24):2319-29.
48. Wilt TJ, Jones KM, Barry MJ, et al. Follow-up of Prostatectomy versus Observation for Early Prostate Cancer. *N Engl J Med* 2017;377(2):132-42.
49. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998;280(11):969-74.
50. Mottet N, Bellmunt J, Briers E, et al. *EAU – ESTRO – ESUR – SIOG Guidelines on Prostate Cancer*. presented at the EAU Annual Congress Amsterdam 2020 ed. Arnhem, The Netherlands: EAU Guidelines Office; 2020.
51. Epstein JI, Walsh PC, Carmichael M, et al. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. *JAMA* 1994;271(5):368-74.
52. Freedland SJ, Terris MK, Csathy GS, et al. Preoperative model for predicting prostate specific antigen recurrence after radical prostatectomy using percent of biopsy tissue with cancer, biopsy Gleason grade and serum prostate specific antigen. *J Urol* 2004;171(6 Pt 1):2215-20.
53. Gnanapragasam VJ, Lophatananon A, Wright KA, et al. Improving Clinical Risk Stratification at Diagnosis in Primary Prostate Cancer: A Prognostic Modelling Study. *PLoS Med* 2016;13(8):e1002063.
54. Zumsteg ZS, Chen Z, Howard LE, et al. Number of Unfavorable Intermediate-Risk Factors Predicts Pathologic Upstaging and Prostate Cancer-Specific Mortality Following Radical Prostatectomy: Results From the SEARCH Database. *Prostate* 2017;77(2):154-63.

55. Zumsteg ZS, Spratt DE, Pei I, et al. A new risk classification system for therapeutic decision making with intermediate-risk prostate cancer patients undergoing dose-escalated external-beam radiation therapy. *Eur Urol* 2013;64(6):895-902.

56. Partin AW, Kattan MW, Subong EN, et al. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *JAMA* 1997;277(18):1445-51.

57. Partin AW, Mangold LA, Lamm DM, et al. Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium. *Urology* 2001;58(6):843-8.

58. Eifler JB, Feng Z, Lin BM, et al. An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011. *BJU Int* 2013;111(1):22-9.

59. Tosoian JJ, Chappidi M, Feng Z, et al. Prediction of pathological stage based on clinical stage, serum prostate-specific antigen, and biopsy Gleason score: Partin Tables in the contemporary era. *BJU Int* 2017;119(5):676-83.

60. Cooperberg MR, Pasta DJ, Elkin EP, et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *J Urol* 2005;173(6):1938-42.

61. Kattan MW, Eastham JA, Stapleton AM, et al. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998;90(10):766-71.

62. Kattan MW, Cuzick J, Fisher G, et al. Nomogram incorporating PSA level to predict cancer-specific survival for men with clinically localized prostate cancer managed without curative intent. *Cancer* 2008;112(1):69-74.

63. Stephenson AJ, Scardino PT, Eastham JA, et al. Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Natl Cancer Inst* 2006;98(10):715-7.

64. Stephenson AJ, Kattan MW, Eastham JA, et al. Prostate cancer-specific mortality after radical prostatectomy for patients treated in the prostate-specific antigen era. *J Clin Oncol* 2009;27(26):4300-5.

65. Gamito EJ, Crawford ED. Artificial neural networks for predictive modeling in prostate cancer. *Curr Oncol Rep* 2004;6(3):216-21.

66. Roffman DA, Hart GR, Leapman MS, et al. Development and Validation of a Multiparameterized Artificial Neural Network for Prostate Cancer Risk Prediction and Stratification. *JCO Clin Cancer Inform* 2018;2:1-10.

67. Tewari A, Issa M, El-Galley R, et al. Genetic adaptive neural network to predict biochemical failure after radical prostatectomy: a multi-institutional study. *Mol Urol* 2001;5(4):163-9.

68. Shariat SF, Karakiewicz PI, Suardi N, et al. Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature. *Clin Cancer Res* 2008;14(14):4400-7.

69. Shariat SF, Capitanio U, Jeldres C, et al. Can nomograms be superior to other prediction tools? *BJU Int* 2009;103(4):492-5; discussion 5-7.

70. Chun FK, Karakiewicz PI, Briganti A, et al. A critical appraisal of logistic regression-based nomograms, artificial neural networks, classification and regression-tree models, look-up tables and risk-group stratification models for prostate cancer. *BJU Int* 2007;99(4):794-800.

71. Boorjian SA, Karnes RJ, Rangel LJ, et al. Mayo Clinic validation of the D'amico risk group classification for predicting survival following radical prostatectomy. *J Urol* 2008;179(4):1354-60; discussion 60-1.

72. Cooperberg MR, Broering JM, Carroll PR. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. *J Natl Cancer Inst* 2009;101(12):878-87.

73. Vainshtein JM, Schipper M, Vance S, et al. Limitations of the Cancer of the Prostate Risk Assessment (CAPRA) Prognostic Tool for Prediction of Metastases and Prostate Cancer-specific Mortality in Patients Treated With External Beam Radiation Therapy. *Am J Clin Oncol* 2016;39(2):173-80.

74. Campbell JM, Raymond E, O'Callaghan ME, et al. Optimum Tools for Predicting Clinical Outcomes in Prostate Cancer Patients Undergoing Radical Prostatectomy: A Systematic Review of Prognostic Accuracy and Validity. *Clin Genitourin Cancer* 2017;15(5):e827-e34.

75. Capitanio U, Briganti A, Gallina A, et al. Predictive models before and after radical prostatectomy. *Prostate* 2010;70(12):1371-8.

76. Chun FK, Karakiewicz PI, Briganti A, et al. Prostate cancer nomograms: an update. *Eur Urol* 2006;50(5):914-26; discussion 26.

77. Lughezzani G, Briganti A, Karakiewicz PI, et al. Predictive and prognostic models in radical prostatectomy candidates: a critical analysis of the literature. *Eur Urol* 2010;58(5):687-700.

78. Martin NE, Mucci LA, Loda M, et al. Prognostic determinants in prostate cancer. *Cancer J* 2011;17(6):429-37.

79. Lamy PJ, Allory Y, Gauchez AS, et al. Prognostic Biomarkers Used for Localised Prostate Cancer Management: A Systematic Review. *Eur Urol Focus* 2018;4(6):790-803.

80. Saini S. PSA and beyond: alternative prostate cancer biomarkers. *Cell Oncol (Dordr)* 2016;39(2):97-106.

81. Bjartell A, Montironi R, Berney DM, et al. Tumour markers in prostate cancer II: diagnostic and prognostic cellular biomarkers. *Acta Oncol* 2011;50 Suppl 1:76-84.

82. Lam TBL, MacLennan S, Willemse PM, et al. EAU-EANM-ESTRO-ESUR-SIOG Prostate Cancer Guideline Panel Consensus Statements for Deferred Treatment with Curative Intent for Localised Prostate Cancer from an International Collaborative Study (DETECTIVE Study). *Eur Urol* 2019;76(6):790-813.

83. Memorial Sloan Kettering Center. Dynamic Prostate Cancer Nomogram: Coefficients. (https://www.mskcc.org/nomograms/prostate/pre_op/coefficients). (Accessed 21.01.2019).

84. Shariat SF, Semjonow A, Lilja H, et al. Tumor markers in prostate cancer I: blood-based markers. *Acta Oncol* 2011;50 Suppl 1:61-75.

85. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* 2004;350(22):2239-46.

86. Paner GP, Stadler WM, Hansel DE, et al. Updates in the Eighth Edition of the Tumor-Node-Metastasis Staging Classification for Urologic Cancers. *Eur Urol* 2018;73(4):560-9.

87. de Rooij M, Hamoen EH, Witjes JA, et al. Accuracy of Magnetic Resonance Imaging for Local Staging of Prostate Cancer: A Diagnostic Meta-analysis. *Eur Urol* 2016;70(2):233-45.

88. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep* 1966;50(3):125-8.

89. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* 1974;111(1):58-64.

90. Mostofi FK. Grading of prostatic carcinoma. *Cancer Chemother Rep* 1975;59(1):111-7.

91. Epstein JI, Allsbrook WC, Jr., Amin MB, et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* 2005;29(9):1228-42.

92. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* 2016;40(2):244-52.

93. Danneman D, Drevin L, Robinson D, et al. Gleason inflation 1998-2011: a registry study of 97,168 men. *BJU Int* 2015;115(2):248-55.

94. Helpap B, Egevad L. The significance of modified Gleason grading of prostatic carcinoma in biopsy and radical prostatectomy specimens. *Virchows Arch* 2006;449(6):622-7.

95. Epstein JI, Amin MB, Reuter VE, et al. Contemporary Gleason Grading of Prostatic Carcinoma: An Update With Discussion on Practical Issues to Implement the 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* 2017;41(4):e1-e7.

96. Pierorazio PM, Walsh PC, Partin AW, et al. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU Int* 2013;111(5):753-60.

97. Humphrey PA, Moch H, Cubilla AL, et al. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours. *Eur Urol* 2016;70(1):106-19.

98. Offermann A, Hupe MC, Sailer V, et al. The new ISUP 2014/WHO 2016 prostate cancer grade group system: first resume 5 years after introduction and systemic review of the literature. *World J Urol* 2019.

99. Billis A, Guimaraes MS, Freitas LL, et al. The impact of the 2005 international society of urological pathology consensus conference on standard Gleason grading of prostatic carcinoma in needle biopsies. *J Urol* 2008;180(2):548-52; discussion 52-3.

100. Uemura H, Hoshino K, Sasaki T, et al. Usefulness of the 2005 International Society of Urologic Pathology Gleason grading system in prostate biopsy and radical prostatectomy specimens. *BJU Int* 2009;103(9):1190-4.

101. Delahunt B, Lamb DS, Srigley JR, et al. Gleason scoring: a comparison of classical and modified (international society of urological pathology) criteria using nadir PSA as a clinical end point. *Pathology* 2010;42(4):339-43.

102. Offermann A, Hohensteiner S, Kuempers C, et al. Prognostic Value of the New Prostate Cancer International Society of Urological Pathology Grade Groups. *Front Med (Lausanne)* 2017;4:157.

103. McLean M, Srigley J, Banerjee D, et al. Interobserver variation in prostate cancer Gleason scoring: are there implications for the design of clinical trials and treatment strategies? *Clin Oncol (R Coll Radiol)* 1997;9(4):222-5.

104. Allsbrook WC, Jr., Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol* 2001;32(1):74-80.

105. Glaessgen A, Hamberg H, Pihl CG, et al. Interobserver reproducibility of percent Gleason grade 4/5 in prostate biopsies. *J Urol* 2004;171(2 Pt 1):664-7.

106. Oyama T, Allsbrook WC, Jr., Kurokawa K, et al. A comparison of interobserver reproducibility of Gleason grading of prostatic carcinoma in Japan and the United States. *Arch Pathol Lab Med* 2005;129(8):1004-10.

107. Allsbrook WC, Jr., Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. *Hum Pathol* 2001;32(1):81-8.

108. Salmo EN. An audit of inter-observer variability in Gleason grading of prostate cancer biopsies: The experience of central pathology review in the North West of England. *Integr Cancer Sci Therap* 2015;2.

109. Rodriguez-Urrego PA, Cronin AM, Al-Ahmadie HA, et al. Interobserver and intraobserver reproducibility in digital and routine microscopic assessment of prostate needle biopsies. *Hum Pathol* 2011;42(1):68-74.

110. Griffiths DF, Melia J, McWilliam LJ, et al. A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility. *Histopathology* 2006;48(6):655-62.

111. Melia J, Moseley R, Ball RY, et al. A UK-based investigation of inter- and intra-observer reproducibility of Gleason grading of prostatic biopsies. *Histopathology* 2006;48(6):644-54.

112. Goodman M, Ward KC, Osunkoya AO, et al. Frequency and determinants of disagreement and error in gleason scores: a population-based study of prostate cancer. *Prostate* 2012;72(13):1389-98.

113. Harnden P, Coleman D, Moss S, et al. Evaluation of the use of digital images for a national prostate core external quality assurance scheme. *Histopathology* 2011;59(4):703-9.

114. Mulay K, Swain M, Jaiman S, et al. Gleason scoring of prostatic carcinoma: impact of a web-based tutorial on inter- and intra-observer variability. *Indian J Pathol Microbiol* 2008;51(1):22-5.

115. Singh RV, Agashe SR, Gosavi AV, et al. Interobserver reproducibility of Gleason grading of prostatic adenocarcinoma among general pathologists. *Indian J Cancer* 2011;48(4):488-95.

116. Ozkan TA, Eruyar AT, Cebeci OO, et al. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol* 2016;50(6):420-4.

117. Al Nemer AM, Elsharkawy T, Elshawarby M, et al. The updated grading system of prostate carcinoma: an inter-observer agreement study among general pathologists in an academic practice. *APMIS* 2017;125(11):957-61.

118. Egevad L, Delahunt B, Berney DM, et al. Utility of Pathology Imagebase for standardisation of prostate cancer grading. *Histopathology* 2018;73(1):8-18.

119. Egevad L. Reproducibility of Gleason grading of prostate cancer can be improved by the use of reference images. *Urology* 2001;57(2):291-5.

120. Mikami Y, Manabe T, Epstein JI, et al. Accuracy of gleason grading by practicing pathologists and the impact of education on improving agreement. *Hum Pathol* 2003;34(7):658-65.

121. Nakai Y, Tanaka N, Shimada K, et al. Review by urological pathologists improves the accuracy of Gleason grading by general pathologists. *BMC Urol* 2015;15:70.

122. Abdollahi A, Sheikhbahaei S, Meysamie A, et al. Inter-observer reproducibility before and after web-based education in the Gleason grading of the prostate adenocarcinoma among the Iranian pathologists. *Acta Med Iran* 2014;52(5):370-4.

123. Egevad L, Algaba F, Berney DM, et al. Interactive digital slides with heat maps: a novel method to improve the reproducibility of Gleason grading. *Virchows Arch* 2011;459(2):175-82.

124. Bondarenko HD, Zanaty M, Harmouch SS, et al. External validation of the novel International Society of Urological Pathology (ISUP) Gleason grading groups in a large contemporary Canadian cohort. *Can Urol Assoc J* 2018.

125. Dell'Oglio P, Karnes RJ, Gandaglia G, et al. The New Prostate Cancer Grading System Does Not Improve Prediction of Clinical Recurrence After Radical Prostatectomy: Results of a Large, Two-Center Validation Study. *Prostate* 2017;77(3):263-73.

126. Epstein JI, Zelefsky MJ, Sjoberg DD, et al. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol* 2016;69(3):428-35.

127. Kirmiz S, Qi J, Babitz SK, et al. Grade Groups Provide Improved Predictions of Pathological and Early Oncologic Outcomes Compared with Gleason Score Risk Groups. *J Urol* 2019;201(2):278-83.

128. Loeb S, Folkvaljon Y, Robinson D, et al. Evaluation of the 2015 Gleason Grade Groups in a Nationwide Population-based Cohort. *Eur Urol* 2016;69(6):1135-41.

129. Mathieu R, Moschini M, Beyer B, et al. Prognostic value of the new Grade Groups in Prostate Cancer: a multi-institutional European validation study. *Prostate Cancer Prostatic Dis* 2017;20(2):197-202.

130. Schulman AA, Howard LE, Tay KJ, et al. Validation of the 2015 prostate cancer grade groups for predicting long-term oncologic outcomes in a shared equal-access health system. *Cancer* 2017;123(21):4122-9.

131. Spratt DE, Cole AI, Palapattu GS, et al. Independent surgical validation of the new prostate cancer grade-grouping system. *BJU Int* 2016;118(5):763-9.

132. Yeong J, Sultana R, Teo J, et al. Gleason grade grouping of prostate cancer is of prognostic value in Asian men. *J Clin Pathol* 2017;70(9):745-53.

133. Beckmann KR, Vincent AD, O'Callaghan ME, et al. Oncological outcomes in an Australian cohort according to the new prostate cancer grading groupings. *BMC Cancer* 2017;17(1):537.

134. Berney DM, Beltran L, Fisher G, et al. Validation of a contemporary prostate cancer grading system using prostate cancer death as outcome. *Br J Cancer* 2016;114(10):1078-83.

135. Chen C, Chen Y, Hu LK, et al. The performance of the new prognostic grade and stage groups in conservatively treated prostate cancer. *Asian J Androl* 2018;20(4):366-71.

136. Delahunt B, Egevad L, Srigley JR, et al. Validation of International Society of Urological Pathology (ISUP) grading for prostatic adenocarcinoma in thin core biopsies using TROG 03.04 'RADAR' trial clinical data. *Pathology* 2015;47(6):520-5.

137. He J, Albertsen PC, Moore D, et al. Validation of a Contemporary Five-tiered Gleason Grade Grouping Using Population-based Data. *Eur Urol* 2017;71(5):760-3.

138. Leapman MS, Cowan JE, Simko J, et al. Application of a Prognostic Gleason Grade Grouping System to Assess Distant Prostate Cancer Outcomes. *Eur Urol* 2017;71(5):750-9.

139. Pompe RS, Davis-Bondarenko H, Zaffuto E, et al. Population-Based Validation of the 2014 ISUP Gleason Grade Groups in Patients Treated With Radical Prostatectomy, Brachytherapy, External Beam Radiation, or no Local Treatment. *Prostate* 2017;77(6):686-93.

140. Spratt DE, Jackson WC, Abugharib A, et al. Independent validation of the prognostic capacity of the ISUP prostate cancer grade grouping system for radiation treated patients with long-term follow-up. *Prostate Cancer Prostatic Dis* 2016;19(3):292-7.

141. De Nunzio C, Pastore AL, Lombardo R, et al. The new Epstein gleason score classification significantly reduces upgrading in prostate cancer patients. *Eur J Surg Oncol* 2018;44(6):835-9.

142. Veloso SG, Lima MF, Salles PG, et al. Interobserver agreement of Gleason score and modified Gleason score in needle biopsy and in surgical specimen of prostate cancer. *Int Braz J Urol* 2007;33(5):639-46; discussion 47-51.

143. Abdollahi A, Meysamie A, Sheikhbahaei S, et al. Inter/intra-observer reproducibility of Gleason scoring in prostate adenocarcinoma in Iranian pathologists. *Urol J* 2012;9(2):486-90.

144. Qureshi A, Lakhtakia R, M ALB, et al. Gleason's Grading of Prostatic Adenocarcinoma: Inter-Observer Variation Among Seven Pathologists at a Tertiary Care Center in Oman. *Asian Pac J Cancer Prev* 2016;17(11):4867-8.

145. Egevad L, Judge M, Delahunt B, et al. Dataset for the reporting of prostate carcinoma in core needle biopsy and transurethral resection and enucleation specimens: recommendations from the International Collaboration on Cancer Reporting (ICCR). *Pathology* 2019;51(1):11-20.

146. Hoogland AM, Kweldam CF, van Leenders GJ. Prognostic histopathological and molecular markers on prostate cancer needle-biopsies: a review. *Biomed Res Int* 2014;2014:341324.

147. Harnden P, Shelley MD, Naylor B, et al. Does the extent of carcinoma in prostatic biopsies predict prostate-specific antigen recurrence? A systematic review. *Eur Urol* 2008;54(4):728-39.

148. Briganti A, Chun FK, Hutterer GC, et al. Systematic assessment of the ability of the number and percentage of positive biopsy cores to predict pathologic stage and biochemical recurrence after radical prostatectomy. *Eur Urol* 2007;52(3):733-43.

149. Quintal MM, Meirelles LR, Freitas LL, et al. Various morphometric measurements of cancer extent on needle prostatic biopsies: which is predictive of pathologic stage and biochemical recurrence following radical prostatectomy? *Int Urol Nephrol* 2011;43(3):697-705.

150. Brimo F, Vollmer RT, Corcos J, et al. Prognostic value of various morphometric measurements of tumour extent in prostate needle core tissue. *Histopathology* 2008;53(2):177-83.

151. Verhoef EI, Kweldam CF, Kummerlin IP, et al. Comparison of Tumor Volume Parameters on Prostate Cancer Biopsies. *Arch Pathol Lab Med* 2020.

152. Freedland SJ, Aronson WJ, Terris MK, et al. The percentage of prostate needle biopsy cores with carcinoma from the more involved side of the biopsy as a predictor of prostate specific antigen recurrence after radical prostatectomy: results from the Shared Equal Access Regional Cancer Hospital (SEARCH) database. *Cancer* 2003;98(11):2344-50.

153. Linson PW, Lee AK, Doytchinova T, et al. Percentage of core lengths involved with prostate cancer: does it add to the percentage of positive prostate biopsies in predicting postoperative prostate-specific antigen outcome for men with intermediate-risk prostate cancer? *Urology* 2002;59(5):704-8.

154. Qian Y, Feng FY, Halverson S, et al. The percent of positive biopsy cores improves prediction of prostate cancer-specific death in patients treated with dose-escalated radiotherapy. *Int J Radiat Oncol Biol Phys* 2011;81(3):e135-42.

155. Rajab R, Fisher G, Kattan MW, et al. Measurements of cancer extent in a conservatively treated prostate cancer biopsy cohort. *Virchows Arch* 2010;457(5):547-53.

156. Freedland SJ, Aronson WJ, Csathy GS, et al. Comparison of percentage of total prostate needle biopsy tissue with cancer to percentage of cores with cancer for predicting PSA recurrence after radical prostatectomy: results from the SEARCH database. *Urology* 2003;61(4):742-7.

157. Cuzick J, Fisher G, Kattan MW, et al. Long-term outcome among men with conservatively treated localised prostate cancer. *Br J Cancer* 2006;95(9):1186-94.

158. Vance SM, Stenmark MH, Blas K, et al. Percentage of cancer volume in biopsy cores is prognostic for prostate cancer death and overall survival in patients treated with dose-escalated external beam radiotherapy. *Int J Radiat Oncol Biol Phys* 2012;83(3):940-6.

159. Bul M, Zhu X, Valdagni R, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. *Eur Urol* 2013;63(4):597-603.

160. Montironi R, Hammond EH, Lin DW, et al. Consensus statement with recommendations on active surveillance inclusion criteria and definition of progression in men with localized prostate cancer: the critical role of the pathologist. *Virchows Arch* 2014;465(6):623-8.

161. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. *J Clin Oncol* 2011;29(16):2185-90.

162. Luo X, Khurana JS, Jhala N, et al. The Association of Invasive Cribriform Lesions With Adverse Prostatic Adenocarcinoma Outcomes: An Institutional Experience, Systematic Review, and Meta-analysis. *Arch Pathol Lab Med* 2019;143(8):1012-21.

163. Iczkowski KA, Torkko KC, Kotnis GR, et al. Digital quantification of five high-grade prostate cancer patterns, including the cribriform pattern, and their association with adverse outcome. *Am J Clin Pathol* 2011;136(1):98-107.

164. Dong F, Yang P, Wang C, et al. Architectural heterogeneity and cribriform pattern predict adverse clinical outcome for Gleason grade 4 prostatic adenocarcinoma. *Am J Surg Pathol* 2013;37(12):1855-61.

165. McKenney JK, Wei W, Hawley S, et al. Histologic Grading of Prostatic Adenocarcinoma Can Be Further Optimized: Analysis of the Relative Prognostic Strength of Individual Architectural Patterns in 1275 Patients From the Canary Retrospective Cohort. *Am J Surg Pathol* 2016;40(11):1439-56.

166. Choy B, Pearce SM, Anderson BB, et al. Prognostic Significance of Percentage and Architectural Types of Contemporary Gleason Pattern 4 Prostate Cancer in Radical Prostatectomy. *Am J Surg Pathol* 2016;40(10):1400-6.

167. Keefe DT, Schieda N, El Hallani S, et al. Cribriform morphology predicts upstaging after radical prostatectomy in patients with Gleason score 3 + 4 = 7 prostate cancer at transrectal ultrasound (TRUS)-guided needle biopsy. *Virchows Arch* 2015;467(4):437-42.

168. Kweldam CF, Kummerlin IP, Nieboer D, et al. Presence of invasive cribriform or intraductal growth at biopsy outperforms percentage grade 4 in predicting outcome of Gleason score 3+4=7 prostate cancer. *Mod Pathol* 2017;30(8):1126-32.

169. Kimura K, Tsuzuki T, Kato M, et al. Prognostic value of intraductal carcinoma of the prostate in radical prostatectomy specimens. *Prostate* 2014;74(6):680-7.

170. Van der Kwast T, Al Daoud N, Collette L, et al. Biopsy diagnosis of intraductal carcinoma is prognostic in intermediate and high risk prostate cancer patients treated by radiotherapy. *Eur J Cancer* 2012;48(9):1318-25.

171. Trudel D, Downes MR, Sykes J, et al. Prognostic impact of intraductal carcinoma and large cribriform carcinoma architecture after prostatectomy in a contemporary cohort. *Eur J Cancer* 2014;50(9):1610-6.

172. Kweldam CF, Kummerlin IP, Nieboer D, et al. Disease-specific survival of patients with invasive cribriform and intraductal prostate cancer at diagnostic biopsy. *Mod Pathol* 2016;29(6):630-6.

173. van Leenders G, Kweldam CF, Hollemans E, et al. Improved Prostate Cancer Biopsy Grading by Incorporation of Invasive Cribriform and Intraductal Carcinoma in the 2014 Grade Groups. *Eur Urol* 2020;77(2):191-8.

174. Cole AI, Morgan TM, Spratt DE, et al. Prognostic Value of Percent Gleason Grade 4 at Prostate Biopsy in Predicting Prostatectomy Pathology and Recurrence. *J Urol* 2016;196(2):405-11.

175. Huang CC, Kong MX, Zhou M, et al. Gleason score 3 + 4=7 prostate cancer with minimal quantity of gleason pattern 4 on needle biopsy is associated with low-risk tumor in radical prostatectomy specimen. *Am J Surg Pathol* 2014;38(8):1096-101.

176. Perlis N, Sayyid R, Evans A, et al. Limitations in Predicting Organ Confined Prostate Cancer in Patients with Gleason Pattern 4 on Biopsy: Implications for Active Surveillance. *J Urol* 2017;197(1):75-83.

177. Dean LW, Assel M, Sjoberg DD, et al. Clinical Usefulness of Total Length of Gleason Pattern 4 on Biopsy in Men with Grade Group 2 Prostate Cancer. *J Urol* 2019;201(1):77-82.

178. Kir G, Seneldir H, Gumus E. Outcomes of Gleason score 3 + 4 = 7 prostate cancer with minimal amounts (<6%) vs >/=6% of Gleason pattern 4 tissue in needle biopsy specimens. *Ann Diagn Pathol* 2016;20:48-51.

179. Berney DM, Beltran L, Sandu H, et al. The percentage of high-grade prostatic adenocarcinoma in prostate biopsies significantly improves on Grade Groups in the prediction of prostate cancer death. *Histopathology* 2019;75(4):589-97.

180. Sauter G, Steurer S, Clauditz TS, et al. Clinical Utility of Quantitative Gleason Grading in Prostate Biopsies and Prostatectomy Specimens. *Eur Urol* 2016;69(4):592-8.

181. Ayala GE, Dai H, Ittmann M, et al. Growth and survival mechanisms associated with perineural invasion in prostate cancer. *Cancer Res* 2004;64(17):6082-90.

182. Zhang LJ, Wu B, Zha ZL, et al. Perineural invasion as an independent predictor of biochemical recurrence in prostate cancer following radical prostatectomy or radiotherapy: a systematic review and meta-analysis. *BMC Urol* 2018;18(1):5.

183. Harnden P, Shelley MD, Clements H, et al. The prognostic significance of perineural invasion in prostatic cancer biopsies: a systematic review. *Cancer* 2007;109(1):13-24.

184. Meng Y, Liao YB, Xu P, et al. Perineural invasion is an independent predictor of biochemical recurrence of prostate cancer after local treatment: a meta-analysis. *Int J Clin Exp Med* 2015;8(8):13267-74.

185. Wu S, Lin X, Lin SX, et al. Impact of biopsy perineural invasion on the outcomes of patients who underwent radical prostatectomy: a systematic review and meta-analysis. *Scand J Urol* 2019;53(5):287-94.

186. Strom P, Nordstrom T, Delahunt B, et al. Prognostic value of perineural invasion in prostate needle biopsies: a population-based study of patients treated by radical prostatectomy. *J Clin Pathol* 2020.

187. DeLancey JO, Wood DP, Jr., He C, et al. Evidence of perineural invasion on prostate biopsy specimen and survival after radical prostatectomy. *Urology* 2013;81(2):354-7.

188. Loeb S, Epstein JI, Humphreys EB, et al. Does perineural invasion on prostate biopsy predict adverse prostatectomy outcomes? *BJU Int* 2010;105(11):1510-3.

189. de la Taille A, Rubin MA, Bagiella E, et al. Can perineural invasion on prostate needle biopsy predict prostate specific antigen recurrence after radical prostatectomy? *J Urol* 1999;162(1):103-6.

190. Saeter T, Bogaard M, Vlatkovic L, et al. The relationship between perineural invasion, tumor grade, reactive stroma and prostate cancer-specific mortality: A clinicopathologic study on a population-based cohort. *Prostate* 2016;76(2):207-14.

191. Zareba P, Flavin R, Isikbay M, et al. Perineural Invasion and Risk of Lethal Prostate Cancer. *Cancer Epidemiol Biomarkers Prev* 2017;26(5):719-26.

192. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2013;137(12):1710-22.

193. Goacher E, Randell R, Williams B, et al. The Diagnostic Concordance of Whole Slide Imaging and Light Microscopy: A Systematic Review. *Arch Pathol Lab Med* 2017;141(1):151-61.

194. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology* 2017;70(1):134-45.

195. Al-Janabi S, Huisman A, Van Diest PJ. Digital pathology: current status and future perspectives. *Histopathology* 2012;61(1):1-9.

196. Rocha R, Vassallo J, Soares F, et al. Digital slides: present status of a tool for consultation, teaching, and quality control in pathology. *Pathol Res Pract* 2009;205(11):735-41.

197. Weinstein RS, Graham AR, Richter LC, et al. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Hum Pathol* 2009;40(8):1057-69.

198. Chargari C, Comperat E, Magne N, et al. Prostate needle biopsy examination by means of virtual microscopy. *Pathol Res Pract* 2011;207(6):366-9.

199. Helin H, Lundin M, Lundin J, et al. Web-based virtual microscopy in teaching and standardizing Gleason grading. *Hum Pathol* 2005;36(4):381-6.

200. Fine JL, Grzybicki DM, Silowash R, et al. Evaluation of whole slide image immunohistochemistry interpretation in challenging prostate needle biopsies. *Hum Pathol* 2008;39(4):564-72.

201. Van Hemelrijck M, Wigertz A, Sandin F, 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.

202. Tomic K, Berglund A, Robinson D, et al. Capture rate and representativity of The National Prostate Cancer Register of Sweden. *Acta Oncol* 2015;54(2):158-63.

203. Hagel E, Garmo H, Bill-Axelson A, et al. PCBaSe Sweden: a register-based resource for prostate cancer research. *Scand J Urol Nephrol* 2009;43(5):342-9.

204. National Board of Health and Welfare, Centre for Epidemiology. The National Patient Register. Stockholm. (https://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish). (Accessed 17.06.2020.).

205. Barlow L, Westergren K, Holmberg L, et al. The completeness of the Swedish Cancer Register: a sample survey for year 1998. *Acta Oncol* 2009;48(1):27-33.

206. Statistics Sweden. The Total Population Register. (https://www.scb.se/contentassets/8f66bcf5abc34d0b98afa4fcbfc0e060/rtb-bar-2016-eng.pdf). (Accessed 17.06.2020.).

207. Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 2016;31(2):125-36.

208. Brooke HL, Talback M, Hornblad J, et al. The Swedish cause of death register. *Eur J Epidemiol* 2017;32(9):765-73.

209. Fall K, Stromberg F, Rosell J, et al. Reliability of death certificates in prostate cancer patients. *Scand J Urol Nephrol* 2008;42(4):352-7.

210. Godtman R, Holmberg E, Stranne J, et al. 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.

211. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.

212. Statistics Sweden. Longitudinal integrated database for health insurance and labour market studies (LISA). (https://www.scb.se/en/services/guidance-for-researchers-and-universities/vilka-mikrodata-finns/longitudinella-register/longitudinal-integrated-database-for-health-insurance-and-labour-market-studies-lisa/). (Accessed 16.06.2020.).

213. Ludvigsson JF, Svedberg P, Olen O, et al. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *Eur J Epidemiol* 2019;34(4):423-37.

214. Mohler JL, Armstrong AJ, Bahnson RR, et al. Prostate Cancer, Version 1.2016. *J Natl Compr Canc Netw* 2016;14(1):19-30.

215. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol* 2010;10:112.

216. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30(4):377-99.

217. Johansson S, Astrom L, Sandin F, et al. Hypofractionated proton boost combined with external beam radiotherapy for treatment of localized prostate cancer. *Prostate Cancer* 2012;2012:654861.

218. Zelic R, Zugna D, Bottai M, et al. Estimation of Relative and Absolute Risks in a Competing-Risks Setting Using a Nested Case-Control Study Design: Example From the ProMort Study. *Am J Epidemiol* 2019;188(6):1165-73.

219. Allan C, Burel JM, Moore J, et al. OMERO: flexible, model-driven data management for experimental biology. *Nat Methods* 2012;9(3):245-53.

220. Lianas L, Piras ME, Musu E, et al. CyTest - An Innovative Open-source Platform for Training and Testing in Cythopathology. *Procd Soc Behv* 2016;228:674-81.

221. Kim RS. Analysis of Nested Case-Control Study Designs: Revisiting the Inverse Probability Weighting Method. *Communications for Statistical Applications and Methods* 2013;20(6):455–66.

222. Kim RS. A new comparison of nested case-control and case-cohort designs and methods. *Eur J Epidemiol* 2015;30(3):197-207.

223. Langholz B, Borgan O. Estimation of absolute risk from nested case-control data. *Biometrics* 1997;53(2):767-74.

224. Salim A, Delcoigne B, Villaflores K, et al. Comparisons of risk prediction methods using nested case-control data. *Stat Med* 2017;36(3):455-65.

225. Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika* 1997;84(2):379-94.

226. Stoer NC, Samuelsen SO. Inverse probability weighting in nested case-control studies with additional matching--a simulation study. *Stat Med* 2013;32(30):5328-39.

227. Kim RS, Kaplan RC. Analysis of secondary outcomes in nested case-control study designs. *Stat Med* 2014;33(24):4215-26.

228. Stoer NC, Meyer HE, Samuelsen SO. Reuse of controls in nested case-control studies. *Epidemiology* 2014;25(2):315-7.

229. Saarela O, Kulathinal S, Arjas E, et al. Nested case-control data utilized for multiple outcomes: a likelihood approach and alternatives. *Stat Med* 2008;27(28):5991-6008.

230. Stoer NC, Samuelsen SO. Comparison of estimators in nested case-control studies with multiple outcomes. *Lifetime Data Anal* 2012;18(3):261-83.

231. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol* 2012;41(3):861-70.

232. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009;170(2):244-56.

233. Wolkewitz M, Cooper BS, Palomar-Martinez M, et al. Nested case-control studies in cohorts with competing events. *Epidemiology* 2014;25(1):122-5.

234. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999;94(446):496-509.

235. Cox DR. Regression Models and Life-Tables. *J R Stat Soc B* 1972;34(2):187-+.

236. Hinchliffe SR, Lambert PC. Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions. *BMC Medical Research Methodology* 2013;13(1):13.

237. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002;21.

238. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata J* 2009;9(2):265-90.

239. Hinchliffe SR, Lambert PC. Extending the flexible parametric survival model for competing risks. *Stata J* 2013;13(2):344-55.

240. Borgan O, Keogh R. Nested case-control studies: should one break the matching? *Lifetime Data Anal* 2015;21(4):517-41.

241. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013;13:33.

242. Royston P, Parmar M, Altman DG. External validation and updating of a prognostic survival model. Department of statistical science, University College London, 2010.

243. Coviello V, Boggess M. Cumulative incidence estimation in the presence of competing risks. *Stata J* 2004;4(2):103-12.

244. Newson RB. Comparing the predictive powers of survival models using Harrell's C or Somers' D. *Stata J* 2010;10(3):339-58.

245. Wolbers M, Koller MT, Witteman JC, et al. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009;20(4):555-61.

246. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009;9:57.

247. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* 1960;20(1):37-46.

248. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70(4):213-20.

249. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.

250. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.

251. Bamber D. Area above Ordinal Dominance Graph and Area Below Receiver Operating Characteristic Graph. *J Math Psychol* 1975;12(4):387-415.

252. Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. *J Clin Oncol* 2015;33(3):272-7.

253. Zelic R, Garmo H, Zugna D, et al. Predicting Prostate Cancer Death with Different Pretreatment Risk Stratification Tools: A Head-to-head Comparison in a Nationwide Cohort Study. *Eur Urol* 2020;77(2):180-8.

254. Langholz B, Goldstein L. Risk set sampling in epidemiologic cohort studies. *Stat Sci* 1996;11(1):35-53.

255. Liddell FDK, Mcdonald JC, Thomas DC. Methods of Cohort Analysis - Appraisal by Application to Asbestos Mining. *J Roy Stat Soc a Sta* 1977;140:469-91.

256. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98(9):683-90.

257. Støer NC, Samuelsen SO. MultipleNCC: Inverse probability weighting of nested case-control data. *R Journal* 2016;8(2):5-18.

258. Daskivich TJ, Chamie K, Kwan L, et al. Comorbidity and competing risks for mortality in men with prostate cancer. *Cancer* 2011;117(20):4642-50.

259. Epstein MM, Edgren G, Rider JR, et al. Temporal trends in cause of death among Swedish and US men with prostate cancer. *J Natl Cancer Inst* 2012;104(17):1335-42.

260. Stattin P, Holmberg E, Johansson JE, et al. Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study. *J Natl Cancer Inst* 2010;102(13):950-8.

261. Langholz B, Borgan O. Counter-Matching - a Stratified Nested Case-Control Sampling Method. *Biometrika* 1995;82(1):69-79.

262. Ohneberg K. *Sampling Designs for Complex Time-to-event Data in Clinical and Epidemiological Studies*. Albert-Ludwigs-Universität Freiburg; 2019.

263. Cooperberg MR. Clinical risk-stratification for prostate cancer: Where are we, and where do we need to go? *Can Urol Assoc J* 2017;11(3-4):101-2.

264. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115(7):928-35.

265. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 2007;26(30):5512-28.

266. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 1st ed. New York, New York: Springer; 2009.

267. Janssen KJ, Moons KG, Kalkman CJ, et al. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;61(1):76-86.

268. Steyerberg EW, Borsboom GJ, van Houwelingen HC, et al. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;23(16):2567-86.

269. Memorial Sloan Kettering Center. Dynamic Prostate Cancer Nomogram: Coefficients. (https://www.mskcc.org/nomograms/prostate/pre_op/coefficients). (Accessed 23.07.2020).

270. Cowley LE, Farewell DM, Maguire S, et al. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagn Progn Res* 2019;3:16.

271. Rosella LC, Corey P, Stukel TA, et al. The influence of measurement error on calibration, discrimination, and overall estimation of a risk prediction model. *Popul Health Metr* 2012;10(1):20.

272. Khudyakov P, Gorfine M, Zucker D, et al. The impact of covariate measurement error on risk prediction. *Stat Med* 2015;34(15):2353-67.

273. Luijken K, Groenwold RHH, Van Calster B, et al. Impact of predictor measurement heterogeneity across settings on the performance of prediction models: A measurement error perspective. *Stat Med* 2019;38(18):3444-59.

274. Pajouheshnia R, van Smeden M, Peelen LM, et al. How variation in predictor measurement affects the discriminative ability and transportability of a prediction model. *J Clin Epidemiol* 2019;105:136-41.

275. Luijken K, Wynants L, van Smeden M, et al. Changing predictor measurement procedures affected the performance of prediction models in clinical examples. *J Clin Epidemiol* 2020;119:7-18.

276. Kweldam CF, Nieboer D, Algaba F, et al. Gleason grade 4 prostate adenocarcinoma patterns: an interobserver agreement study among genitourinary pathologists. *Histopathology* 2016;69(3):441-9.

277. Sadimin ET, Khani F, Diolombi M, et al. Interobserver Reproducibility of Percent Gleason Pattern 4 in Prostatic Adenocarcinoma on Prostate Biopsies. *Am J Surg Pathol* 2016;40(12):1686-92.

278. Shah RB, Li J, Cheng L, et al. Diagnosis of Gleason pattern 5 prostate adenocarcinoma on core needle biopsy: an interobserver reproducibility study among urologic pathologists. *Am J Surg Pathol* 2015;39(9):1242-9.

279. Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 2009;119(17):2408-16.

280. Davidsson S, Fiorentino M, Andren O, et al. Inflammation, focal atrophic lesions, and prostatic intraepithelial neoplasia with respect to risk of lethal prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2011;20(10):2280-7.

281. Tolonen TT, Kujala PM, Tammela TL, et al. Overall and worst gleason scores are equally good predictors of prostate cancer progression. *BMC Urol* 2011;11:21.

282. Trpkov K, Sangkhamanon S, Yilmaz A, et al. Concordance of "Case Level" Global, Highest, and Largest Volume Cancer Grade Group on Needle Biopsy Versus Grade Group on Radical Prostatectomy. *Am J Surg Pathol* 2018;42(11):1522-9.

283. van Houwelingen H, Putter H. *Dynamic Prediction in Clinical Survival Analysis*. Baton Rouge, UNITED STATES: Taylor & Francis Group; 2011.

284. Lenert MC, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless. *J Am Med Inform Assoc* 2019;26(12):1645-50.

285. Siregar S, Nieboer D, Vergouwe Y, et al. Improved Prediction by Dynamic Modeling: An Exploratory Study in the Adult Cardiac Surgery Database of the Netherlands Association for Cardio-Thoracic Surgery. *Circ Cardiovasc Qual Outcomes* 2016;9(2):171-81.

286. Hickey GL, Grant SW, Caiado C, et al. Dynamic prediction modeling approaches for cardiac surgery. *Circ Cardiovasc Qual Outcomes* 2013;6(6):649-58.

287. Toll DB, Janssen KJ, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61(11):1085-94.

288. Debray TP, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;68(3):279-89.

289. Jenkins DA, Sperrin M, Martin GP, et al. Dynamic models to predict health outcomes: current status and methodological challenges. *Diagn Progn Res* 2018;2:23.

290. Strobl AN, Vickers AJ, Van Calster B, et al. Improving patient prostate cancer risk assessment: Moving from static, globally-applied to dynamic, practice-specific risk calculators. *J Biomed Inform* 2015;56:87-93.

291. Halabi S, Li C, Luo S. Developing and Validating Risk Assessment Models of Clinical Outcomes in Modern Oncology. *JCO Precis Oncol* 2019;3.

292. Bull LM, Lunt M, Martin GP, et al. Harnessing repeated measurements of predictor variables for clinical risk prediction: a review of existing methods. *Diagn Progn Res* 2020;4:9.

293. Nicolaie MA, van Houwelingen JC, de Witte TM, et al. Dynamic prediction by landmarking in competing risks. *Stat Med* 2013;32(12):2031-47.

294. Nicolaie MA, van Houwelingen JC, de Witte TM, et al. Dynamic pseudo-observations: a robust approach to dynamic prediction in competing risks. *Biometrics* 2013;69(4):1043-52.

295. Liu Q, Tang G, Costantino JP, et al. Landmark Proportional Subdistribution Hazards Models for Dynamic Prediction of Cumulative Incidence Functions [electronic article]. Advance Access: April 01, 2019.

296. Proust-Lima C, Taylor JM. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* 2009;10(3):535-49.

297. Taylor JM, Park Y, Ankerst DP, et al. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* 2013;69(1):206-13.

298. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 2017;389(10071):815-22.

299. Faria R, Soares MO, Spackman E, et al. Optimising the Diagnosis of Prostate Cancer in the Era of Multiparametric Magnetic Resonance Imaging: A Cost-effectiveness Analysis Based on the Prostate MR Imaging Study (PROMIS). *Eur Urol* 2018;73(1):23-30.

300. Elkhoury FF, Felker ER, Kwan L, et al. Comparison of Targeted vs Systematic Prostate Biopsy in Men Who Are Biopsy Naive: The Prospective Assessment of Image Registration in the Diagnosis of Prostate Cancer (PAIREDCAP) Study. *JAMA Surg* 2019;154(9):811-8.

301. Rouviere O, Puech P, Renard-Penna R, et al. Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. *Lancet Oncol* 2019;20(1):100-9.

302. Kenigsberg AP, Renson A, Rosenkrantz AB, et al. Optimizing the Number of Cores Targeted During Prostate Magnetic Resonance Imaging Fusion Target Biopsy. *Eur Urol Oncol* 2018;1(5):418-25.

303. Groenwold RH, Moons KG, Pajouheshnia R, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *J Clin Epidemiol* 2016;78:90-100.

304. Pajouheshnia R, Peelen LM, Moons KGM, et al. Accounting for treatment use when validating a prognostic model: a simulation study. *BMC Med Res Methodol* 2017;17(1):103.

305. Sperrin M, Martin GP, Pate A, et al. Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. *Stat Med* 2018;37(28):4142-54.

306. Pajouheshnia R, Schuster NA, Groenwold RHH, et al. Accounting for time-dependent treatment use when developing a prognostic model from observational data: A review of methods. *Stat Neerl* 2019.

307. van Geloven N, Swanson SA, Ramspek CL, et al. Prediction meets causal inference: the role of treatment in clinical prediction models. *Eur J Epidemiol* 2020.

308. ICH E9 working group. ICH E9 (R1): addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. EMA/CHMP/ICH/436221/2017, 2020,

309. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144(3):201-9.

310. Sachs MC, Sjolander A, Gabriel EE. Aim for Clinical Utility, Not Just Predictive Accuracy. *Epidemiology* 2020;31(3):359-64.

311. Clarke MG, Ewings P, Hanna T, et al. How accurate are doctors, nurses and medical students at predicting life expectancy? *Eur J Intern Med* 2009;20(6):640-4.

312. Sammon JD, Abdollah F, D'Amico A, et al. Predicting Life Expectancy in Men Diagnosed with Prostate Cancer. *Eur Urol* 2015;68(5):756-65.

313. Bhatt NR, Davis NF, Breen K, et al. Life expectancy calculation in urology: Are we equitably treating older patients? *Cent European J Urol* 2017;70(4):368-71.

314. Daskivich TJ, Fan KH, Koyama T, et al. Effect of age, tumor risk, and comorbidity on competing risks for survival in a U.S. population-based cohort of men with prostate cancer. *Ann Intern Med* 2013;158(10):709-17.

315. Albertsen PC, Moore DF, Shih W, et al. Impact of comorbidity on survival among men with localized prostate cancer. *J Clin Oncol* 2011;29(10):1335-41.

316. Daskivich TJ. The Importance of Accurate Life Expectancy Prediction in Men with Prostate Cancer. *Eur Urol* 2015;68(5):766-7.

317. Kalra S, Basourakos S, Abouassi A, et al. The implications of ageing and life expectancy in prostate cancer treatment. *Nat Rev Urol* 2016;13(5):289-95.

318. Kent M, Vickers AJ. A systematic literature review of life expectancy prediction tools for patients with localized prostate cancer. *J Urol* 2015;193(6):1938-42.