Pre-treatment risk stratification of prostate cancer patients: A critical review

George Rodrigues, MD, FRCPC, MSc;† Padraig Warde, FRCPC, MB;‡ Tom Pickles, MD, FRCPC;§ Juanita Crook, MD, FRCPC;¶ Michael Brundage, MD, FRCPC, MSc;¶ Luis Souhami, MD, FRCPC;* Himu Lukka, MD, FRCPC;§ on behalf of the Genitourinary Radiation Oncologists of Canada

Abstract

Introduction: The use of accepted prostate cancer risk stratification groups based on prostate-specific antigen, T stage and Gleason score assists in therapeutic treatment decision-making, clinical trial design and outcome reporting. The utility of integrating novel prognostic factors into an updated risk stratification schema is an area of current debate. The purpose of this work is to critically review the available literature on novel pre-treatment prognostic factors and alternative prostate cancer risk stratification schema to assess the feasibility and need for changes to existing risk stratification systems.

Methods: A systematic literature search was conducted to identify original research publications and review articles on prognostic factors and risk stratification in prostate cancer. Search terms included risk stratification, risk assessment, prostate cancer or neoplasms, and prognostic factors. Abstracted information was assessed to draw conclusions regarding the potential utility of changes to existing risk stratification schema.

Results: The critical review identified three specific clinically relevant potential changes to the most commonly used three-group risk stratification system: (1) the creation of a very-low risk category; (2) the splitting of intermediate-risk into a low- and high-intermediate risk groups; and (3) the clarification of the interface between intermediate- and high-risk disease. Novel pathological factors regarding high-grade cancer, subtypes of Gleason score 7 and percentage biopsy cores positive were also identified as potentially important risk-stratification factors.

Conclusions: Multiple studies of prognostic factors have been performed to create currently utilized prostate cancer risk stratification systems. We propose potential changes to existing systems.

Introduction

Predictive modelling is a process by which a mathematical construct is created and ideally validated to predict the future probability of an outcome. While the concept and application of predictive modelling is not unique to medical decision-making, it is frequently used in various clinical scenarios for patients and physicians to weigh therapeutic options on comparative benefits and risks in domains, such as life expectancy, disease control, health-related quality-of-life and treatment complications. Other common goals for the clinical use of predictive modelling include the reduction of inconsistent management due to physician bias, or differences in experience and knowledge, as well as the rational design of clinical trials in terms of inclusion/exclusion criteria and stratification.

Various methodological options exist in the production and presentation of predictive models. Predictive modelling options include (in general order of increasing accuracy and complexity): look-up tables, risk stratification/classification analyses, regression-tree analyses, nomograms and artificial neural networks. The choice of a predictive model depends on the nature of the outcome to be predicted, the predictive variables available and the ultimate purpose of the model in terms of how the end-user is likely to best interact with the model information. In addition, all prediction models ideally should be assessed for various methodological domains, including model accuracy, validity and complexity.

Risk stratification systems serve multiple purposes, including guiding decision-making, providing clinical trial design stratification options or facilitating inter-physician or organizational data/outcome exchange by the establishment of a common nomenclature. In the oncology context, risk stratification can identify clinical situations where multimodality therapy should be used in high-risk scenarios and conversely, either single modality and/or observational strategies in lower-risk patients. Risk stratification groups are routinely used to define a class of patients to be included (or excluded) from clinical trials that are designed to investigate therapies targeted for a specific risk group (e.g., adjuvant chemotherapy for high-risk patients or active surveillance strategies for patients with low risk of progression). In clinical
trials with patients with a broad range of risk characteristics, risk groups can be used for stratification. Just as important, the establishment of a risk stratification schema creates a common nomenclature to allow patients, physicians, institutions, government agencies and clinical trial cooperative groups to present and/or compare outcome data in a variety of clinical and academic scenarios.

The Genitourinary Radiation Oncologists of Canada (GUROC) have published a consensus statement on the management of prostate cancer in 2001. This consensus statement included details on a risk stratification scheme for a common definition of low-intermediate- and high-risk non-metastatic prostate cancer in the pre-therapy setting. Since this initial risk stratification consensus statement, a plethora of pre- and post-treatment predictive models, including both currently employed and proposed risk stratification systems, have been published relating to clinical and biochemical prostate cancer outcomes. At the biennial November 2009 GUROC meeting, GUROC members had preliminary discussions on either reaffirming or altering the Canadian Consensus Prostate Cancer Risk Stratification Guidelines. This review was commissioned prior to any change to critically assess the evidence for the introduction of non-classical (i.e., non-initial prostate-specific antigen [PSA], biopsy Gleason score and clinical T stage) prognostic factors into pre-treatment risk stratification systems, as well as to review all current and proposed risk stratification systems in the medical literature.

Methods

A PubMed search was conducted to identify original or review articles addressing pre-treatment assessment of prognostic factors and risk stratification in prostate cancer from 1966 to the present. Relevant search terms included: risk stratification, risk assessment, prostate cancer or neoplasms and prognostic factors. Articles assessing post-treatment prognostic factors/risk stratification schema were not considered. Relevant information abstracted from identified studies was then assessed regarding the utility of the information to inform potential changes to the three-group risk stratification schema previously adopted by GUROC.

Results: evidence review

Classical prognostic factors

Various parameters have been routinely considered to assess outcome as it relates to prostate cancer treatment. These prostate cancer endpoints are usually clinical (overall survival, disease-free survival, metastasis-free survival), surgical (rates of extracapsular disease, seminal vesicle involvement, positive margins and lymph node positivity) or biochemical (PSA biochemical-free failure). Pre-treatment PSA, clinical T category and biopsy-based Gleason score have been shown to be independently predictive of various combinations of prostate cancer related endpoints in a variety of treatment scenarios in the non-metastatic setting. A variety of clinical staging systems have been used to assign T categories over the past few decades, thus making comparisons of various publications and predictive models challenging. Specifically, the 1997 American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (UICC) system simplifies unilateral versus bilateral organ-confined but palpable disease into T2a and T2b categories; however, other tumour, node, metastases (TNM) staging systems prior to and since 1997 used three categories for organ-confined palpable disease: (1) T2a – unilateral one-half of one lobe or less; (2) T2b – unilateral greater than one-half of one lobe; and (3) T2c – bilateral disease. Another staging system commonly used is one described by Jewett. This system stratifies patients into four progressively more advanced stage groupings based on various combinations of T, N and M categories.

Since 1966, prostate tumour grading based on glandular differentiation and architecture on biopsies has been routinely used as a prognostic indicator of outcome. Gleason grading can be reported in a variety of ways including: primary, secondary, and tertiary, total Gleason score (sum of primary and secondary grades), modified Gleason score (e.g., 3+4 with tertiary pattern 5 equals overall Gleason 8 cancer) and subclassifications of Gleason scores (e.g., Gleason 7; 4+3 vs. 3+4). Various interpretations can be applied to determine final Gleason scores to be assigned to a biopsy set prior to a management decision. Final Gleason scores can be assigned either as maximum Gleason score from a single biopsy core, from a specific prostate region, or a “pathologist consensus” Gleason score for the entire biopsy set. This may explain, in part, the significant levels of intra- and inter-pathologist variability in Gleason grading, as well as poor correlation between biopsy and post-prostatectomy Gleason scores. Despite these well-documented challenges, Gleason grading remains the main pathological reporting system for prostate cancer due a consistent strong association with prostate cancer outcome.

Pre-treatment initial PSA is routinely available prior to a treatment decision due to the role of PSA as a screening tool. Increasing levels of PSA prior to treatment have been shown to be associated with increasing tumour volume/stage and Gleason score, the risk of extracapsular/semenal vesicle or lymph node involvement or positive margins, and ultimately with prostate cancer outcomes. The relationships between PSA, Gleason score and clinical T stage have been described by many investigators in terms of progressively worse biochemical and prostate cancer outcomes.
Risk stratification in prostate cancer

with increasing PSA level. Early reports by Partin and colleagues and D’Amico and colleagues defined important PSA cutpoints of <10, 10.1-20, and >20 ng/mL that are still used to define prostate cancer risk stratification groups.

Over 20 pre-treatment predictive models (probability graphs, nomograms, look-up tables, and neural networks) using various combinations of these three classical prognostic factors exist. The goal of the first published classical predictive model was to predict the risk of post-prostatectomy seminal vesicle invasion based on a cohort of 2953 patients. In 1997, Partin and colleagues published the first look-up tables relating initial PSA, Gleason score and clinical T stage with rates of organ-confined disease, positive margins, and the risk of seminal vesicle and lymph node positivity after radical prostatectomy. The Partin tables have been subjected to various validation investigations with mixed results in terms of their statistical operating characteristics. Other important predictive models include the Kattan and Stephenson nomograms. The Kattan nomogram relates the classical prognostic factors with biochemical recurrence, while the recently reported Stephenson nomogram relates the three classical prognostic factors to 15-year cancer-specific mortality in a cohort of 12,677 radical prostatectomy patients with a reported internal accuracy of 82%.

Common prostate cancer risk stratification systems

Based on the prognostic power of initial PSA, biopsy Gleason score and clinical T stage, a variety of pre-treatment prostate cancer risk stratification systems have been reported. In 1998, D’Amico and colleagues first proposed a three-group risk stratification system to predict post-treatment biochemical failure after radical prostatectomy and external-beam radiotherapy. This system divided non-metastatic patients into low-, intermediate-, and high-risk based on initial PSA, clinical T stage and biopsy Gleason score. Low-risk prostate cancer was defined as 1992 AJCC T1/T2a, and PSA ≤10 ng/mL, and Gleason score ≤6. Intermediate-risk prostate cancer was defined as 1992 AJCC T2b, and/or PSA 10-20 ng/mL and/or Gleason 7 disease. High-risk disease was classified as having any one of the following high-risk features: 1992 AJCC ≥T2c, PSA >20 ng/mL or Gleason 8-10 disease. In 2001, the GUROC published the results of a consensus meeting on the topics of risk assessment, conformal radiotherapy, brachytherapy and combined hormonal therapy. A consensus for prostate cancer risk stratification was reached at this meeting; consensus was based on a review and discussion of the available literature of clinical risk factors related to biochemical failure. It was noted that differences existed in the literature on the exact definitions of risk categorization. However, consensus was achieved around three categories: (1) low – 1997 AJCC T1-T2a, PSA ≤10 ng/mL and Gleason ≤6; (2) intermediate – 1997 AJCC T1-T2, PSA ≤20 ng/mL and Gleason ≤7 not otherwise low-risk; and (3) high-risk – 1997 AJCC T3-T4 or PSA >20 ng/mL or Gleason 8-10. Additional institutional risk stratification definitions, in conjunction with reports of radiotherapy outcomes, have been reported.

Organizational classification systems have been developed by various cancer and urological organizations (Table 1), including the National Comprehensive Cancer Network (NCCN, USA), National Institute for Health and Clinical Excellence (NICE, UK), European Society of Medical Oncology (ESMO), American Urological Association (AUA) and the European Association of Urology (EAU). The NICE guidelines are consistent with the Canadian Consensus classification system, while the AUA and EAU have adopted classification systems consistent with that of D’Amico. Two important contemporary classification systems have been recently updated and have added additional risk groupings to the traditional three-group classifications. The NCCN

Common prostate cancer risk stratification systems

Based on the prognostic power of initial PSA, biopsy Gleason score and clinical T stage, a variety of pre-treatment prostate cancer risk stratification systems have been reported. In 1998, D’Amico and colleagues first proposed a three-group risk stratification system to predict post-treatment biochemical failure after radical prostatectomy and external-beam radiotherapy. This system divided non-metastatic patients into low-, intermediate-, and high-risk based on initial PSA, clinical T stage and biopsy Gleason score. Low-risk prostate cancer was defined as 1992 AJCC T1/T2a, and PSA ≤10 ng/mL, and Gleason score ≤6. Intermediate-risk prostate cancer was defined as 1992 AJCC T2b, and/or PSA 10-20 ng/mL and/or Gleason 7 disease. High-risk disease was classified as having any one of the following high-risk features: 1992 AJCC ≥T2c, PSA >20 ng/mL or Gleason 8-10 disease. In 2001, the GUROC published the results of a consensus meeting on the topics of risk assessment, conformal radiotherapy, brachytherapy and combined hormonal therapy. A consensus for prostate cancer risk stratification was reached at this meeting; consensus was based on a review and discussion of the available literature of clinical risk factors related to biochemical failure. It was noted that differences existed in the literature on the exact definitions of risk categorization. However, consensus was achieved around three categories: (1) low – 1997 AJCC T1-T2a, PSA ≤10 ng/mL and Gleason ≤6; (2) intermediate – 1997 AJCC T1-T2, PSA ≤20 ng/mL and Gleason ≤7 not otherwise low-risk; and (3) high-risk – 1997 AJCC T3-T4 or PSA >20 ng/mL or Gleason 8-10. Additional institutional risk stratification definitions, in conjunction with reports of radiotherapy outcomes, have been reported.

Organizational classification systems have been developed by various cancer and urological organizations (Table 1), including the National Comprehensive Cancer Network (NCCN, USA), National Institute for Health and Clinical Excellence (NICE, UK), European Society of Medical Oncology (ESMO), American Urological Association (AUA) and the European Association of Urology (EAU). The NICE guidelines are consistent with the Canadian Consensus classification system, while the AUA and EAU have adopted classification systems consistent with that of D’Amico. Two important contemporary classification systems have been recently updated and have added additional risk groupings to the traditional three-group classifications. The NCCN
guidelines currently include very low-risk (T1c, and Gleason score ≤6, PSA ≤10 ng/mL, <3 positive biopsy cores each ≤50% involved and PSA density of <0.15 ng/mLg) and very high-risk (T3b-T4) strata. Additionally, the AJCC and UICC have introduced initial PSA level and Gleason score into the staging system generally consistent with the D’Amico classification; however, they have retained stage III (T3a, T3b) and stage IV (T4, N1 or M1 groupings), which are consistent with previous editions of the staging manual. This major change to the AJCC/UICC was first proposed by Roach and colleagues in 2007 after results from a pooled analysis of 912 patients from the University of California-San Francisco and the University of Michigan. This data demonstrated the superiority of a three classical prognostic factor system versus one based on T stage classification alone in terms of predicting important clinical outcomes, including overall survival, disease-free survival and freedom from PSA failure.

**Novel prognostic factors**

The literature investigating novel prognostic factors in prostate cancer is vast; however, a recent systematic review performed by Sutcliffe and colleagues has synthesized the available information into a UK-based Health Technology Assessment (HTA) report. This review included assessments of clinical and pathological prognostic factors and models in the radical non-metastatic treatment setting. The authors looked for associations with important clinical endpoints, such as overall survival, disease-specific survival and biochemical and/or clinical recurrence. Various exclusion criteria (including the removal of studies with fewer than 200 patients, less than 5 years median follow-up, and abstract/review publications) were applied to the search strategy to evaluate clinical studies with strong methodology. A total of 21,947 abstracts were initially retrieved; ultimately, only 30 final reports met all inclusion and exclusion criteria. A variety of putative genetic (β-catenin, androgen CAG repeats receptor, CYP3A4 genotype, DNA ploidy, vitamin D receptor, Ki67, Bcl-2, p53, syndecan-1, CD10, stat5 activation), biochemical (acid phosphatase, creatinine), PSA related (PSA velocity and doubling time), pathological (Gleason variants 3+4 vs. 4+3, high grade components, modified Gleason score, % biopsy involvement, % cancer in surgical specimen) and clinical prognostic factors (maximum tumour dimension and tumour volume) were evaluated by the HTA research team. The following factors were promising due to adequate study quality and consistency of observed relationships between the factor and patient outcome: acid phosphatase level, PSA kinetics (velocity and doubling time), type of Gleason 7 (3+4 vs. 4+3), percentage positive biopsy cores and amount of high-grade (Gleason pattern 4/5) cancer. In the abstract and executive summary of this systematic review, the authors concluded that most of the literature assessing novel prognostic factors and models is poor and has significant methodological issues, including limited or missing validation procedures. However, of particular note, the authors singled out PSA velocity as having particularly strong evidence and significant hazard ratios (HR 9.8 for pre-treatment and HR 12.8 for post-prostatectomy clinical scenarios). However, the use of PSA velocity may be limited to active surveillance patients and cannot likely be used in risk stratification systems due to variability in testing frequency and number (i.e., patient may be treated after only one value and thus a PSA velocity cannot be calculated). Regarding the amount of high-grade cancer, Nanda and colleagues demonstrated that patients with Gleason score 7 with tertiary pattern 5 had similar biochemical outcome to patients with Gleason score 9-10. This confirmed the observation that the amount of high-grade cancer is an important prognostic factor to consider.

**Alternative risk stratification systems**

A variety of alternative risk stratification systems, with the occasional direct comparison of models, have been described. The literature primarily consists of variations of the existing risk stratification groups, but also includes score-based and nomogram-based systems. Two pre-treatment risk stratification models were evaluated in conjunction with the HTA report. Cowen and colleagues created a risk nomogram to predict overall survival at 5, 10 and 15 years and median survival. Nomogram components included various patient (age, Charlson score, blood pressure, smoking status, marital status, body mass), biochemical (initial PSA), tumour (T stage and Gleason score) and treatment (surgery vs. other) factors. This model was validated by comparison with other available prediction tools with an overall agreement statistic (C-statistic) of 0.73. Advantages of this model include the use of classical factors and strong methodology and validation; however, novel cancer-related factors were not included. Han and colleagues retrospectively performed a multivariable analysis on over 2000 radical prostatectomy patients to develop nomogram and table-based models to predict for 3, 5, 7 and 10-year biochemical recurrence free-survival. This investigation assessed all three classical factors including an assessment of the different primary and secondary patterns of Gleason 7 biopsies. This investigation confirmed the importance of all three classical factors with modification of the Gleason score to include two Gleason 7 categories in the prediction of biochemical outcomes (final model included initial PSA, T stage and Gleason score – 5, 6, 3+4, 4+3, vs. 8-10). However, this investigation was limited by several factors including restriction to prostatectomy patients, retrospective data collection, no overall survival analysis and no validation or measure of performance.
Additional investigations have assessed two alternative risk stratification systems. First, a previously validated four-group risk stratification system, as developed by the Radiation Therapy Oncology Group (RTOG) with a low-risk cohort (T1-T2 and GS2-6 with no PSA cut-offs) and three progressively higher-risk cohorts (RTOG 2 – T1T2GS7 or T3GS2-6; RTOG 3 – T1-T2GS8-10 or T3GS7; and RTOG 4 – T3GS8-10). Secondly, Roach and colleagues subsequently reported on an adapted version of this model for intermediate- and high-risk prostate cancer with integration of the Kattan nomogram for the prediction of metastatic disease. This investigation demonstrated further possible risk stratification particularly in RTOG groups 2 and 3 by the use of three Kattan risk groups (<8.5%, 8.5-15% and >15% predicted risk of metastatic disease). Although the use of this validated four-group risk stratification is potentially useful for clinical trial stratification and design, Roach and colleagues did not propose any specific alteration to the four-group-based system that was previously developed by the RTOG.

Williams and Beasley have published validation analyses regarding a five-group risk stratification system including low (PSA <7.5, Gleason score ≤6), low-intermediate (PSA 7.5-15, Gleason score ≤6), high-intermediate (PSA 15-20, Gleason score ≥6 or PSA ≤10, Gleason score ≥7), high (PSA 20-30, Gleason score ≥6 or PSA 10-20, Gleason score ≥7), and extreme (PSA >20, Gleason score ≥7 or PSA >30, Gleason score ≥6) groups. In the initial assessment and validation of this stratification system and corresponding nomogram, a comparison of this new five-group recursive partitioning analysis based system was compared to the contemporary NCCN three-group risk stratification system and the Kattan nomogram. This comparison was done to predict biochemical failure based on the definition by the American Society of Therapeutic Radiology and Oncology (ASTRO) as PSA nadir plus 2 ng/mL. This five-group system demonstrated superior discrimination ability when compared to the other systems; however, there was still some difficulty in the misclassification of the intermediate-risk groups. Additionally, the confounding interaction of hormonal therapy with biochemical (and other outcomes) was not assessed in this initial model building and validation exercise. In follow-up to this investigation, Beasley and colleagues assessed the interaction of hormonal therapy and the five-group model to determine whether any subgroups would benefit from combined hormonal therapy with external-beam radiotherapy. This investigation confirmed the benefit of hormonal therapy in the high-intermediate, high and extreme groups of the stratification system that suggests that current risk stratification systems should consider modification of intermediate-risk into two groups.

Discussion

Options for implementing changes in risk stratification systems

Existing risk stratification systems should ideally be modified based on the evidence-based information within a consensus approach to demonstrate that a change is warranted based on risk and clinical management considerations. Proposed changes should be consistent with management practices and ideally should be coherent with older/current risk stratification systems to be able to compare outcomes over time. If a new prognostic factor is introduced, literature describing the implementation and operating characteristics of this new prognostic factor should be well-defined and reliable so that they can be easily reproduced by clinicians, have demonstrated internal and external validity, and ideally be easily attainable with acceptable cost.

A critical review of the literature has identified two specific potential changes to the existing three-group risk stratification system that would be clinically relevant. First, analogous to the recent update to the NCCN guidelines, the inclusion of a very-low risk category to identify patients that may be entered into observation/surveillance protocols should be considered. Secondly, intermediate-risk prostate cancer as it is currently defined has been demonstrated to be a heterogeneous group of patients in terms of outcome and treatment management. A division of the intermediate-risk group into a low-intermediate risk group and a high-intermediate risk group is appropriate given the observation of hormonal therapy benefit for the higher-risk group as described by Beasley and colleagues. However, the authors’ five-group stratification system did not integrate factors, such as the amount of high-grade cancer, Gleason pattern 4+3 versus 3+4 and percentage positive biopsy cores. Ideally, these factors should be considered for inclusion in any possible changes to the existing risk-stratification systems.

In the ongoing RTOG 0815 intermediate-risk prostate cancer trial assessing radiation therapy with short-term hormonal therapy versus no hormonal therapy, patients with Gleason 7, PSA 10-20, T2b/c disease, as well as ≥ 50% biopsy cores, are not considered intermediate-risk in this clinical trial. The definition of the interface between intermediate- and high-risk disease may also need to consider published evidence by Nanda and colleagues showing that patients with tertiary pattern five have biochemical outcomes similar to Gleason score 9-10 patients. Current trends in multi-institutional clinical trial designs already reflect these developments; therefore, risk stratification systems need to be periodically reassessed and updated to reflect the available evidence, current clinical practice and modern clinical trial design.

The GUROC will undertake a project to create a multi-institutional database to delineate the appropriate defini-
tions of low-intermediate, high-intermediate and high-risk prostate cancer within a five-group risk stratification system. Candidate factors to be assessed in conjunction with this database project will be new PSA cut-off values (15 ng/mL), amount of high-grade cancer, Gleason pattern 3+4 versus 4+3, percentage positive biopsy cores and T stage (i.e., presence of T2b/c disease). Available Canadian databases considered for combined analysis will include the British Columbia Cancer Agency prostate cancer database, the National Cancer Institute of Canada PR5 intermediate-risk dose (and dose-per-fraction) fractionation study, as well as the Princess Margaret Hospital prostate cancer dose-escalation and brachytherapy clinical databases. Ultimately, validation of the five-group risk stratification system will need to occur and may utilize maturing or currently accruing clinical trial databases, including those from the RTOG 9910 and 0126 and from the Ontario Clinical Oncology Group PROFIT (Prostate Fractionated Irradiation Trial) study assessing 78 Gy/39 fractions versus 60 Gy/20 fractions for intermediate-risk prostate cancer.

Conclusion

Multiple studies of prognostic factors have been performed to create currently utilized prostate cancer risk stratification systems. Subsequent studies have evaluated novel prognostic factors and alterations to the traditional three-group risk stratification systems; however, a coherent set of investigations to advise organizations on an appropriate strategy to alter existing risk stratification systems is lacking. Ultimately, changes will likely be driven by consensus-based approaches linked to current practice patterns (surveillance, short-term and long-term hormonal therapy integrated with radiation therapy) and informed by the available evidence. Ongoing evaluation of important biochemical and clinical outcomes before and after any changes will be vital to assess any implemented changes.

Competing interests: None declared.

This paper has been peer-reviewed.

References

Risk stratification in prostate cancer


Correspondence: Dr. George Rodrigues, 790 Commissioners Rd E, London, ON N6A 4L6; george.rodrigues@lhsc.on.ca