Title
Prostate cancer risk assessment

Permalink
https://escholarship.org/uc/item/3jf4t6pp

Journal
Cancer, 113(11)

ISSN
0008-543X

Author
Cooperberg, Matthew R

Publication Date
2008-12-01

DOI
10.1002/cncr.23920

Peer reviewed

Prostate Cancer Risk Assessment: Choosing the Sharpest Tool in the Shed

Matthew R. Cooperberg, MD, MPH

University of California, San Francisco, Box 1695, 1600 Divisadero St, A-607, San Francisco, CA 94143-1695, tel (415) 885-3660, fax (415) 885-7443

Matthew R. Cooperberg: mcooperberg@urology.ucsf.edu

Abstract

Prostate cancer risk assessment has become increasingly complex, with a long and growing menu of risk assessment tools available. Instrument selection for clinical practice and research must balance accuracy, validation, and applicability.

Keywords

prostate neoplasms; risk assessment; nomograms; reproducibility of results; prognosis

Introduction

This year in the United States, 28,660 men are expected to die due to prostate cancer, a mortality figure among men surpassed only by lung cancer, yet dwarfed by the 186,320 expected new diagnoses.[1] Most men diagnosed with prostate cancer will ultimately die of other causes, and the natural history of the disease is relatively protracted even in cases which are eventually lethal. Given this frequently indolent tumor behavior and the potential toxicity of all available treatments,[2] clinicians and researchers alike increasingly recognize the importance of risk stratification of prostate cancer patients, and by extension the adoption of risk-adapted treatment strategies. Even younger patients with lower risk disease are now eligible for trials of active surveillance at a growing number of institutions, those with low- to intermediate-risk disease may be treated with local monotherapy, those with intermediate- to high-risk disease should in many cases receive multimodal therapy, and those at highest risk should ideally be offered enrollment in clinical trials given the high rates of recurrence with current standard therapies.[3]

While this basic framework is supported by a broad consensus of prostate cancer experts, the question of how exactly patients should be risk stratified, in clinical practice and for research trials, has become increasingly complex. In 2001, Ross et al published a literature review identifying 42 risk prediction tools for prostate cancer; since that time, the field has expanded greatly. In this issue of *Cancer*, Shariat et al summarize 109 nomograms and other risk prediction instruments, covering multiple decision points in prostate cancer care: prediction of positive biopsy with or without a prior negative biopsy, prediction before surgery of pathological outcomes, prediction before and after surgery of biochemical endpoints, prediction before radiation therapy of biochemical and clinical endpoints, and prediction of metastases and survival among patients with recurrent disease after primary treatment.

Of note, the term "nomogram" denotes a graphical representation of a predictive formula, not the formula itself. Nomograms in oncology are generally derived from Cox proportional hazards regression analyses—including, in many instances in the case of prostate cancer, a cubic spline or other transformation of the prostate specific antigen level—the results of which can be represented in a variety of ways. Nomograms are one popular option; alternatives include lookup tables[4] and categorized point systems.[5] Readily available statistical software has made generation of a nomogram based on a Cox model relatively trivial. The result is a profusion of nomograms—there were eleven new prostate cancer nomograms introduced at the 2008 meeting of the American Urological Association alone—which can be confusing to clinicians, researchers, and patients alike. Which should be used in clinical practice or in research: the "latest and greatest" or the "tried and true"? Which are the sharpest tools, and which are best suited to improve treatment decision-making?

Evaluating risk instruments

Shariat et al discuss some of the key features of predictive tools relevant to their implementation, including accuracy, calibration, generalizability, and parsimony—though they do not evaluate most of the instruments subsequently reviewed by these criteria. Aspects of the criteria merit further discussion. Shariat et al consider validation in the context of accuracy and calibration; it is also a critical aspect of generalizability. Many risk instruments are developed based on patients treated at one or a few, usually high-volume, academic institutions, by a relatively small number of clinicians under uniform protocols. Many distinct nomograms are based, in fact, on updated assessment of a few large, previously analyzed cohorts.

An instrument's performance in different settings among other patient groups cannot be assumed to be equivalent, for which reason newly developed risk instruments must be subjected to validation studies. Such studies ideally should be performed with cohorts of patients external to the center which developed the instrument, and in the best case the analysis should be performed by researchers entirely independent of the original center. Few instruments have in fact been evaluated to this level of scrutiny; examples include the Partin tables,[6] the original preoperative and postoperative Kattan nomograms,[7, 8] and the UCSF Cancer of the Prostate Risk Assessment (CAPRA) score,[9, 10] which have been validated externally and independently, in both American and European cohorts.

Risk prediction tools tend to perform at higher levels of accuracy among academic cohorts compared to community-based cohorts, likely for the reasons of uniformity of practice in the academic setting noted above. Thus, for example, the original Kattan preoperative nomogram performs with a c-index between 0.74 and 0.79 in academic cohorts,[10-12] but somewhat lower at 0.68 in a community-based cohort.[13] Conversely, the CAPRA score, which yielded a c-index of 0.66 in the original community-based development studies,[5] performed with a c-index as high as 0.81 in academic center-based validation studies.[9, 10] A final point with respect to validation is that because distribution and interpretation of prostate cancer risk factors, particularly Gleason grading, has changed significantly over time with artifactual changes in pathologists' grading practices,[14] it is important that the relevance of a given instrument to contemporary patients be periodically verified.

Parsimony, also called level of complexity in the review by Shariat et al, also warrants additional comment under the more general rubric of applicability. Unlike accuracy and validation, parsimony and applicability are qualitative rather than quantitative criteria, and can be considered separately for clinical practice and research. A clinically applicable instrument should be readily understood by clinicians and patients, and rapidly calculable during a patient encounter. An instrument applicable for risk stratification in clinical research, in turn, must be

easily calculable for hundreds or thousands of patients, and must be able to stratify them to a manageable number of risk strata.

Risk instruments have come to incorporate increasing numbers of variables over time, such that the paper nomograms require up to eight steps to use, which can be somewhat cumbersome in practice. For this reason, and given the growing number of instruments applicable to practice, web-based calculators for selected instruments have been made available as noted in the Shariat et al's review. These computer-based tools are easier to use than the paper tables, but still require navigation of several screens for each patient. The web calculators also are somewhat of a black box, as the derivations of the nomogram scores are not obvious to users or patients. Indeed, our group developed the CAPRA score specifically in an effort to balance the accuracy of a instrument based on a Cox model with the applicability of a scoring system based on simple addition and requiring neither paper tables nor software.[5]

Another potential problem with the software approach is that the calculators run multiple nomograms simultaneously, which creates a temptation to give an individual patient comparative likelihoods of outcomes following both surgery and radiation—the same temptation faces a patient who accesses the website directly. The nomograms cannot fairly be used this way; each was derived with a different cohort of patients, using different definitions of recurrence. Kattan et al, in fact, explicitly stated in the original preoperative nomogram paper that the tool was meant to be used only among men who have already elected radical prostatectomy,[12] not among those considering treatment alternatives. The instructions for the original nomogram also specified that the calculated likelihood of recurrence should be presented to the patient with ± 10% error[12]; inclusion of the confidence intervals is indeed important, but is generally omitted both from more recent paper nomograms and from the web-based calculators. Finally, for research purposes, the web calculators would be nearly as difficult as the paper nomograms to calculate for large numbers of patients.

Prediction of distal endpoints and across treatments

Most nomograms and other instruments designed to predict cancer outcomes following treatment have been evaluated for prediction biochemical outcomes following prostatectomy or radiation therapy. However, consistently defining biochemical recurrence is notoriously challenging—by one recent count 152 different definitions were used in studies published between 1991 and 2004—53 in prostatectomy series and 99 in radiation series.[15] Moreover, biochemical recurrence does not consistently predict more distal clinical endpoints, such as metastasis, cancer-specific mortality (CSM), and all-cause mortality (ACM).[4]

These clinical endpoints may be defined more consistently, and are more relevant to patients than biochemical outcomes. Relatively few pre-treatment instruments reported to date have been evaluated at these endpoints, however. Exceptions include the D'Amico 3-level classification, which was shown to predict CSM after surgery or radiation[16]; a nomogram by Kattan et al which predicts metastasis following external-beam radiation therapy,[17] the CAPRA score, recently demonstrated to predict metastasis, CSM, and ACM after surgery, radiation therapy, or androgen deprivation therapy[18]; and novel nomograms introduced by Stephenson et al to predict CSM at 15 years following prostatectomy[19] and Duan et al to predict ACM at 10 years following prostatectomy or radiation therapy.[20]

An ideal instrument would be able to risk stratify patients regardless of primary treatment choice; a patient with more aggressive disease features faces increased risk of recurrence regardless of treatment choice, and an instrument should covey a measure of relative risk independent of treatment modality. Pooling or comparing patients undergoing prostatectomy with those receiving radiation therapy is difficult in studies with biochemical endpoints due to the difficulties in combining definitions of biochemical failure applied to patients with different

primary treatments. Furthermore, as noted above, most published nomograms are intended to be used among patients who have already made a treatment selection. Notable exceptions, mentioned above, include the D'Amico classification (tested for prostatectomy and radiation patients),[16] the CAPRA score (prostatectomy, radiation, and androgen deprivation patients), [18] and the nomogram described by Duan et al (prostatectomy and radiation patients).[20]

Conclusions and future directions

The only risk assessment instrument endorsed by the American Urological Association's 2007 practice guideline for prostate cancer is the 3-level D'Amico classification.[3] This classification system has been used extensively in practice and research, and certainly excels in terms of parsimony and ease of use. More recent multivariable instruments offer improved accuracy and discrimination, as reviewed by Shariat et al, yet most are significantly more difficult to use, and the plethora of novel instruments runs the risk of producing more confusion than illumination. An open question, in fact, is to what extent—if any—these various tools are being widely adopted in routine contemporary practice.

Changes could be made in the reporting of novel nomograms which may result in easier interpretation and implementation. Confidence intervals should be included with the tools as was done with original Kattan preoperative nomogram.[12] It would be useful to report the predictive formula derived from the Cox model along with the nomogram itself, allowing investigators interested in testing, validating, or adopting the nomogram to incorporate it into statistical programs with relative ease. This change would facilitate the head-to-head comparisons among instruments which Shariat et al appropriately characterize as the best means of objectively ascertaining which have the best performance characteristics. Indeed, a case can be made that at this point so many nomograms have been put forward that any new model should be compared explicitly to existing instruments intended for the same clinical scenario (e.g., prior to prostatectomy), in order that the incremental value of the new tool can be assessed objectively. To date, some papers reporting new nomograms have done this, while the majority have not.

Finally, the question of risk stratification in the research setting should be addressed more formally. Shariat et al rightly advise that risk prediction tools be used more widely in clinical trial design. The score derived from most nomograms, however, is a likelihood of reaching the endpoint in question, expressed on a 100 point scale. This output is well suited for counseling an individual patient, but results in an impractical number of strata for research purposes. Thus, if a nomogram is to be used to identify a cohort of high risk patients for a given protocol, for example, limits need to be chosen to determine which scores correspond to high risk. Absent clear *a priori* establishment and validation of these thresholds, different thresholds might be used for each study—or worse, could be chosen post hoc to choose subgroups to suit the investigators' purposes. Grouping of scores has been validated for the CAPRA score[9, 10] but not for most nomograms, which may be a barrier to their broader use in the research setting.

Clearly, no risk instrument should be used in isolation to direct patients to or away from treatment alternatives. Factors not measured by current models, such as baseline quality of life, comorbidity, and life expectancy are all considered qualitatively in decision-making. At least as important are patient treatment preferences, both for immediate vs. deferred treatment and for one form of treatment over another. In fact, the clarity with which patients can understand a risk prediction tool's inputs and results, and thus the extent to which the tool facilitates their decision-making, is another important—and generally overlooked—consideration in tool selection.

With numerous prostate cancer biomarkers working their way from bench to bedside, it seems reasonable to hope that the next few years will see novel markers sufficiently validated to be integrated into other risk prediction tools. In the interim, clinicians and researchers are faced with the question of which clinically-based instruments to use and in what settings. Near the end of their review, Shariat et al call for the development of more nomograms. Equally important will be applying consistent criteria for evaluating both existing and novel nomograms, to determine which are in fact the best tools for the job. Good accuracy and calibration are critical—the tools need to be sharp. What is less clear is what increment in accuracy, as measured by the c-index or other metric, is clinically meaningful. Assuming comparable accuracy among alternatives, the best tools for clinical practice and research should be those which are reliable and easy to handle—that is, those which have been most extensively validated, are easiest to calculate, and make both effective and efficient use of the available clinical data.

Acknowledgments

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. CA Cancer J Clin Mar-Apr;2008 58(2):71–96. [PubMed: 18287387]

2. Wilt TJ, MacDonald R, Rutks I, Shamliyan TA, Taylor BC, Kane RL. Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. Ann Intern Med Mar 18;2008 148(6):435–448. [PubMed: 18252677]

3. American Urological Association Prostate Cancer Clinical Guidelines Panel. Guideline for the management of clinically localized prostate cancer: 2007 update. American Urological Association Education and Research, Inc.; 2007.

4. Freedland SJ, Humphreys EB, Mangold LA, et al. Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. JAMA Jul 27;2005 294(4):433–439. [PubMed: 16046649]

5. 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 Jun;2005 173(6):1938–1942. [PubMed: 15879786]

6. Penson DF, Grossfeld GD, Li YP, Henning JM, Lubeck DP, Carroll PR. How well does the Partin nomogram predict pathological stage after radical prostatectomy in a community based population? Results of the cancer of the prostate strategic urological research endeavor. J Urol Apr;2002 167(4): 1653–1657. discussion 1657-1658. [PubMed: 11912382]

7. Graefen M, Karakiewicz PI, Cagiannos I, et al. A validation of two preoperative nomograms predicting recurrence following radical prostatectomy in a cohort of European men. Urol Oncol Jul-Aug;2002 7 (4):141–146. [PubMed: 12474529]

8. Graefen M, Karakiewicz PI, Cagiannos I, et al. Validation study of the accuracy of a postoperative nomogram for recurrence after radical prostatectomy for localized prostate cancer. J Clin Oncol Feb 15;2002 20(4):951–956. [PubMed: 11844816]

9. May M, Knoll N, Siegsmund M, et al. Validity of the CAPRA score to predict biochemical recurrence-free survival after radical prostatectomy. Results from a European multicenter survey of 1,296 patients. J Urol 2007;178(5):1957–1962. [PubMed: 17868719]

10. Zhao KH, Hernandez DJ, Han M, Humphreys EB, Mangold LA, Partin AW. External Validation of University of California, San Francisco, Cancer of the Prostate Risk Assessment Score. Urology. Mar 25;2008 10.1016/j.urology.2007.11.165

11. Graefen M, Karakiewicz PI, Cagiannos I, et al. International validation of a preoperative nomogram for prostate cancer recurrence after radical prostatectomy. J Clin Oncol Aug 1;2002 20(15):3206–3212. [PubMed: 12149292]

12. Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J Natl Cancer Inst May 20;1998 90(10):766–771. [PubMed: 9605647]

13. Greene KL, Meng MV, Elkin EP, et al. Validation of the Kattan preoperative nomogram for prostate cancer recurrence using a community based cohort: results from Cancer of the Prostate Strategic Urological Research Endeavor (CaPSURE). J Urol Jun;2004 171(6, Part 1):2255–2259. [PubMed: 15126797]

14. Cooperberg MR, Lubeck DP, Mehta SS, Carroll PR. Time trends in clinical risk stratification for prostate cancer: implications for outcomes (data From CaPSURE). J Urol Dec;2003 170(6 Pt 2):S21–S27. [PubMed: 14610406]

15. Cookson MS, Aus G, Burnett AL, et al. Variation in the definition of biochemical recurrence in patients treated for localized prostate cancer: the American Urological Association Prostate Guidelines for Localized Prostate Cancer Update Panel report and recommendations for a standard in the reporting of surgical outcomes. J Urol Feb;2007 177(2):540–545. [PubMed: 17222629]

16. D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, Chen MH. Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. J Clin Oncol Jun 1;2003 21(11):2163–2172. [PubMed: 12775742]

17. Kattan MW, Zelefsky MJ, Kupelian PA, et al. Pretreatment nomogram that predicts 5-year probability of metastasis following three-dimensional conformal radiation therapy for localized prostate cancer. J Clin Oncol Dec 15;2003 21(24):4568–4571. [PubMed: 14673043]

18. Cooperberg MR, Broering JM, Carroll PR. The UCSF Cancer of the Prostate Risk Assessment score accurately predicts metastasis, prostate cancer mortality, and all-cause mortality. J Urol 2008;179 (4S) Abstract 321.

19. 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 Urol 2008;179(4S) Abstract 1889.

20. Duan JJ, Piccirillo JF, Vlahiotis A, Luly JR, Kibel AS. Preoperative nomogram for overall 10-year survival after radical prostatectomy or radiation therapy for prostate cancer. J Urol 2008;179(4S) Abstract 1437.