EUROPEAN UROLOGY 76 (2019) 268–272

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Platinum Opinion

The State of the Science on Prostate Cancer Biomarkers: The San Francisco Consensus Statement

Matthew R. Cooperberg [a,b,*], *Peter R. Carroll* [a], *Marc A. Dall'Era* [c], *Benjamin J. Davies* [d], *John W. Davis* [e], *Scott E. Eggener* [f], *Felix Y. Feng* [a,g], *Daniel W. Lin* [h], *Todd M. Morgan* [i], *Alicia K. Morgans* [j], *Daniel E. Spratt* [k], *Samir S. Taneja* [l], *David F. Penson* [m,n]

[a] *Department of Urology, UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA;* [b] *Department of Epidemiology & Biostatistics, University of California, San Francisco, CA, USA;* [c] *Department of Urology, University of California, Davis, Sacramento, CA, USA;* [d] *Department of Urology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA;* [e] *MD Anderson Cancer Center, Houston, TX, USA;* [f] *Department of Surgery, University of Chicago, Chicago, IL, USA;* [g] *Department of Radiation Oncology, University of California, San Francisco, CA, USA;* [h] *Department of Urology, University of Washington and Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;* [i] *University of Michigan Department of Urology, Ann Arbor, MI, USA;* [j] *Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA;* [k] *Department of Radiation Oncology, University of Michigan School of Medicine, Ann Arbor, MI, USA;* [l] *NYU Langone Health, New York, NY, USA;* [m] *Department of Urology, Vanderbilt University School of Medicine, Nashville, TN, USA;* [n] *Tennessee Valley Health Care System, Nashville, TN, USA*

Introduction

At every point along the prostate cancer continuum of care, from early detection through to advanced disease, a man must face decisions that may have a direct impact on his long-term outcomes. These decisions are increasingly complex as an ever-growing array of management options emerge, and highly complex data drive both individual patient care and clinical studies. Consequently, much uncertainty remains around how to optimize decisions for a given patient. At the root of the challenge for men diagnosed with prostate cancer is prostate cancer's profound clinical heterogeneity, ranging from entirely indolent to rapidly progressive disease. Risk stratification based on standard clinical and pathologic features often has suboptimal performance; for example, the ability to discriminate patients who will develop metastatic disease is usually only 0.5–0.7 [1,2] (0.5 being equivalent to a coin flip). Robust multivariable clinical models can increase the performance to 0.65–0.80 depending on the cohort analyzed and clinical scenario, but even then there remains prognostic uncertainty, especially on the extremes of the risk spectrum [2,3]. The degree to which a patient's prostate biopsy represents the overall tumor biology remains a concern for patients and physicians, and is likely a driver of discordant care such as overtreatment of low-risk disease or undertreatment of higher-risk disease.

The available set of tissue, blood, urine, and imaging biomarkers has exploded in recent years, and well over a dozen different tests now compete in this tightening marketplace. Several are analytically and clinically validated to improve on clinical parameters alone by more accurately estimating prognosis and specific oncologic outcomes [2,4–8]. Most such studies have confirmed markers' independent *prognostic* value. Whether a marker may be *predictive* of a response to a given management strategy is a different and more complex question. A prognostic marker may help a man with "intermediate-risk" cancer better estimate his likelihood of cancer progression; a predictive one would tell that man whether his cancer is more or less likely to respond to, for example, radiation therapy (RT).

Retrospective evidence suggests that these markers may guide management decisions to intensify or deintensify therapy, but none to date has been validated prospectively, in a clinically meaningful context, to serve as a true predictor of response to any specific therapy. Furthermore, the optimal clinical decision point, interpretation, and incorporation of these biomarkers into clinical decision

* Corresponding author. 550 16th St, Box 1695, San Francisco, CA 94143, USA. Tel. +1 415 885 3660.
E-mail address: matthew.cooperberg@ucsf.edu (M.R. Cooperberg).

https://doi.org/10.1016/j.eururo.2019.05.013
0302-2838/Published by Elsevier B.V. on behalf of European Association of Urology.

making remain incompletely defined; and available markers tend to be very expensive, especially considering the dearth of prospective evidence. With the goals of defining the current state of science and high-priority objectives for near-term research, we convened an independent, multidisciplinary panel of prostate cancer biomarker thought leaders in urologic, radiation, and medical oncology, concurrent with the 2018 American Urological Association annual meeting in San Francisco. Key questions were proposed and agreed upon via iterative e-mail discussions prior to the meeting, and no external entity provided any support for or input into the panel's deliberations. The following summarizes the key questions and discussion by the panel.

What are the potential clinical scenarios after prostate cancer diagnosis in which a biomarker could have utility?

Biomarkers may potentially improve risk stratification across a range of critical prostate cancer decision points: whether to undergo an initial biopsy, whether to undergo a repeat biopsy after a prior negative biopsy, whether to manage a favorable-risk cancer with active surveillance or definitive treatment, how to select or combine treatments for intermediate- and high-risk localized disease (eg, addition or duration of androgen deprivation therapy to RT) and after prostatectomy (eg, addition of adjuvant RT), and how to modulate intensity of follow-up surveillance. Currently, available markers have been well validated for some, but not all, of these clinical states. What remains unclear, however, is how to *optimize* marker use in clinical practice. Just because a marker *may* apply to a given patient's situation does not necessarily mean that the patient *should* undergo testing. As with any clinical intervention, a given test should be used only when the result may realistically be expected to change management and/or improve clinical outcomes.

Guidelines are consistent in asserting that most men with a low-volume grade group 1 (Gleason 3 + 3 = 6) tumor and a relatively low prostate-specific antigen (PSA) level should undergo active surveillance. Although undersampling occurs, this fact does not mean that every man necessarily needs immediate additional testing [9,10]. Given the excellent results of low-risk patients in active surveillance series that do not utilize genomic biomarkers or magnetic resonance imaging (MRI), the bar to prove that these tests improve long-term, clinically meaningful outcomes for these patients is quite high.

Appropriate endpoint selection is crucial. Each gene expression classifier was developed and initially optimized to predict a different endpoint. By now, many markers have been shown to predict multiple endpoints, including adverse pathology, biochemical recurrence, metastasis-free survival, and/or cancer-specific mortality. However, one endpoint is not a clear surrogate for another. Most men with recurrence after treatment do not die of disease, with or without additional treatment, and given the prolonged natural history of prostate cancer, it can be difficult to ascertain how the possibility of adverse pathology at prostatectomy influences whether a man with an otherwise low-risk tumor should or should not proceed directly to treatment. The temporal associations of most markers with outcomes—and how wide the windows of opportunity are for effective intervention—remain obscure and will require additional, likely prospective, studies.

Now that a growing number of markers have been shown to improve risk stratification retrospectively, prospective studies are necessary to validate their use in decision making (ie, utility), but perhaps more importantly, to document improvement in clinically meaningful outcomes (ie, decreased morbidity from disease and/or from treatment). Prediction of adverse pathology may be a useful intermediate outcome but ultimately is insufficient, as a marker must be able to identify men in whom we can meaningfully reduce their intensity of surveillance, or alter their management to improve outcomes in regard to tumor control and/or quality of life. Currently, biomarker-based decisions are heterogeneous and often based on unvalidated risk thresholds, further warranting the need for prospective validation.

To what clinical standards should biomarkers be compared, and how should markers be combined with other clinical parameters, imaging tests, and prognostic instruments?

A critical principle of biomarker research—often overlooked—is that any putative marker (or imaging test) must be shown to add value above and beyond an existing *multivariable*, ideally linear, clinical standard. Thus, a prediagnostic test should improve not only on PSA (free and total, since both are inexpensive and widely available), but also on a validated model reflecting PSA, age, family history, race/ethnicity, previous biopsy results, etc. A postdiagnosis test, likewise, must improve not only on Gleason score or risk groups, but also on a validated, truly multivariable model that incorporates multiple known clinical risk factors. Clinical risk stratification without biomarkers can be quite accurate depending on the context, and novel biomarkers must be shown to add value to the existing, already obtained information. Additionally, clinical models are variably accurate depending on a patient's disease risk. Low-risk patients very rarely develop metastatic disease, and thus it is hard to improve prognostic models beyond what is already known for this population. In contrast, about 10–30% of high-risk patients will develop metastatic disease, and clinical models are not able to sufficiently discriminate between those who will or will not develop metastatic disease.

No consensus exists regarding what incremental improvement in accuracy should be considered "adequate." Changes in area under the curve or c-index can be modest, especially when the cohort reflects a restricted clinical risk range. Decision curve analysis [11] can be more illustrative of the range of clinical threshold probabilities for intervention over which a test may be useful. Ultimately, the balance between

improvement in accuracy or discrimination and the cost for a given test remains unclear, and accuracy improvement is the only criterion by which clinical utility should be judged.

Optimal integration of biomarkers together with imaging tests—in particular multiparametric MRI (mpMRI)—is also not well defined. Logic dictates that where a tumor is visible on either ultrasound or mpMRI, the lesion should be targeted to obtain samples for potential biomarker analysis, although generally speaking, the sample with the highest Gleason grade is the one profiled, whether taken from a targeted or a systematic core. Multiparametric MRI is increasingly available and widely used in the USA and other health care systems, but is performed and interpreted to highly variable quality. Recent analyses show that Prostate Imaging Reporting and Data System version 2 is not sufficient to assure consistency across radiologists [12,13]. The mean cost of MRI in the USA is $2500, which varies widely depending on region and may even exceed that of tissue-based biomarkers [14]. Which endpoints are best predicted by MRI and which by other biomarkers, and in which situations markers and imaging should be sequenced or combined, remain unanswered questions and high priorities for future research.

Which endpoints should be assessed in studies investigating the value of a candidate marker?

Ultimately, the use of a biomarker—as with any other intervention—must lead to improved quantity and/or quality of life through better tailoring of treatment timing and intensity. Biomarkers have been tested for a wide range of outcomes, including adverse pathology at prostatectomy, biochemical recurrence after treatment, distant metastasis, cancer-specific survival, and overall survival. None of these is perfect; near-term endpoints imperfectly predict ultimate mortality, yet mortality endpoints require decades of follow-up. Furthermore, it can be difficult or impossible to base a decision on, for example, an absolute mortality difference of 2%—yet a 10-yr mortality probability of 2% versus 4% represents a typical range for favorable-risk patients and a doubling of absolute risk.

Regardless of the specific endpoint, a useful marker must be proved in the short term to drive changes in decisions and, more importantly, in the long term to drive better outcomes. In an era of value-driven health care, simply providing ill-defined reassurance or confidence reinforcing a decision already supported by multiple guidelines (eg, initial selection of active surveillance for a low-grade tumor) is insufficient. Furthermore, most of the extant literature on the "clinical utility" of biomarkers and imaging tests tends to be of low quality, prone to multiple biases, and lacking in meaningful clinical outcomes.

Which study designs should be used to establish prognostic value and clinical utility?

Currently, most commercial tests are being developed by US companies, and multiple major payers in the USA make decisions about reimbursement for biomarkers based on the criteria proposed by Simon et al. [15]. Under this framework, the "gold standard" design is a randomized, controlled trial (RCT) designed specifically to evaluate the marker. No such trials have been reported for any prostate cancer markers, although one (G-MINOR—NCT02783950) has recently completed accrual across the state of Michigan.

Simon et al.'s [15] criteria give second highest weight to studies in which biomarkers are assessed in RCTs that were designed for other purposes. Two such studies of a given marker are required for a "level I" designation. Nonrandomized studies, no matter how carefully designed, can at most support a "level II" designation. In this regard, the framework of Simon et al. [15] may be notably flawed; the biobanking component of an RCT addressing another clinical question may still be prone to bias, and the quality of specimen collection, archiving, and annotation is no less important than a priori randomization. Furthermore, clinical utility—that is, change in decisions—is often employed as the key criterion by the MolDx program in determining Medicare coverage; yet, these decisions may or may not be leading to better long-term outcomes. Other criteria such as REMARK [16] can provide a useful alternative framework for assigning levels of evidence and should be considered while making coverage decisions. Multiple studies assessing tissue-based biomarkers from prior RTOG/NRG randomized trials are underway, and select ongoing trials (NRG GU002 and NRG GU006) have incorporated prospective gene expression testing into the trial design.

Many (although not all) existing biomarker studies have been performed among men who are mostly Caucasian and often of relatively high socioeconomic status. Future validation studies must emphasize recruitment of diverse cohorts both to confirm biomarker performance among non-Caucasian men and to explore any potentially discernable biological differences across subpopulations.

More RCTs focused specifically on biomarkers are essential, and both public and industry funding should be focused on these. Multiple designs could be considered [17], and some may yield clinically meaningful endpoints in a reasonable time frame. While long-term endpoints are challenging, intermediate/surrogate endpoints can be informative, and the length of follow-up time should not be an excuse to forgo the critical studies needed to inform biomarker use in localized prostate cancer. Emerging prospective registries that collect data pervasively at the point of care may also prove invaluable as a means of collecting credible data on patterns of care and decisions in the presence or absence of marker results.

How are biomarker results best expressed and shared with patients?

Establishing analytic and statistical validity is only part of the challenge in developing a test with true clinical utility. Equally important is the way in which the results, which are commonly complex, polygenic models, are distilled into reports for clinician or patient use. Commercial clinical

reports are sometimes designed with as much input from marketing divisions as from scientific teams, and in some cases, the score is presented in a format that has little to no relationship with the actual validation studies.

Many prostate cancer patients have poor numeracy, a problem that can acutely be exacerbated by the stress of a new diagnosis and the task of integrating multiple scores, percentages, and other figures into their understanding of their disease. Even clinicians routinely fail to understand complex numerical data adequately, and may be uncomfortable communicating results [18]. Incomplete understanding of risk through low health numeracy can contribute to poor medication adherence [19], potentially contributing to worse health outcomes [19]. Overly complex or poorly contrived reports may exacerbate these problems, with prostate cancer being particularly challenging as risk reflects a biological continuum rather than binary states of "high" and "low" risk. Most biomarkers (including PSA), therefore, cannot be expressed accurately as "positive" or "negative" based on single thresholds. Some may be best expressed with two thresholds: a low threshold that maximizes negative predictive value and a high threshold that maximizes positive predictive value, in an effort to reduce challenges with risk communication [20].

A robust literature identifies best practices with respect to presenting this type of information [21], and researchers and biomarker companies should consider partnering with communication experts to develop reports that are both accurate and easy for patients with varied educational backgrounds to understand. Regardless of the report format, clinicians ordering these tests must be able to discuss them in the context of the overall clinical picture with patients with a wide range of literacy and numeracy.

Concluding thoughts

Biomarker calculations and the statistical basis of their integration with clinical parameters, whether before or after diagnosis, should be made public to facilitate the independent validation studies that are essential for the long-term success of a marker. While commercial providers may find this proposition challenging in the short term, in the long run both companies and patients would benefit substantially from faster optimization and validation of next-generation markers. Biomarker cost and reimbursement can be quite arbitrary, and its value must be proved along with efficacy.

The authors of this consensus statement have led a substantial proportion of the investigations, which have driven the biomarker field to its current state. We summarize our conclusions and recommendations in Table 1. While many of our comments above may suggest skepticism about markers' relevance or value, substantial progress has been made in the past few years, and we are universally optimistic that appropriately designed studies with meaningful endpoints, next-generation iterations of current markers, and development of entirely novel tests will ultimately drive the proof of their optimal implementation. Much work remains to be done, but biomarkers have tremendous potential to optimize the care of men at risk for or diagnosed with prostate cancer.

Table 1 – San Francisco Consensus statements

1. To be clinically useful, a putative biomarker must be shown to improve on an existing, validated, multivariable model reflecting all available clinical information—not simply on a single variable or nonlinear risk grouping.
2. Biomarker study endpoints should be tailored to the disease state of interest, and should capture clinically meaningfully changes that ultimately impact a patient's quantity and/or quality of life.
3. Current postdiagnosis biomarkers were developed and validated mostly in men who underwent prostatectomy or other active treatments, complicating efforts to extrapolate these results to untreated, contemporary men.
4. Randomized controlled trials specifically designed to answer the key clinical questions are a critical and necessary companion to ongoing observational studies that use prospectively collected and carefully annotated biospecimens. More randomized trials of markers and imaging tests should be prioritized and funded, and must recruit diverse populations of men.
5. Clinicians ordering biomarker studies for men with prostate cancer must be prepared to discuss the results in the context of the individual's medical situation with patients from a wide range of health literacy and numeracy levels.

Conflicts of interest: Relevant financial disclosures (24 mo): Cooperberg: MDx Health (advisory board). Carroll: none. Dall'Era: Genomic Health (speaker), MDx Health (clinical trial). Davies: none. Davis: GenomeDx (research study). Eggener: OPKO, Genomic Health, MDx Health (research support). Feng: Blue Earth Diagnostics (advisory board). Lin: GenomeDx, Genomic Health, MDxHealth (research funding). Morgan: Myriad Genetics (advisory board, research funding), GenomeDx (research funding), MDx Health (research funding). Morgans: Seattle Genetics (honorarium). Spratt: Blue Earth (advisory boards). Taneja: MDx Health (study participant). Penson: none.

References

[1] Berlin A, Murgic J, Hosni A, et al. Genomic classifier for guiding treatment of intermediate-risk prostate cancers to dose-escalated image-guided radiotherapy without hormone therapy. Int J Radiat Oncol Biology Phys 2019;103:84–91.

[2] Spratt DE, Zhang J, Santiago-Jiménez M, et al. Development and validation of a novel integrated clinical-genomic risk group classification for localized prostate cancer. J Clin Oncol 2018;36:581–90.

[3] Cooperberg MR, Broering JM, Carroll PR. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. JNCI J National Cancer Inst 2009;101:878–87.

[4] Cooperberg MR, Simko JP, Cowan JE, et al. Validation of a cell-cycle progression gene panel to improve risk stratification in a contemporary prostatectomy cohort. J Clin Oncol 2013;31:1428–34.

[5] Klein EA, Cooperberg MR, Magi-Galluzzi C, et al. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. Eur Urol 2014;66:550–60.

[6] Erho N, Crisan A, Vergara IA, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS One 2013;8:e66855.

[7] Neste L, Hendriks RJ, Dijkstra S, et al. Detection of high-grade prostate cancer using a urinary molecular biomarker-based risk score. Eur Urol 2016;70:740–8.

[8] Parekh DJ, Punnen S, Sjoberg DD, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. Eur Urol 2015;68:464–70.
[9] Chen RC, Rumble BR, Loblaw AD, et al. Active surveillance for the management of localized prostate cancer (Cancer Care Ontario guideline): American Society of Clinical Oncology clinical practice guideline endorsement. J Clin Oncol 2016;34:2182–90.
[10] 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;199:683–90.
[11] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.
[12] Sonn GA, Fan RE, Ghanouni P, et al. Prostate magnetic resonance imaging interpretation varies substantially across radiologists. Eur Urol Focus. In press. https://doi.org/10.1016/j.euf.2017.11.010.
[13] Greer MD, Shih JH, Barrett T, et al. All over the map: An interobserver agreement study of tumor location based on the PI-RADSv2 sector map. J Magn Reson Imaging 2018;48:482–90.
[14] Kim S, Vickers AJ, Hu JC. Challenges in adopting level 1 evidence for multiparametric magnetic resonance imaging as a biomarker for prostate cancer screening. JAMA Oncol 2018;4:1663–4.
[15] Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. JNCI J National Cancer Inst 2009;101:1446–52.
[16] McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies. J Clin Oncol 2005;23:9067–72.
[17] Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. JNCI J National Cancer Inst 2010;102:152–60.
[18] Gramling R, Irvin JE, Nash J, Sciamanna C, Culpepper L. Numeracy and medicine: key family physician attitudes about communicating probability with patients. J Am Board Fam Pract 2004;17:473.
[19] Nandyala AS, Nelson LA, Lagotte AE, Osborn CY. An analysis of whether health literacy and numeracy are associated with diabetes medication adherence. HLRP Heal Lit Res Pract 2018;2:e15–20.
[20] Wei JT, Feng Z, Partin AW, et al. Can urinary PCA3 supplement PSA in the early detection of prostate cancer? J Clin Oncol 2014;32:4066–72.
[21] Hamstra DA, Johnson SB, Daignault S, et al. The impact of numeracy on verbatim knowledge of the longitudinal risk for prostate cancer recurrence following radiation therapy. Med Decis Making 2014;35:27–36.