European Association of Urology

Platinum Priority – Prostate Cancer – Editor's Choice
Editorial by Peter C. Albertsen on pp. 908–909 of this issue

askMUSIC: Leveraging a Clinical Registry to Develop a New Machine Learning Model to Inform Patients of Prostate Cancer Treatments Chosen by Similar Men

Gregory B. Auffenberg [a], Khurshid R. Ghani [b], Shreyas Ramani [c], Etiowo Usoro [c], Brian Denton [b,d], Craig Rogers [e], Benjamin Stockton [f], David C. Miller [b], Karandeep Singh [c,g,h,], for the Michigan Urological Surgery Improvement Collaborative*

[a] Department of Urology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; [b] Department of Urology, University of Michigan Medical School, Ann Arbor, MI, USA; [c] School of Information, University of Michigan, Ann Arbor, MI, USA; [d] Department of Industrial & Operations Engineering, University of Michigan College of Engineering, Ann Arbor, MI, USA; [e] Department of Urology, Henry Ford Health System, Detroit, MI, USA; [f] Lakeside Urology PC, St. Joseph, MI, USA; [g] Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA; [h] Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

Article info

Abstract

Background: Clinical registries provide physicians with a means for making data-driven decisions but few opportunities exist for patients to interact with registry data to help make decisions.

Objective: We sought to develop a web-based system that uses a prostate cancer (CaP) registry to provide newly diagnosed men with a platform to view predicted treatment decisions based on patients with similar characteristics.

Design, setting, and participants: The Michigan Urological Surgery Improvement Collaborative (MUSIC) is a quality improvement consortium of urology practices that maintains a prospective registry of men with CaP. We used registry data from 45 MUSIC urology practices from 2015 to 2017 to develop and validate a random forest machine learning model. After fitting the random forest model to a derivation cohort consisting of a random two-thirds sample of patients after stratifying by practice location, we evaluated the model performance in a validation cohort consisting of the remaining one-third of patients using a multiclass area under the curve (AUC) measure and calibration plots.

Results and limitations: We identified 7543 men diagnosed with CaP, of whom 45% underwent radical prostatectomy, 30% surveillance, 17% radiation therapy, 5.6% androgen deprivation, and 1.8% watchful waiting. The personalized prediction for patients in the validation cohort was highly accurate (AUC 0.81).

Conclusions: Using clinical registry data and machine learning methods, we created a web-based platform for patients that generates accurate predictions for most CaP treatments.

Patient summary: We have developed and tested a tool to help men newly diagnosed with prostate cancer to view predicted treatment decisions based on similar patients from our registry. We have made this tool available online for patients to use.

* Corresponding author. Department of Learning Health Sciences, University of Michigan Medical School, 300N. Ingalls Street, Ann Arbor, MI 48109, USA. Tel. +1 734 9361649; Fax: +1 734 6473914.
E-mail address: kdpsingh@umich.edu (K. Singh).

1.   Introduction

Physicians and other health care stakeholders are increasingly using clinical data registries to enhance their own decision-making and improve health care quality [1–3]. Recognizing their potential, many national medical organizations have invested heavily in clinical registries as a foundation for quality improvement efforts such as the American College of Surgery's National Surgical Quality Improvement Program and the American Urological Association's AQUA registry. The Centers for Medicare and Medicaid Services established the qualified clinical data registries program as part of the Physician Quality Reporting System in 2014 and further elevated registries in the 2015 Medicare Access and CHIP Reauthorization Act to provide a greater financial incentive for participation in registries, recognizing that they can serve as a unique tool "to foster improvement in the quality of care provided to patients" [4].

Although registry initiatives have gained prominence during a period notable for increasing focus on patient access to personal health information [5], there are few examples whereby patients can directly interact with clinical registries. The information in these large data repositories, if presented appropriately, may represent an important avenue for patients to find individualized, evidence-based answers to questions about a new diagnosis or prescribed treatment. Strategies commonly used to distill data into provider-facing tools to guide decision-making (eg, predictive modeling, machine learning) [1,6–8] may prove powerful for generating patient-facing content from registry data that can specifically address patient-centered questions. Men with localized prostate cancer often face multiple viable treatment options. In the setting of uncertainty, high-quality decisions are those that are informed, values-based, and implemented without undue delay [9].

In this context, we used data from a large clinical registry of men diagnosed with prostate cancer maintained by the Michigan Urological Surgery Improvement Collaborative (MUSIC) and input from patient advocates to develop and validate a model to help patients focus their self-education on the treatments most likely to be recommended by their urologist. Although use of such a model will not replace the shared decision-making process, it may complement the information that patients receive from traditional educational materials, potentially reducing decisional uncertainty for patients. As a second aim, we worked to integrate this model into a web-based platform to generate a user-friendly interface that patients can use to interact with the predictive model to augment their decision-making.

2.   Patients and methods

2.1.   Data source

MUSIC was established in 2011 in partnership with Blue Cross Blue Shield of Michigan. The quality improvement collaborative currently comprises 45 diverse community and academic urology practices representing approximately 90% of the urologists in the state. For all men seen in MUSIC practices who undergo a prostate biopsy, trained data abstractors prospectively enter a standardized set of data elements into the registry database. These elements include demographic and clinicopathologic information related to prostate biopsies, subsequent diagnoses, treatments, and follow-up care, from which both the patient characteristics and treatment strategies are derived. Prior reports have described the data acquisition and quality control activities for MUSIC, which include annual data audits at each practice and validation analyses based on insurance claims [10,11]. Each MUSIC practice obtained an exemption or approval for collaborative participation from a local institutional review board.

2.2.   Study cohort

The cohort in the present study included all men in the MUSIC registry newly diagnosed with prostate cancer between November 4, 2015 and November 3, 2017 for whom a subsequent treatment decision was documented. Although MUSIC registry data for prostate cancer date back to 2011, we excluded data from before 2015 because increasing utilization of active surveillance among patients with favorable-risk prostate cancer may lead to outdated treatment probabilities if older data are included.

We excluded patients with metastatic disease using clinical staging information documented within 150 d of the initial biopsy and patients with a missing date of birth or Gleason score or with a weight outside the range of 80–700 lbs. We also excluded patients who had undergone rare treatments, including chemotherapy, cryotherapy, high-intensity focused ultrasound, immunotherapy, and radical cystoprostatectomy.

We divided patients into derivation and validation cohorts using 2:1 random sampling of the overall cohort, stratified by practice location. Thus, each of the 45 urology practices was represented in both the derivation and validation cohorts.

2.3.   Model development

Using age at biopsy, prediagnosis prostate-specific antigen level, number of positive cores on biopsy, total cores taken at biopsy, and weight as continuous predictors; primary, secondary, and overall biopsy Gleason scores as numerical values; and history of myocardial infarction and diabetes status as binary predictors, we fitted a multinomial random forest model to predict the probability of receiving a given primary treatment. Random forest is a nonparametric method described by Leo Breiman for classification (binary or multinomial outcome) and regression (continuous outcome) problems [12].

The predicted primary treatment outcomes included: radical prostatectomy, radiation therapy (either external beam, brachytherapy, or both), primary androgen deprivation therapy, active surveillance, and watchful waiting. Active surveillance and watchful waiting are coded as distinct separate entities in the registry, as described elsewhere [13]. Additional details are discussed in the Supplementary material.

2.4.   Model validation

After fitting the random forest model using the derivation cohort, the discrimination of the model was evaluated in the validation cohort using a multiclass area under the curve (AUC) measure [14]. Model calibration was evaluated using a calibration plot comparing deciles of predicted probabilities for each outcome with observed probabilities.

2.5.   Missing data

Variables with missing values were assumed to carry information. During model training, optimal binary splits were determined by minimizing the error using nonmissing data. After a variable split was

determined, missing values for that variable were assigned to the direction minimizing the error. When generating predictions, missing values followed the assigned direction [15].

2.6. Development of the askMUSIC web platform

The askMUSIC platform was developed using the Shiny web app platform in the R language. The website was designed with input from the study team and four patients with prostate cancer. Patients reviewed the initial version of the tool and communicated their opinions and recommendations first via a conference call and then by e-mail. This was facilitated by a urologist (G.B.A.) and an informatician (K.S.).

2.7. Software

We used R 3.4.3 for all analyses [16]. The random forest model was implemented using the *h2o* R package [17], which links to the *h2o* Java program (version 3.16.0.2).

3. Results

We identified 7543 men with newly diagnosed localized prostate cancer who met the inclusion and exclusion criteria (Fig. 1). Of these men, 3413 (45%) underwent radical prostatectomy, 2289 (30%) active surveillance, 1280 (17%) radiation, 422 (5.6%) androgen deprivation therapy, and 139 (1.8%) watchful waiting. After stratifying by practice location, we randomly assigned 5016 men to the derivation cohort and 2527 to the validation cohort. There was a statistically significant but clinically minor difference in biopsy Gleason score between the two cohorts (Table 1). The difference in treatment choices between the two cohorts was not statistically significant (Table 2).

3.1. Model discrimination and calibration

The model achieved excellent discrimination, with an overall multiclass AUC of 0.81 in the validation cohort (Table 3). The easiest treatments to distinguish from one another were active surveillance and androgen deprivation therapy, while the most difficult to distinguish were active surveillance and watchful waiting. The model was well calibrated in the validation cohort (Fig. 2).

3.2. Model characteristics

Although the model parameters are not readily shareable in a random forest, the relative importance of variables can be derived using a permutation method [18]. Age is the most important variable, followed by the number of positive cores and overall biopsy Gleason score ($p < 0.001$; Supplementary Table 1). Relationships between the predictors and treatment options are demonstrated using a partial dependence plot in Supplementary Figure 1. In general, higher age is associated with a lower probability of radical prostatectomy and a higher probability of alternative strategies. A higher number of positive cores and a lower number of total cores are both associated with greater probability of undergoing radical prostatectomy. Higher biopsy Gleason score is associated with a lower
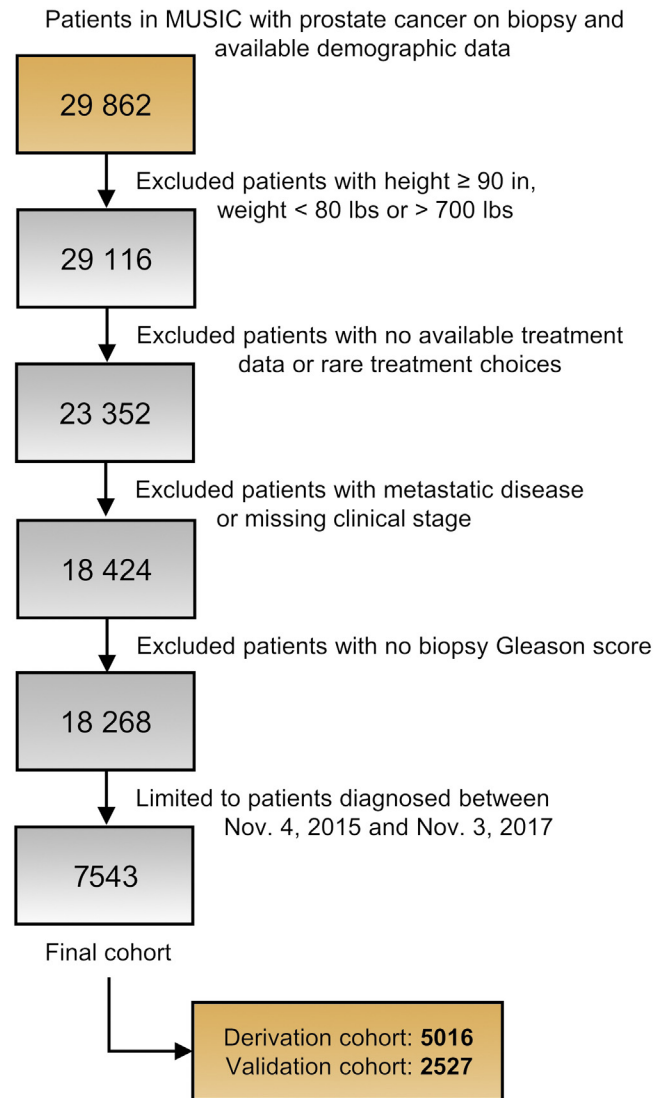

Fig. 1 – Flow chart of patients meeting the inclusion and exclusion criteria.

probability of active surveillance (Supplementary Figs. 1 and 2).

3.3. Model deployment using the askMUSIC platform

We held a teleconference with four patient advocates who are part of the MUSIC collaborative on August 5, 2016 to obtain feedback on the presentation of the model results. The key suggestions that have been incorporated into the tool are summarized in Supplementary Table 2. The model is accessible to patients at http://ask.musicurology.com (Fig. 3). A link to its code is available in Supplementary Table 3.

4. Discussion

We developed a multinomial random forest model to predict the probability of receiving a given treatment for localized prostate cancer based on common clinicopathologic and

Table 1 – Baseline characteristics for the model derivation and validation cohorts

Characteristic	Derivation cohort	Validation cohort	p value [a]	Missing (%)
Patients (n)	5016	2527		
Median age, yr (IQR)	66 (60–71)	66 (60–71)	0.787	0.0
Biopsy Gleason score, n (%)			0.047	0.0
≤6	1762 (35)	874 (35)		
3 + 4	1639 (33)	790 (31)		
4 + 3	733 (15)	428 (17)		
8	502 (10)	239 (9.5)		
9	336 (6.7)	183 (7.2)		
10	44 (0.9)	13 (0.5)		
Median PSA, ng/ml (IQR)	6.1 (4.6–8.9)	6.2 (4.7–9.1)	0.198	3.2
Median positive cores, n (IQR)	4 (2–6)	4 (2–6)	0.825	1.6
Median total cores, n (IQR)	12 (12–12)	12 (12–12)	0.550	1.6
Diabetes mellitus, n (%)	648 (13)	341 (14)	0.507	0.0
History of MI, n (%)	161 (3.2)	87 (3.4)	0.640	0.0
Median weight, lbs (IQR)	200 (177–225)	200 (177–225)	0.997	0.0

IQR = interquartile range; PSA = prostate-specific antigen; MI = myocardial infarction.
[a] A χ^2 test was used to compare categorical variables and a Wilcoxon test for continuous variables.

Table 2 – Primary treatment received by cohort [a]

Primary treatment	Patients, n (%)	
	Derivation cohort	Validation cohort
Active surveillance	1535 (31)	754 (30)
Androgen deprivation therapy	283 (5.6)	139 (5.5)
Radiation	829 (17)	451 (18)
Radical prostatectomy	2263 (45)	1150 (46)
Watchful waiting	106 (2.1)	33 (1.3)

[a] p = 0.088, χ^2 test.

Table 3 – Model discrimination between differing primary treatments

Primary treatments	Validation cohort AUC
Active surveillance vs androgen deprivation therapy	0.96
Active surveillance vs radiation	0.89
Active surveillance vs radical prostatectomy	0.90
Active surveillance vs watchful waiting	0.64
Androgen deprivation therapy vs radiation	0.67
Androgen deprivation therapy vs radical prostatectomy	0.84
Androgen deprivation therapy vs watchful waiting	0.80
Radiation vs radical prostatectomy	0.68
Radiation vs watchful waiting	0.80
Radical prostatectomy vs watchful waiting	0.92
Multiclass AUC	0.81

AUC = area under the curve.

demographic factors. The model exhibited excellent discrimination and calibration between treatment options in the validation cohort. Integration of the model into a web-based platform allows patients to access this tool on the Internet.

Predictive modeling has long been leveraged to aid physicians looking to use large data sets to augment real-time decision-making [1,2,6,8]; applying these principles to create patient-facing models may represent a meaningful way to provide patients with insights from registry data while also addressing data security issues. Merging high-quality registry data, random forest predictive modeling, and a web platform to allow simple interaction with the models may help patients to rapidly identify the information in a registry most relevant to their situation to augment decision-making. Furthermore, as patients using this tool interact with an interface to query a model as opposed to raw data, it should largely eliminate many of the usability and data security concerns that may arise from allowing interactions with raw data. This tool may provide patients with an objective source of additional information to inform decision-making. Examples of possible uses are as follows: (1) men could use the model to gain perspective on the treatment options that their physician is likely to recommend; (2) the model may enable patients to ask physicians about discrepancies in individual recommendations in comparison to treatments received by similar men as they work to make a decision about what is the best road forward for management; and (3) physicians wanting a patient to

seriously consider an alternative management strategy to the patient's initial choice may benefit from referring patients to the model so they can understand how common certain strategies are for a given stage of disease.

Our findings should be considered in the context of several limitations. First, predictive models learned from human behavior can reinforce human biases [19]. For instance, if active surveillance is appropriate but underutilized in the data, then a model derived from the data may encourage continued underutilization of active surveillance. Similarly, primary androgen deprivation therapy monotherapy is generally not an appropriate treatment for localized prostate cancer, so the model finding that similar men receive this treatment does not necessarily mean that it is medically appropriate. We have tried to avoid treatment suggestions that are based on out-of-date or nonrepresentative practice patterns by discarding older data before model fitting and including all MUSIC practices in the derivation cohort. Nonetheless, the model does not account for personal preferences with respect to survival and the side effects of different treatments, so while this tool has potential to serve as an important support to decision-making, it should not replace physician-led counseling and
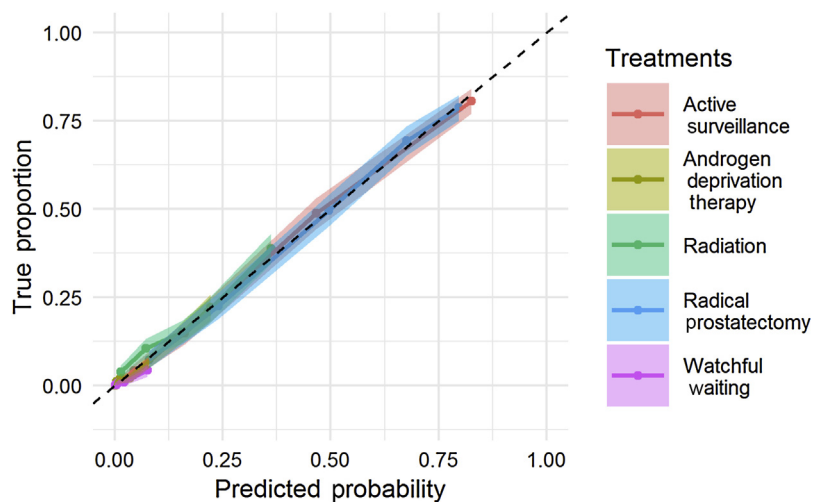
Fig. 2 – Calibration plot for each treatment in the validation cohort, with shaded 95% confidence intervals.
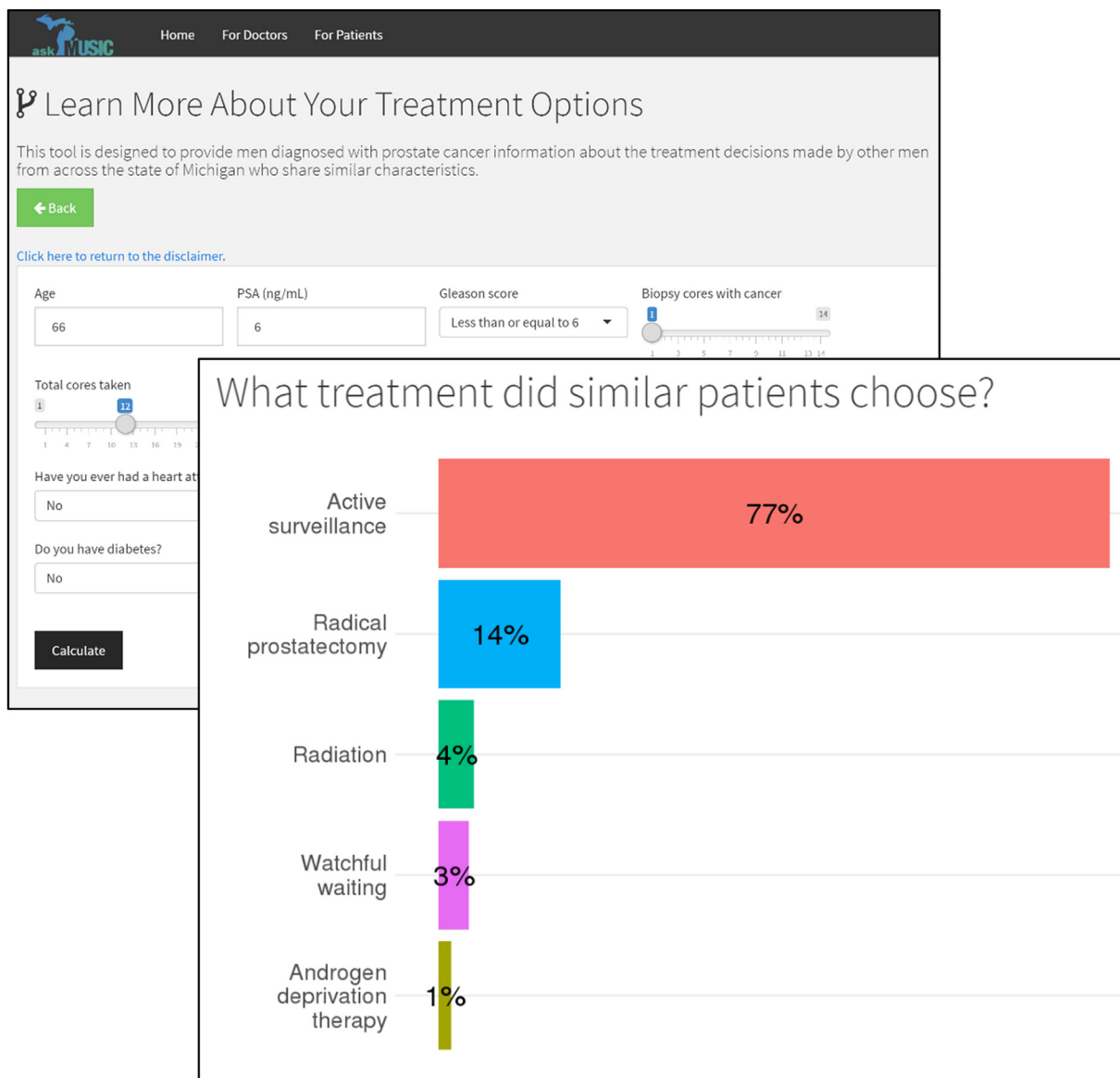

Fig. 3 – Screenshot of the askMUSIC web platform.

shared decision-making. Second, how patient use of the predictive model impacts their decision-making is not yet known. In the absence of an objectively correct treatment decision, there is reason to believe that aggregated patient data—either directly or through a model—may be useful in supporting decision-making [20]. In the face of an overwhelming amount of information available online, the model may help patients to focus their self-education on a narrower set of treatment options. Third, we have not externally validated this model outside the state of Michigan. Both our derivation and validation cohorts are representative of overall contemporary practice patterns within MUSIC. Finally, the model performed poorly in discriminating between active surveillance and watchful waiting. Although these management options have different definitions [21], these terms are sometimes applied interchangeably and may explain some of the difficulty with discrimination.

Despite these limitations, our work has important implications. Treatment recommendations can vary substantially across different risk strata and there is often more than one viable treatment option [22]. The last 10–20 yr have been notable for significant changes in how often various management strategies are used for prostate cancer [23]. Deciding on a course of treatment may lead to significant anxiety for patients [24] and may in part explain why many patients favor overtreatment [25,26] in the face of pressure from family members and physicians to select more aggressive treatments than may be warranted [27]. One approach that has proven successful in leveraging patient experiences to drive a better understanding of treatment options and decision-making about health is the use of patient-led online communities [28]. Providing patients with a tool to help them understand decisions made by similar men serves a similar purpose, although the effect of the tool on patient decision-making remains to be seen.

In the future we plan to disseminate the askMUSIC tool across our collaboration. Understanding how the tool impacts patient decision-making and decisional uncertainty will be an important next step, and particularly whether the model leads patients to choose less aggressive treatments for low-risk cancers or vice versa.

5. Conclusions

We have developed a predictive model using the random forest machine learning method that uses clinicopathologic and demographic characteristics to provide an individualized prediction of treatment in the state of Michigan. Predictive modeling may represent an important mechanism for unlocking information in data registries to allow patients to learn directly from these valuable repositories.

The abstract on which this manuscript is based was presented at the American Urological Association 2017 annual meeting.

Author contributions: Karandeep Singh had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Auffenberg, Ghani, Usoro, Denton, Rogers, Stockton, Miller, Singh.
Acquisition of data: Auffenberg, Singh.
Analysis and interpretation of data: Auffenberg, Ghani, Usoro, Ramani, Denton, Rogers, Stockton, Miller, Singh.
Drafting of the manuscript: Auffenberg, Singh.
Critical revision of the manuscript for important intellectual content: Auffenberg, Ghani, Usoro, Ramani, Denton, Rogers, Stockton, Miller, Singh.
Statistical analysis: Ramani, Singh.
Obtaining funding: Auffenberg, Denton, Miller, Singh.
Administrative, technical, or material support: Usoro.
Supervision: Singh, Ghani.
Other: None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.eururo.2018.09.050.

References

[1] 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;173:1938–42. http://dx.doi.org/10.1097/01.ju.0000158155.33890.e7.

[2] Desai NR, Kennedy KF, Cohen DJ, et al. Contemporary risk model for inhospital major bleeding for patients with acute myocardial infarction: the Acute Coronary Treatment and Intervention Outcomes Network (ACTION) registry®–Get With The Guidelines (GWTG)®. Am Heart J 2017;194:16–24. http://dx.doi.org/10.1016/j.ahj.2017.08.004.

[3] Share DA, Campbell DA, Birkmeyer N, et al. How a regional collaborative of hospitals and physicians in Michigan cut costs and im-

proved the quality of care. Health Aff 2011;30:636–45. http://dx.doi.org/10.1377/hlthaff.2010.0526.

[4] Centers for Medicare & Medicaid Services. 2016 Physician Quality Reporting System (PQRS): Qualified Clinical Data Registry (QCDR) participation made simple. www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2016PQRS_QCDR_MadeSimple.pdf

[5] The Office of the National Coordinator for Health Information Technology. Electronic capabilities for patient engagement among U.S. non-federal acute care hospitals: 2012–2015. https://dashboard.healthit.gov/evaluations/data-briefs/hospitals-patient-engagement-electronic-capabilities-2015.php

[6] Yeh RW, Secemsky EA, Kereiakes DJ, et al. Development and validation of a prediction rule for benefit and harm of dual antiplatelet therapy beyond 1 year after percutaneous coronary intervention. JAMA 2016;315:1735. http://dx.doi.org/10.1001/jama.2016.3775.

[7] Pencina MJ, Peterson ED. Moving from clinical trials to precision medicine: the role for predictive modeling. JAMA 2016;315:1713–4. http://dx.doi.org/10.1001/jama.2016.4839.

[8] Stephenson AJ, Scardino PT, Eastham J, et al. Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. J Clin Oncol 2005;23:7005–12. http://dx.doi.org/10.1200/JCO.2005.01.867.

[9] Berry DL, Halpenny B, Hong F, et al. The Personal Patient Profile-Prostate decision support for men with localized prostate cancer: a multi-center randomized trial. Urol Oncol 2013;31:1012–21. http://dx.doi.org/10.1016/j.urolonc.2011.10.004.

[10] Womble PR, Linsell SM, Gao Y, et al. A statewide intervention to reduce hospitalizations after prostate biopsy. J Urol 2015;194:403–9. http://dx.doi.org/10.1016/j.juro.2015.03.126.

[11] Womble PR, Montie JE, Ye Z, Linsell SM, Lane BR, Miller DC. Contemporary use of initial active surveillance among men in Michigan with low-risk prostate cancer. Eur Urol 2015;67:44–50. http://dx.doi.org/10.1016/J.EURURO.2014.08.024.

[12] Breiman L. Random forests. Mach Learn 2001;45:5–32. http://dx.doi.org/10.1023/A:1010933404324.

[13] Luckenbaugh AN, Auffenberg GB, Hawken SR, et al. Variation in guideline concordant active surveillance followup in diverse urology practices. J Urol 2017;197:621–6. http://dx.doi.org/10.1016/j.juro.2016.09.071.

[14] Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach Learn 2001;45:171–86. http://dx.doi.org/10.1023/A:1010920819831.

[15] H2O.ai. Distributed random forest (DRF)—H2O 3.16.0.4 documentation. http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/drf.html

[16] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

[17] Aiello S, Kraljevic T, Maj P, H2O.ai team. h2o: R interface for H2O 2016. https://mran.microsoft.com/snapshot/2016-05-21/web/packages/h2o/h2o.pdf

[18] Altmann A, Tolosi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics 2010;26:1340–7. http://dx.doi.org/10.1093/bioinformatics/btq134.

[19] Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science 2017;356:183–6. http://dx.doi.org/10.1126/science.aal4230.

[20] Longhurst CA, Harrington RA, Shah NH. A "green button" for using aggregate patient data at the point of care. Health Aff 2014;33:1229–35. http://dx.doi.org/10.1377/hlthaff.2014.0099.

[21] Cher ML, Dhir A, Auffenberg GB, et al. Appropriateness criteria for active surveillance of prostate cancer. J Urol 2017;197:67–74. http://dx.doi.org/10.1016/j.juro.2016.07.005.

[22] Mohler JL, Armstrong AJ, Bahnson RR, et al. Prostate cancer, version 1.2016. J Natl Compr Cancer Netw 2016;14:19–30.

[23] Cooperberg MR, Carroll PR. Trends in management for patients with localized prostate cancer, 1990–2013. JAMA 2015;314:80. http://dx.doi.org/10.1001/jama.2015.6036.

[24] Lin GA, Aaronson DS, Knight SJ, Carroll PR, Dudley RA. Patient decision aids for prostate cancer treatment: a systematic review of the literature. CA Cancer J Clin 2009;59:379–90. http://dx.doi.org/10.3322/caac.20039.

[25] Moses KA, Penson DF, Du L, Ahmed I, Morgans AK. Patient expectations from systemic treatment of metastatic prostate cancer (mPC). J Clin Oncol 2017;35(6 Suppl):218. http://dx.doi.org/10.1200/JCO.2017.35.6_suppl.218.

[26] Xu J, Janisse J, Ruterbusch JJ, et al. Patients' survival expectations with and without their chosen treatment for prostate cancer. Ann Fam Med 2016;14:208–14. http://dx.doi.org/10.1370/afm.1926.

[27] Xu J, Neale AV, Dailey RK, Eggly S, Schwartz KL. Patient perspective on watchful waiting/active surveillance for localized prostate cancer. J Am Board Fam Med 2012;25:763–70. http://dx.doi.org/10.3122/jabfm.2012.06.120128.

[28] Wicks P, Massagli M, Frost J, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12:e19. http://dx.doi.org/10.2196/jmir.1549.