AWARD NUMBER: CDMRPL-16-0-DM160500

TITLE: **Development of Clinical Decision Support Tools for Patients with Advanced Prostate Cancer**

PRINCIPAL INVESTIGATOR: Dr. Jonathan Forsberg, MD, PhD

CONTRACTING ORGANIZATION: Uniformed Services University of the Health Sciences (USUHS), Bethesda MD

REPORT DATE: January 2021

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
January 2021	FINAL	15Mar2017-14Sep2020

4. TITLE

Development of Clinical Decision Support Tools for Patients with Advanced Prostate Cancer

5a. CONTRACT NUMBER
CDMRPL-16-0-DM160500

5b. GRANT NUMBER
DHA-16-CRII-PMRA

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
CAPT Jonathan Forsberg, MD PhD

E-Mail: jonathan.a.forsberg.mil@mail.mil

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Uniformed Services University
4301 Jones Bridge Rd
Bethesda MD 20814

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project has improved our understanding of advanced prostate cancer and skeletal-related events (SRE's - defined as pain or fracture due to metastatic cancer). From the patient's perspective, SRE's adversely affect functional independence and quality of life. From the payer's perspective, these complications are not only quite common, but also disproportionately expensive to treat. The developed *clinical decision support* tools help estimate the likelihood of survival, disease progression, and skeletal related events, at several time-points useful for medical and surgical decision-making in an effort to deliver more focused *Precision Medicine* to men living with advanced prostate cancer. These models have been validated, and formatted for use online (www.pathfx.org).

15. SUBJECT TERMS

Prostate cancer, bone metastases, Bayesian belief network, BBN, random forest, LASSO

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	36	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION:

This proposal will strengthen our understanding of advanced prostate cancer and help prevent skeletal-related events (SRE's)—defined as pain or fracture due to metastatic cancer. From the patient's perspective, SRE's adversely affect functional independence and quality of life. From the payer's perspective, these complications are not only quite common, but also disproportionately expensive to treat. The investigators will develop *clinical decision support* tools designed to estimate the likelihood of survival, disease progression, and skeletal related events, at several time-points useful for medical and surgical decision-making in an effort to deliver more focused *Precision Medicine* to men living with advanced prostate cancer. Given that prognostic indicators, treatment protocols and survival estimates vary considerably by oncologic disease, it is beneficial to develop disease-specific survival models. This would allow for the inclusion of prognostic information unique to men with metastatic prostate cancer, such as the prostate specific antigen concentration at the time of diagnosis of metastatic bone disease (proximal PSA), and the change in alkaline phosphatase concentration over time, or alkaline phosphatase velocity (APV). Our purpose was to determine whether it was possible to develop models capable of estimating the likelihood of survival post-treatment for skeletal related events (SRE) at timepoints relevant to men with metastatic bone disease due to prostate carcinoma. These models will be validated, and formatted for use online (www.pathfx.org) as well as integrated within the electronic health record.

2. KEYWORDS:

Prostate Cancer, Precision medicine, bone metastases, Bayesian belief network, BBN, random forest, Least Absolute shrinkage and Selection Operator, LASSO

3. ACCOMPLISHMENTS:

What were the major goals of the project?

The major aims are to;
SA1: Model Development and validation (24 mos);
Major Task 1: Identify and Validate Data set
Major Task 2: Develop appropriate analytic models designed to estimate study outcomes;
SA2:Model Implementation and validation across datasets (24-36 mos)
Major Task 1:Model Implementation using PATHfx framework with online functionality

To achieve the main goals of the study there are several key elements that must be achieved in the process. 1) Regulatory approval (complete) 2) Set up contracts for analytics vendor (complete) 3) Develop the 3 main models prior to implementation and external validation (BBN, RF, LASSO)(complete) 4) Model refinement (complete) 5) Disseminate project findings (finalizing)

What was accomplished under these goals?

- Local Regulatory approval obtained (USUHS): 1/22/2018
- Secondary (HRPO) Regulatory Approval Obtained: 3/14/2018 (HRPO Log Number A-20109)
- Analytics vendor (Perduco) and contract initiated and API user manuals developed
- Initial data models created; we have written the modeling/validation/UI code using prior data and are preparing various packages in R that will be used on the backend analysis.
- API framework updated and improved PATHfx online functionality.
- Publication of results is peer reviewed journal (in review)
- Implementation on PATHFx API and user interface completed.
- For each time point (12, 24, 26, 48, 60, and 120 months), feature directionality graph was completed using the Lime package in R open software. Graph shows positive and negative influence of included variables (see 12 months graph below).

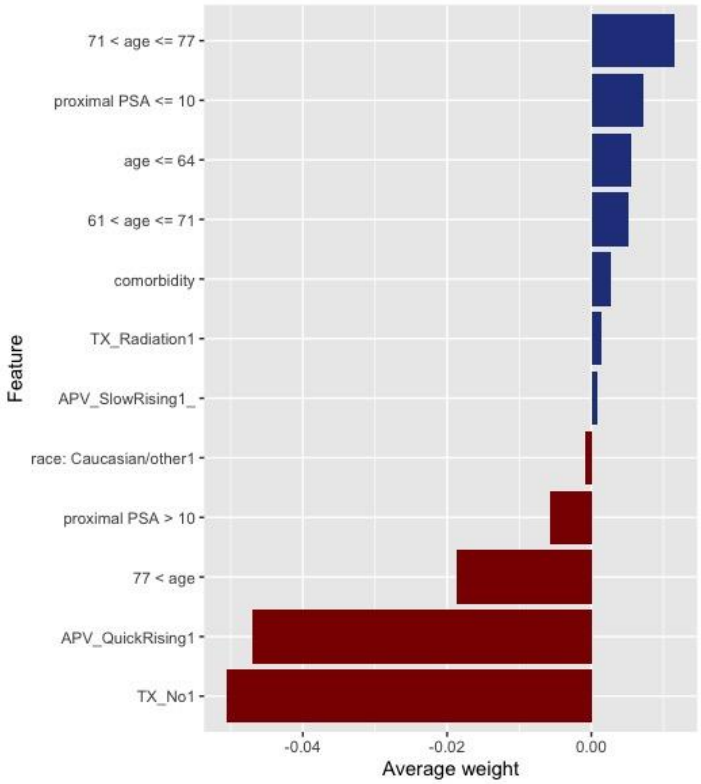

Fig. 6. This figure illustrates the directionality (to support or contradict the outcome of interest) of each level of the model features, ranked by average weight of feature level across all cases. Blue horizontal bars (positive feature weight) are associated with feature subcategories that confirm or support the outcome of interest (Survival12Months). Top positive influence on survival includes younger age and lower proximal PSA values. Red horizontal bars (negative feature weight) are associated with feature subcategories that contradict or refute the outcome of interest. Top negative influence on the outcome includes No treatment, quick rising type APV, and older age.

What opportunities for training and professional development has the project provided?

Nothing to Report

How were the results disseminated to communities of interest?

- Peer Reviewed paper has been submitted to Journal of Bone and Joint and currently in review.

What do you plan to do during the next reporting period to accomplish the goals?

- Final Report – nothing further to report

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

What was the impact on other disciplines?

- *This tool can be more broadly adopted to other cancer types and new models can be developed using similar techniques. It can be used as a working example of machine learning in health care. Second, it broadens the scope of PathFx to include additional cancer types to apply to a wider range of patients/providers.*

What was the impact on technology transfer?

Nothing to Report

What was the impact on society beyond science and technology?

We continue to disseminate information and update the International Bone Metastasis Registry, which will guide data collection efforts in patients with metastatic bone disease due to prostate cancer.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

- Nothing to Report.

Actual or anticipated problems or delays and actions or plans to resolve them

- No additional issues are anticipated
- Final Review and publication of project findings will proceed.

Changes that had a significant impact on expenditures

- none

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

.

Significant changes in use or care of human subjects

- Not applicable

Significant changes in use or care of vertebrate animals

-N/A

Significant changes in use of biohazards and/or select agents

- **N/A**

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

- None yet to report. Publication pending submission in Journal of Bone and Joint.

Books or other non-periodical, one-time publications.

Nothing to Report

Other publications, conference papers and presentations.

Nothing to Report- no recent conferences due to COVID 19 travel/meeting restrictions.

- **Website(s) or other Internet site(s)**

  - www.Pathfx.org

- **Technologies or techniques**

  Nothing to Report

- **Inventions, patent applications, and/or licenses**

Nothing to Report

- **Other Products**

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

Name: *Jonathan Forsberg, MD PhD*
Project Role: *Principle Investigator*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *12*
Contribution to Project: Project oversight.

Name: *Jennifer Cullen*
Project Role: *co-Principle Investigator*
Nearest person month worked: *12*
Contribution to Project: revisions to IRB protocol, training of team to perform QA/QC on data for this study.
Name: *Yongmei Chen*
Project Role: *Principal Biostatistician*
Nearest person month worked: *12*
Contribution to Project: QA/QC for data fields.

Name: *Claire Kuo*
Project Role: *Data Analyst*
Nearest person month worked: *12*
Contribution to Project: QA/QC for data fields

Name: *Clare Grazal*
Project Role: *Data analyst*
Nearest person month worked: *12*
Contribution to Project- Data analytics and modelling.

Name: *Michael Wiley*
Project Role: *Program Manager*
Nearest person month worked: *3*
Contribution to Project- Program and financial management

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to Report

What other organizations were involved as partners?

Nothing to Report

8. **SPECIAL REPORTING REQUIREMENTS**
   COLLABORATIVE AWARDS:
   QUAD CHARTS:

9. **APPENDICES:**

Development of clinical decision support tools for patients with advanced prostate cancer
Clinical Research Intramural Initiative program- Precision Medicine Award; Log# DM 160500

Co-PIs: Forsberg, Jonathan / Cullen, Jennifer **Org:** Uniformed Services University of the Health Sciences **Award Amount:** $666,352

Study/Product Aim(s): In this study, three clinical decision support tools will be developed for use in men with prostate cancer. Each will estimate the likelihood of survival, disease progression, and skeletal related events, at multiple time-points useful for medical and surgical decision-making. These models will be validated, and formatted for use online (www.pathfx.org) as well as integrated within the electronic health record. These tools will benefit tailored decision making and contribute to the goal of precision medicine.

Approach: This protocol will leverage a large, existing military cohort to apply Bayes' conditional probability theorem to develop prostate cancer-specific clinical decision support tools, using patient clinical and treatment characteristics to predict key study outcomes, including overall survival and metastasis bone.

Accomplishment: Updated PATHfx user interface.

Timeline and Cost

Activities CY	17	18	19	20	Total
Aim 1: Identify and Validate Data set					
Aim 1: Develop Models					
Aim 2: Model Implementation and Validation					
Manuscript Preparation and summary findings					
Estimated Budget ($K)	**$257K**	**$262K**	**$147K**		**$666K**

Goals/Milestones

CY17 Goals
- ✓ Obtain regulatory compliance; complete
- ✓ Identify and Validate Data set- Identified

CY18 Goals
- ✓ Update web U/I for PATHFx updated modelling- complete
- ✓ Develop data models (BBN, LASSO, RFP)- complete

CY19 Goals
- ✓ Implement modelling and update to PATHFx 3.0 framework - complete
- ✓ Complete external validation dataset/collaborators) - complete
- ✓ Update models based on external validation- complete
- ✓ Present preliminary analysis at national meetings – N/A due to COVID 19

CY20 Goals
- ✓ Publish conclusions – publication in review
- ✓ Create API for future collaborations

Appendix 1: Publication draft

Estimating 10-Year Survival in Patients with Bone Metastases due to Prostate Cancer: Toward a Disease-Specific Survival Estimation Tool

Ashley B. Anderson, MD, Resident[1], Clare Grazal, MS, Bioinformatician[2], Rikard Wedin, MD, PhD, Associate Professor[3], Claire Kuo, MS, MPH, Biostatistician[4], Yongmei Chen, MS, MD, Biostatistician[4], Bryce R. Christensen, MD, MBA, Resident[5], Jennifer Cullen, MPH, PhD, Epidemiologist[4], and Jonathan A. Forsberg, MD, PhD, Professor[1], Orthopaedic Oncologist[6]

[1]Division of Orthopaedics, Department of Surgery, Uniformed Services University, Walter Reed National Military Medical Center, 8901 Rockville Pike, Bethesda, MD 20889; [2]The Henry Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr, Suite 100 Bethesda, MD 20817; [3]Department of Molecular Medicine and Surgery (MMK), K1, Orthopaedics, Karolinska Institutet, A2:07 171 76, Stockholm, Sweden; [4]Center for Prostate Disease Research, Department of Surgery, Uniformed Services University, Walter Reed National Military Medical Center, 6720A Rockledge Dr, Suite 300, Bethesda MD 20817; [5]Department of Internal Medicine, San Antonio Military Medical Center, 3551 Roger Brooke Dr, San Antonio, TX 78219; [6]Department of Orthopaedic Surgery, The Johns Hopkins University Hospital, 601 N. Caroline St., Baltimore MD 21287

INTRODUCTION

In the United States, prostate cancer is the most common diagnosed malignancy and the second leading cause of cancer death in men.[1,2] The clinical treatment decision-making process is

challenging because prostate cancer is a complex disease. Several tumor markers and biomarkers are associated with prognosis. For example, proximal prostate-specific antigen (PSA) (defined as the most recent value measured at least 6 months before developing metastasis) <10 ng/mL is an independent predictor of metastasis-free survival among men with biochemical recurrence after undergoing radical prostatectomy.[3] In addition, the change in alkaline phosphatase concentration over time, alkaline phosphatase velocity (APV), is a prognostic biomarker associated with overall survival in men with castration-resistant prostate cancer.[4,5] Metwalli et al[5] found that higher APV was also an independent predictor of overall survival, as well as for bone metastasis–free survival in patients with castration-resistant prostate cancer, where APV ≥50 (upper quartile) is "quick rising", APV of 0 is "no rising", and all other APV values are "slow rising." High APV (uppermost quartile of velocity) is also predictive of distant metastasis–free survival in patients who have undergone radical prostatectomy and experienced biochemical recurrence.[4] Despite their association with survival, these biomarkers have not been used as part of a cohesive or otherwise multivariable survival model.

The approach to treating men with metastatic bone disease due to prostate cancer requires balanced consideration of clinical benefits, life expectancy, comorbidities, quality of life, and the risk of adverse effects. Clinical practice guidelines[6] published recently by the Musculoskeletal Tumor Society recommend that physicians consider using clinical support tools, such as PATHFx (www.pathfx.org), to estimate a patient's survival trajectory by estimating survival after treatment for a skeletal-related event (SRE), which is defined as pathologic fracture; spinal cord compression requiring surgical treatment; or nonsurgical treatment, including radiotherapy, cryotherapy, or radiofrequency ablation. PATHFx currently estimates the likelihood of survival at 1, 3, 6, 12, 18 and 24 months after surgical or nonsurgical intervention or an SRE.[7-11] Its use

supports shared decision making by ensuring treatment strategies align with each patient's personal survival trajectory and functional goals.

In April 2019, the U.S. Food and Drug Administration released guidelines to regulate machine learning–based technologies used in patient care.[12,13] To ensure models remain relevant, these guidelines emphasise the importance of a clinical support tool's lifecycle. Although PATHFx has been externally validated in several centers worldwide, continual advances in the treatment of patients with advanced cancer require that the models be updated regularly. For this reason, we updated the six PATHFx models using recent data obtained from patients undergoing contemporary systemic therapy, including targeted agents, and immunotherapy.[14] Validation data for this study were derived from the International Bone Metastasis Registry, which helps ensure that the updated models are applicable to various patient populations worldwide. This commitment to lifecycle management ensures that PATHFx remains applicable as treatment philosophies change and new therapies become available, thereby providing clinicians with the most current, broadly applicable tool to estimate survival in this patient population.

Currently, the PATHFx tool groups cancer diagnoses according to historical data on survival rates. Because prognostic indicators, treatment protocols, and survival estimates vary widely by cancer type, it may be beneficial to develop disease-specific survival models. Such models would make use of prognostic information unique to men with metastatic prostate cancer, such as proximal PSA and APV.

Our primary purpose was to develop models to estimate the duration of survival after treatment for skeletal-related events (SREs) in men with metastatic bone disease due to prostate cancer. Such models could inform the PATHFx clinical decision support tool, which currently

groups cancer types according to historical survival data. Our secondary purpose was to determine disease-specific factors that should be included in an international cancer registry.

METHODS

Guidelines

This retrospective prognostic classification study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines[15] and the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research.[16]

Data Source and Patient Selection Criteria

The study population comprised 29,000 men enrolled in the institutional review board–approved Center for Prostate Disease Research (CPDR) Multicenter National Database Program.[17] We reviewed the records of men who sustained a SRE due to metastatic prostate cancer and subsequently required treatment with radiotherapy or surgery between 1989 and 2017. There were 1,404 men with prostate cancer that metastasised to bone in the data set. Of these, 438 patients had sufficient information to calculate APV, defined as the slope of the linear regression line of alkaline phosphatase values obtained after the diagnosis of metastatic bone disease, plotted against time in years.

Outcome

We developed six models designed to estimate the likelihood of survival at 1, 2, 3, 4, 5, and 10 years after treatment of an SRE.

Demographic, clinical, and pathologic features

Consistent with previous methods of using APV as a prognostic feature, and because of the strong skew and non-normality of the APV distribution, APV was binned into the uppermost quartile ("quick-rising") of all observed values and then compared with the lower 3 quartiles combined ("slow-rising") and zero value ("no-rising").[4,5] Proximal PSA, defined as PSA concentration at the time of diagnosis of metastatic bone disease, was missing in 11% of these records. Data for all other features were complete. For each consecutive time point, the number of patients decreased because of censoring. Patient demographic and clinical data extracted for analysis were as follows: self-reported race (black or white/other), presence of comorbidities, age at first known bone metastasis, proximal PSA, APV values, method of local treatment of the primary tumor (radiotherapy or surgery), adjuvant therapy (radiotherapy, chemotherapy, and hormonal therapy) and date of death.

Categorical and continuous features included in the models and the proportions of missing data are listed in Tables 1 and 2. We used Bayes factor (BF) analysis to compare the cohorts. BF analysis considers the strength of evidence supporting or contradicting the study hypothesis. The analysis is categorised by the following: BF ≥100 indicates strong supporting evidence for the alternative hypothesis; BF <100 indicates strong supporting evidence for the null hypothesis; and BF of approximately 0 indicates no probable difference between the 2 groups.[18,19]

Model Development

We selected gradient boosting machine (GBM) modeling because it is a decision tree machine learning technique that builds an ensemble of shallow and weak trees or learners in succession (rather at than all at once as in random forest machine learning), so each tree learns and improves from the previous iteration. GBM modeling trains models in a gradual, additive, and sequential manner, which strengthens the final product.[20,21] The final model is built on the strength of previous, smaller predictors.

We used Python, version 3.7.4 (Python Software Foundation, Beaverton, OR) to develop the models. For each model, we split the data 80/20 into train and test groups and a further 80/20 split of the train group into train and validation sets stratified by the binary outcome feature across all groups. Data were shuffled to create random order before splitting into listed groups. Because the number of patients at each time point differed, the exact number of records in the train, validation, and test groups vary by time point and cohort. Data types were changed to integer, float, and object as applicable. Missing data were imputed using the multiple imputation by chained equations algorithm from the IterativeImputer package. Our data were preprocessed to scale using the MinMaxScaler package of sklearn.preprocessing. Six GBM models were created, 1 for each of the 6 survival time points, using the train set and the GradientBoostingClassifier package in sklearn.ensemble.

Feature Selection

Because of our limited data set, we made all categorical features binary. This allowed for a more transparent analysis of results. The features included in the model were further characterised for magnitude and direction of each feature's association with the outcome (patient survival) using

the local interpretable model–agnostic explanations (LIME) package in R software for the models.[22]

Model Regulation

GBM models continue improving to minimise error at the risk of overfitting. For internal validation, we used a cross-validated grid search to direct our choice of parameters using GridSearch CV within sklearn.model_selection package (Python Software Foundation). Our scoring measure of interest was the AUC. For each model, we selected parameters that produced the highest AUC in the validation set. Parameters of interest were learning rate, number of estimators, maximum depth of tree, minimum number of samples per node to be considered for splitting, minimum number of samples required in a terminal node, and subsample percentage included for each tree.

Performance Assessment

We created predictive values for each model by using each corresponding test set. First, calibration plots were used to visualise the concurrence of the predicted probabilities with the observed frequencies in the data set. Then, discriminatory ability was determined by estimating the AUC. Next, Brier scores were used to determine overall accuracy of the predictions. The Brier score measures distance between the actual outcome and the predicted probability assigned to the outcome for each observation, where the best possible Brier score for accuracy is 0 and the worst is 1.[23,24] Finally, we determined whether the models possessed clinical utility by using decision curve analysis[25,26], as described previously in this patient population.[14]

RESULTS

Participants

Continuous and categorical features for the train and test sets are listed in Tables 2 and 3, respectively. As expected, we found no difference between continuous features in the train and test sets (BF of approximately 0) (Table 1). When comparing categorical features, we found no difference between the 2 groups (BF of approximately 0) for treatment type and survival (yes/no) at any time point (Table 2).

Model Development and Validation

The AUC was between 0.73 and 0.86 for all 6 models (Table 3). Brier scores were <0.20 and demonstrated the model's predictions were accurate. The relative influence table for the 6 models in Figure 1 shows the degree of influence for each feature on the overall model. Proximal PSA and patient age at the time of first-known SRE consistently had the most influence across all models. Treatment method and APV became increasingly influential with the later time period models.

Global Application (Model-Level Interpretation)

In earlier survival estimate models, proximal PSA and age at diagnosis had more influence on the outcome variable. Notably, APV was an important feature at all time points; quick-rising APV was more influential in later survival estimate mode. Unexpectedly, the method of treating the primary disease also had strong influence; however, treatment-naïve status was more influential on survival than was radiotherapy and/or chemotherapy.

Local Application (Patient-Level Interpretation)

To trust and apply models correctly, clinicians must be able to interpret them at the patient level.[22] The positive and negative directionality of each model is shown in Figure 1. Overall, features positively associated with survival were younger age at metastasis diagnosis, proximal PSA <10 ng/mL, slow-rising APV, no-rising APV, radiotherapy treatment, and hormonal or chemotherapy treatment (Figure 1). Features negatively associated with survival were older age at metastasis diagnosis, proximal PSA >10 ng/mL, quick-rising APV values, and being treatment-naïve (Figure 1). The patient-level interpretations were consistent with global-level model application.

Clinical Utility

Decision curve analysis showed that physicians may achieve better outcomes by using the 6 models described above, rather than assuming all will survive, or none will survive for 1, 2, 3, 4, 5, and 10 years, respectively (Figure 2). Decision curve analysis measures the net benefit of using a clinical support tool across different threshold probabilities defined as the point of equipoise when considering 2 courses of treatment (e.g., nonsurgical vs. surgical for short-term survival estimates, less invasive vs. more invasive for longer-term estimates). Low-threshold probabilities are associated with healthier patients, whereby physicians have a low threshold to offer surgery. In contrast, high-threshold probabilities are associated with patients in which surgery carries greater risk.

DISCUSSION

The duration of survival for prostate cancer patients with metastatic bone disease is difficult to predict. We successfully developed models to estimate survival in patients with prostate cancer who have metastatic bone disease to help clinicians navigate treatment algorithms. Previous studies have shown that APV is predictive of distant metastasis–free survival.[4,5,27] In this study, we showed that machine learning–based models can predict survival in prostate cancer patients, and that these models improve in both discriminatory ability and accuracy with the addition of APV data. Specifically, the patient's primary disease treatment type and APV became increasingly influential in the later time period models. Although further external validation studies are required, these data justify inclusion of these models in the PATHFx tool, an open-source clinical decision-making support tool for survival estimation (https://www.pathfx.org).[7]

Patient race and ethnicity may provide important information on genetic and socioeconomic factors pertaining to disease.[28,29] Race was self-reported by patients at the time of enrollment and divided into 2 broad categories (white/other or black). Using the CPDR database, Cullen et al[30] found self-reported black race was not a predictor of poorer overall survival among participants in the CPDR Multicenter National Database Program undergoing active surveillance, despite race-based differences in baseline clinical risk characteristics.

Although PATHFx is validated, it does not offer disease-specific estimates of survival.[14] The prostate cancer–specific models at 1 and 2 years can be compared directly with the PATHFx 1- and 2-year models in terms of discriminatory ability (AUC) and accuracy of prediction (Brier score). The new 1-year prostate disease–specific model we developed (AUC = 0.85; Brier score = 0.07) was superior to the PATHFx (version 3.0) 1-year model (AUC = 0.78; Brier score = 0.18). However, the 2-year prostate disease–specific model (AUC = 0.80; Brier score = 0.17) was no better than the PATHFx 2-year model (AUC = 0.82; Brier score = 0.12). Based on this

direct comparison, the 1-year prostate disease–specific model could be used independently to accurately determine survival duration in men with metastatic prostate cancer. However, predictive algorithms continue to improve with exposure to more data.[31] Therefore, we believe there is room for improvement by incorporating additional PATHFx variables, such as hemoglobin concentration and absolute lymphocyte count.

Although the classification ability of the prostate-specific models derived in this study is no better than that of the current PATHFx tool,[14] we have developed 4 additional models that estimate survival at 3, 4, 5, and 10 years. Validation statistics and decision curve analysis indicate that these models are suitable for clinical use. Incorporating these prostate cancer–specific models into the PATHFx clinical support tool is part of our continued responsibility to provide accurate estimations of survival to help clinicians and patients navigate complex treatment algorithms. Unlike traditional statistical decision rules, the accuracy of machine learning–based models can be improved over time with better machine learning methods, more data, changes in practice, changes in the patient population, and/or better understanding of disease processes.[31]

When evaluating the results of this study, its limitations must be considered. It is possible that other statistical techniques could be used to develop similar prognostic models for prostate cancer. Our author group has extensive experience using various machine learning techniques. Some techniques are prone to overfitting and produce overly optimistic results. Therefore, we implemented GBM with hyperparameter tuning to mitigate the risk of overfitting. Our study was limited by missing data, which can result in incomplete codification of train data and overfitting; however, we mitigated these effects by using a “holdout” validation set. Despite these results, external validation studies are necessary before these models can be recommended for use in

other patient populations. Beyond APV, there may be other laboratory-related features to consider in the model; however, the data are not complete in the CPDR database. The number of features available for the model was a limitation. Only 31% of the CPDR data had APV data, so we restricted the data set to the 438 records with APV values. Nevertheless, we anticipate that the models will continue to improve as more data become available. Although the addition of APV to the models improved performance, we may see continued improvement in model performance by including additional demographic and laboratory-based patient data. For example, Stattin et al[32] found that a panel of kallikrein marker (human kallikrein-related peptidase 2 [hK2] and total, free, and intact PSA) is strongly predictive of distant metastasis in men with modestly elevated PSA. As these data are collected and added to national and international prostate cancer registries, we could continue to augment the models for survival estimations. Additionally, different mechanisms exist to measure and categorise APV, and previously determined[4,5,27] cut points were used in this analysis.

By including disease-specific information such as APV, we developed a tool that helps predict survival duration in men with metastatic bone disease due to prostate cancer. Although external validation studies are required to support its use in other patient populations, these data justify inclusion of these models in the PATHFx tool. In addition, data used in the GBM model, including APV and proximal PSA, should be included in the International Bone Metastasis Registry.

REFERENCES

1. **Siegel RL, Miller KD, Jemal A.** Cancer statistics, 2019. *CA. Cancer J. Clin.* 2019;69:7-34.

2. **Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ.** Cancer statistics, 2009. *CA. Cancer J. Clin.* 2009;59:225-49.

3. **Markowski MC, Suzman D, Chen Y, Trock BJ, Cullen J, Feng Z, Antonarakis ES, Paller CJ, Han M, Partin AW, Eisenberger MA.** PSA doubling time (PSADT) and proximal PSA value predict metastasis-free survival (MFS) in men with biochemically recurrent prostate cancer (BRPC) after radical prostatectomy (RP). *J. Clin. Oncol.* 2017;35:5075.

4. **Salter CA, Cullen J, Kuo C, Chen Y, Hurwitz L, Metwalli AR, Dimitrakoff J, Rosner IL.** Alkaline Phosphatase Kinetics Predict Metastasis among Prostate Cancer Patients Who Experience Relapse following Radical Prostatectomy. *Biomed Res Int* 2018;2018:4727089.

5. **Metwalli AR, Rosner IL, Cullen J, Chen Y, Brand T, Brassell SA, Lesperance J, Porter C, Sterbis J, McLeod DG.** Elevated alkaline phosphatase velocity strongly predicts overall survival and the risk of bone metastases in castrate-resistant prostate cancer. *Urol Oncol* 2014;32:761-8.

6. **Musculoskeletal Tumor Society (MSTS), American Society for Radiation Oncology (ASTRO), American Society of Clinical Oncology (ASCO).** The Treatment of Metastatic Carcinoma and Myeloma of the Femur: Clinical Practice Guideline. Available at:

https://www.astro.org/ASTRO/media/ASTRO/Patient%20Care%20and%20Research/PDFs/MSTSBonemetsGLPC.pdf. Accessed on Apr 30, 2020.

7. **Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH.** Estimating survival in patients with operable skeletal metastases: An application of a bayesian belief network. *PloS one* 2011;6:e19956.

8. **Forsberg JA, Sjoberg D, Chen QR, Vickers A, Healey JH.** Treating metastatic disease: Which survival model is best suited for the clinic? *Clin Orthop Relat Res* 2013;471:843-50.

9. **Forsberg JA, Wedin R, Bauer HC, Hansen BH, Laitinen M, Trovik CS, Keller JO, Boland PJ, Healey JH.** External validation of the Bayesian Estimated Tools for Survival (BETS) models in patients with surgically treated skeletal metastases. *BMC Cancer* 2012;12:493.

10. **Meares C, Badran A, Dewar D.** Prediction of survival after surgical management of femoral metastatic bone disease - A comparison of prognostic models. *J Bone Oncol* 2019;15:100225.

11. **Ogura K, Gokita T, Shinoda Y, Kawano H, Takagi T, Ae K, Kawai A, Wedin R, Forsberg JA.** Can a multivariate model for survival estimation in skeletal metastases (PATHFx) be externally validated using Japanese patients? *Clin Orthop Relat Res* 2017;475:2263-70.

12. **Hwang TJ, Kesselheim AS, Vokinger KN.** Lifecycle Regulation of Artificial Intelligence- and Machine Learning-Based Software Devices in Medicine. *JAMA* 2019;

13. **U.S. Food & Drug Administration.** Artificial Intelligence and Machine Learning in Software as a Medical Device. Available at https://www.fda.gov/medical-

devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device. Accessed on Apr 3, 2020.

14. **Anderson AB, Wedin R, Fabbri N, Boland P, Healey J, Forsberg JA.** External Validation of PATHFx Version 3.0 in Patients Treated Surgically and Nonsurgically for Symptomatic Skeletal Metastases. *Clin Orthop Relat Res* 2020;478:808-18.

15. **Collins GS, Reitsma JB, Altman DG, Moons KG.** Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann. Intern. Med.* 2015;162:55-63.

16. **Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, Venkatesh S, Berk M.** Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J Med Internet Res* 2016;18:e323.

17. **Brassell SA, Dobi A, Petrovics G, Srivastava S, McLeod D.** The Center for Prostate Disease Research (CPDR): a multidisciplinary approach to translational research. *Urol Oncol* 2009;27:562-9.

18. **Goodman SN.** Toward evidence-based medical statistics. 1: The P value fallacy. *Ann. Intern. Med.* 1999;130:995-1004.

19. **Goodman SN.** Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials* 2005;2:282-90; discussion 301-4, 64-78.

20. **Natekin A, Knoll A.** Gradient boosting machines, a tutorial. *Front Neurorobot* 2013;7:21.

21. **Friedman JH.** Greedy function approximation: A gradient boosting machine. *Ann. Statist* 2001;29:1189-232.

22. **Ribeiro MT, Singh S, Guestrin C.** Model-Agnostic Interpretability of Machine Learning. Available at https://arxiv.org/abs/1606.05386. Accessed on Mar 16, 2020.

23. **Brier GW.** Verification of Forecasts Expressed in Terms of Probability. *Mon Weather Rev* 1950;78:1-3.

24. **Assel M, Sjoberg DD, Vickers AJ.** The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagn Progn Res* 2017;1:19.

25. **Vickers AJ, Van Calster B, Steyerberg EW.** Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;352:i6.

26. **Vickers AJ, Elkin EB.** Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making* 2006;26:565-74.

27. **Hammerich KH, Donahue TF, Rosner IL, Cullen J, Kuo HC, Hurwitz L, Chen Y, Bernstein M, Coleman J, Danila DC, Metwalli AR.** Alkaline phosphatase velocity predicts overall survival and bone metastasis in patients with castration-resistant prostate cancer. *Urol Oncol* 2017;35:460 e21- e28.

28. **Newman LA, Kaljee LM.** Health Disparities and Triple-Negative Breast Cancer in African American Women: A Review. *JAMA Surg* 2017;152:485-93.

29. **Leopold SS, Beadling L, Calabro AM, Dobbs MB, Gebhardt MC, Gioe TJ, Manner PA, Porcher R, Rimnac CM, Wongworawat MD.** Editorial: The Complexity of Reporting Race and Ethnicity in Orthopaedic Research. *Clin Orthop Relat Res* 2018;476:917-20.

30. **Cullen J, Brassell SA, Chen Y, Porter C, L'Esperance J, Brand T, McLeod DG.** Racial/Ethnic patterns in prostate cancer outcomes in an active surveillance cohort. *Prostate Cancer* 2011;2011:234519.

31. **Liu Y, Chen PC, Krause J, Peng L.** How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *JAMA* 2019;322:1806-16.

32. **Stattin P, Vickers AJ, Sjoberg DD, Johansson R, Granfors T, Johansson M, Pettersson K, Scardino PT, Hallmans G, Lilja H.** Improving the Specificity of Screening for Lethal Prostate Cancer Using Prostate-specific Antigen and a Panel of Kallikrein Markers: A Nested Case-Control Study. *Eur. Urol.* 2015;68:207-13.

Table 1. Continuous variables contained within the train and test sets

Variable by Time Point	Median (IQR)			P value*	Bayes Factor
	Whole Cohort	Train Set	Test Set		
Proximal PSA					
1-Year	33.4 (200)	38.6 (202)	23.7 (153)	0.37	0.15
2-Year	35.7 (205)	30.5 (147)	51.7 (294)	0.59	0.14
3-Year	36.7 (211)	34.0 (205)	46.2 (238)	0.38	0.15
4-Year	36.8 (217)	36.8 (232)	37.4 (202)	0.43	0.15
5-Year	39.6 (236)	42.8 (232)	31.5 (242)	0.54	0.14
10-Year	42.8 (252)	44.5 (275)	40.7 (131)	0.21	0.17
Age					
1-Year	71.0 (12.7)	71.1 (12.6)	70.1 (12.9)	0.41	0.19
2-Year	71.0 (12.7)	71.2 (12.6)	70.6 (12.4)	0.96	0.13
3-Year	71.0 (12.7)	71.0 (12.9)	71.1 (11.7)	0.95	0.13
4-Year	71.0 (12.8)	71.0 (12.7)	71.6 (13.4)	0.32	0.22
5-Year	71.0 (12.9)	70.8 (12.7)	71.5 (13.1)	0.46	0.18
10-Year	71.0 (13.0)	70.8 (12.6)	72.5 (13.7)	0.21	0.32

PSA, prostate-specific antigen.

*P values determined using Pearson's chi-squared test.

Table 2. Categorical variables contained within the train and test sets

Variable by Time Point	Whole Cohort		Train Set		Test Set		P value*	Bayes Factor
Comorbidity	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**		
1-Year	145	293	118	232	27	61	0.68	0.20
2-Year	144	286	112	232	32	54	0.49	0.26
3-Year	143	283	119	221	24	62	0.26	0.39
4-Year	139	278	115	218	24	60	0.36	0.31
5-Year	137	265	108	213	29	52	0.81	0.20
10-Year	127	246	103	195	24	51	0.78	0.21
Hormone therapy/chemotherapy								
1-Year	282	156	222	128	60	28	0.48	0.25
2-Year	276	154	230	114	46	40	0.03	2.55
3-Year	272	154	217	123	55	31	>0.99	0.19
4-Year	266	151	213	120	53	31	0.98	0.19
5-Year	260	142	208	113	52	29	>0.99	0.19
10-Year	243	130	197	101	46	29	0.52	0.27
Treatment-naïve								
1-Year	88	350	76	274	12	76	0.12	0.63
2-Year	87	343	66	278	21	65	0.35	0.28
3-Year	87	339	69	271	18	68	>0.99	0.16
4-Year	86	331	66	267	20	64	0.51	0.23
5-Year	82	320	66	255	16	65	0.99	0.16
10-Year	77	296	60	238	17	58	0.75	0.19
Radiotherapy								
1-Year	68	370	52	298	16	72	0.55	0.19
2-Year	67	363	48	296	19	67	0.09	0.76
3-Year	67	359	54	286	13	73	0.99	0.14
4-Year	65	352	54	279	11	73	0.59	0.17
5-Year	60	342	47	274	13	68	0.89	0.15
10-Year	53	320	41	257	12	63	0.76	0.17
APV of 0								
1-Year	188	250	154	196	34	54	0.43	0.28
2-Year	183	247	142	202	41	45	0.34	0.34
3-Year	180	246	149	191	31	55	0.24	0.44
4-Year	176	241	144	189	32	52	0.47	0.27

5-Year	172	230	138	183	34	47	0.97	0.20
10-Year	156	217	123	175	33	42	0.77	0.22
Quick-rising APV								
1-Year	111	327	86	264	25	63	0.55	0.22
2-Year	109	321	90	254	19	67	0.52	0.22
3-Year	109	317	82	258	27	59	0.21	0.44
4-Year	108	309	88	245	20	64	0.73	0.19
5-Year	106	296	87	234	19	62	0.60	0.21
10-Year	103	270	86	212	17	58	0.35	0.32
Slow-rising APV								
1-Year	139	299	110	240	29	59	0.88	0.19
2-Year	138	292	112	232	26	60	0.78	0.19
3-Year	137	289	109	231	28	58	>0.99	0.18
4-Year	133	284	101	232	32	52	0.22	0.46
5-Year	124	278	96	225	28	53	0.50	0.26
10-Year	114	259	89	209	25	50	0.66	0.23
Black								
1-Year	92	346	77	273	15	73	0.38	0.25
2-Year	90	340	74	270	16	70	0.66	0.18
3-Year	90	336	65	275	25	61	0.06	1.16
4-Year	88	329	72	261	16	68	0.71	0.18
5-Year	83	219	70	251	13	68	0.32	0.30
10-Year	78	295	64	234	14	61	0.71	0.19
White or other race								
1-Year	346	92	273	77	73	15	0.38	0.25
2-Year	340	90	270	74	70	16	0.66	0.18
3-Year	336	90	275	65	61	25	0.06	1.16
4-Year	329	88	261	72	68	16	0.71	0.18
5-Year	319	83	251	70	68	13	0.32	0.30
10-Year	295	78	234	64	61	14	0.71	0.19
Survival duration								
1-Year	405	33	324	26	81	7	>0.99	0.10
2-Year	330	100	264	80	66	20	>0.99	0.16
3-Year	269	157	215	125	54	32	>0.99	0.19
4-Year	223	194	178	155	45	39	>0.99	0.19
5-Year	181	221	145	176	36	45	>0.99	0.20
10-Year	71	302	57	241	14	61	>0.99	0.16

*P values determined using Pearson's chi-squared test.

TABLE 3. Summary of the accuracy (AUC) and discriminatory ability (Brier score) of the predictive model at each time period.

Model	**AUC (95% CI)**	**Brier Score (95% CI)**
1-Year	0.76 (0.61–0.91)	0.07 (0.02–0.12)
2-Year	0.73 (0.60–0.85)	0.17 (0.12–0.22)
3-Year	0.86 (0.79–0.94)	0.19 (0.16–0.21)
4-Year	0.82 (0.73–0.91)	0.20 (0.18–0.22)
5-Year	0.79 (0.69–0.89)	0.19 (0.15–0.23)
10-Year	0.79 (0.65–0.93)	0.14 (0.09–0.19)

AUC, area under the receiver operating characteristic curve; CI, confidence interval.

FIGURE LEGENDS

Figure 1 A-F. This figure shows both the relative influence of each feature and whether the feature has a positive or negative association with survival. The directionality (to support or contradict the outcome of interest) of each level of the model features is ranked by average weight of feature level across all cases. Blue bars (positive feature weight) are associated with features that are associated with survival; red bars (negative feature weight) represent features that are negatively associated with survival at (A) 1 year, (B) 2 years, (C) 3 years, (D) 4 years, (E) 5 years, and (F) 10 years.

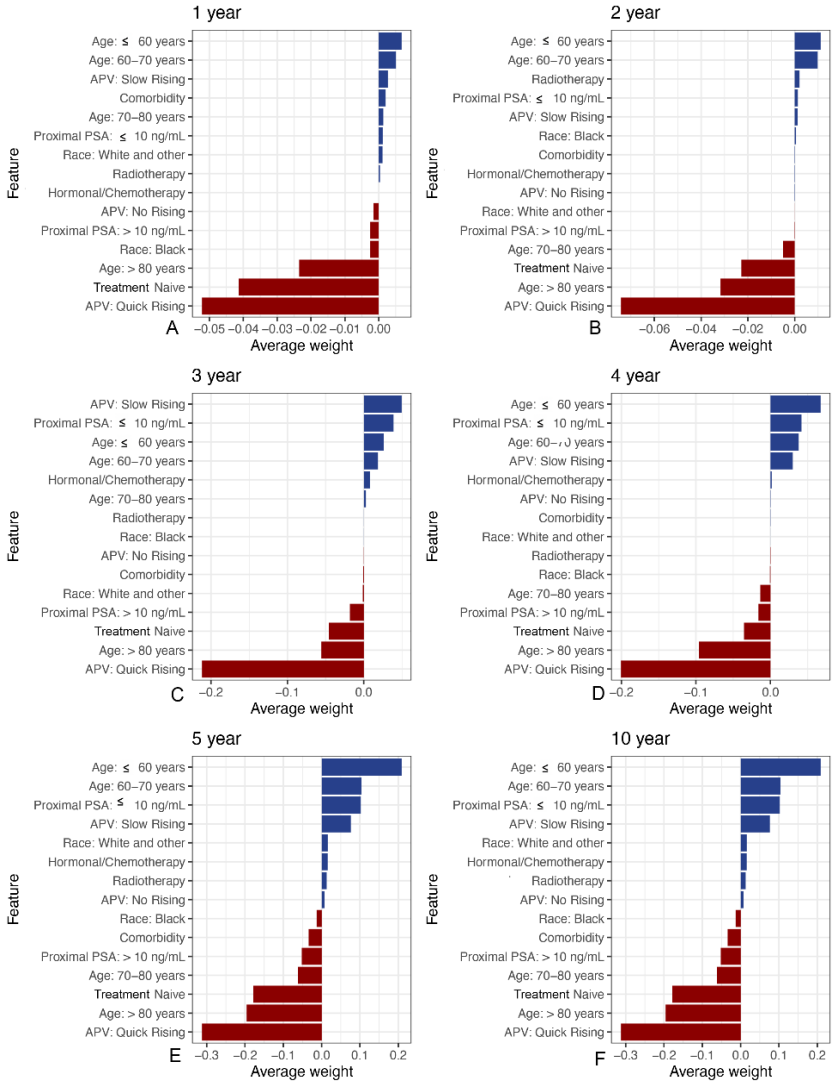

Figure 2 A-F. Decision curve analyses of each of the 6 models designed to estimate patient survival at (A) 1 year, (B) 2 years, (C) 3 years, (D) 4 years, (E) 5 years, and (F) 6 years after treatment or surgery for skeletal-related events due to bone metastasis from prostate cancer. These results suggest that all the models (dotted line) should be used rather than assuming all patients (continuous line) or no patients (thick continuous line) will survive longer than the period of each predictive model.