UNC SCHOOL OF MEDICINE

AUA 2020: PD32-02

What Really Matters When Predicting Other Cause Mortality for Men with Prostate Cancer: A Machine Learning Approach to Variable Selection

Brooke Namboodri Spratte, BS; Xi Zhou, MS; Stephen McMahon, BS; Matthew E. Nielsen, MD, MS; Jennifer Lund, PhD; Alex Sox-Harris, PhD, MS; Angela B. Smith, MD, MS; Ethan Basch, MD, MS; Hung-Jui Tan, MD, MSHPM

Acknowledgement

Brooke Namboodri Spratte was supported by Summer Medical Student Fellowship Program sponsored by the AUA through the support of the Herbert Brendler, MD, Research Fund.

Hung-Jui Tan, MD, MSHPM was supported by a Mentored Research Scholar Grant in Applied and Clinical Research, MRSG-18-193-01-CPPB, from the ACS as well as the NIH Loan Repayment Program.

These funding sources had no role in the design, conduct, analysis, or decision to publish the manuscript.

Prostate Cancer

Figure 3. Leading Sites of New Cancer Cases and Deaths – 2020 Estimates

Estimated New Cases

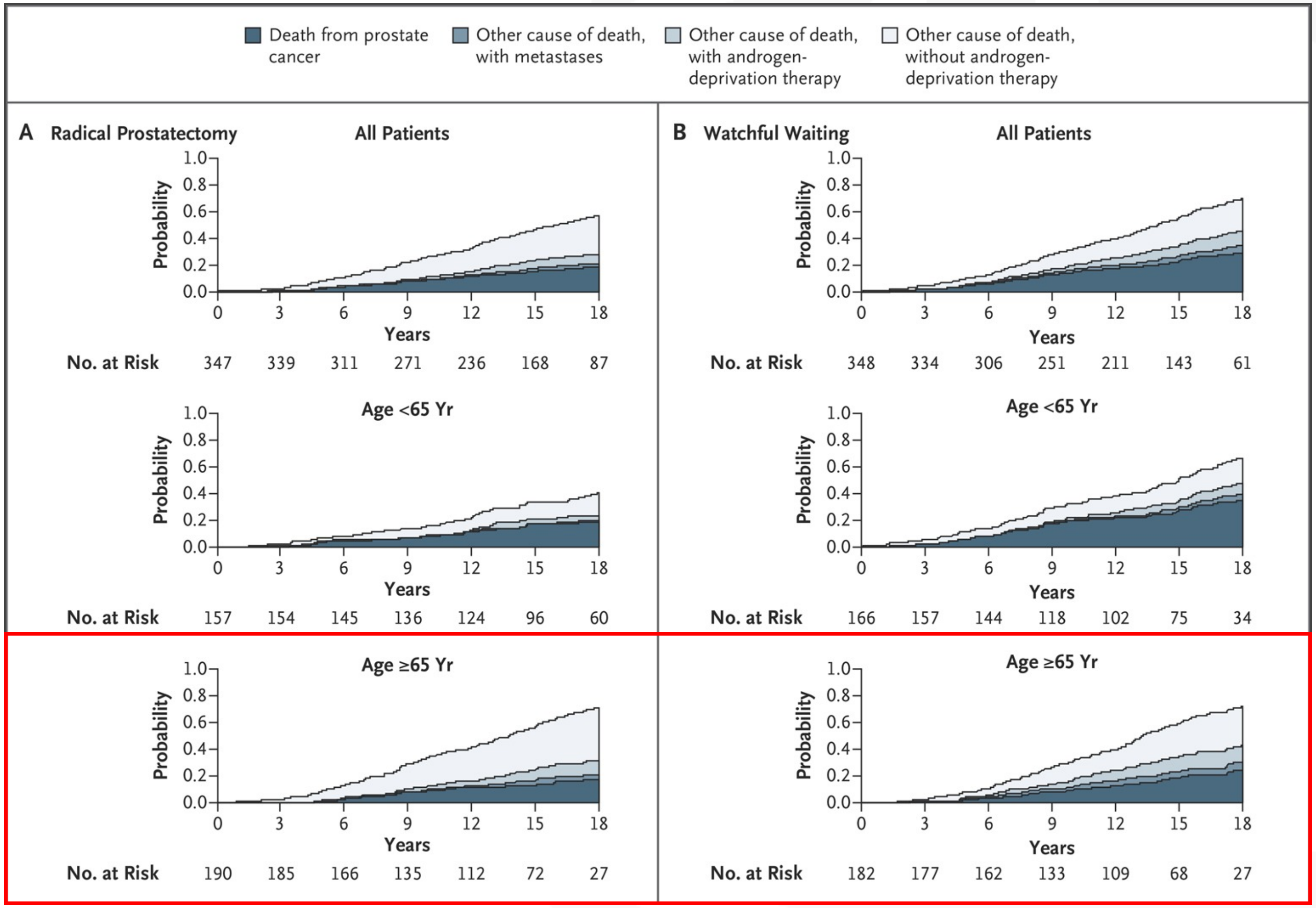
Male			Female		
Prostate	191,930	21%	Breast	276,480	30%
Lung & bronchus	116,300	13%	Lung & bronchus	112,520	12%
Colon & rectum	78,300	9%	Colon & rectum	69,650	8%
Urinary bladder	62,100	7%	Uterine corpus	65,620	7%
Melanoma of the skin	60,190	7%	Thyroid	40,170	4%
Kidney & renal pelvis	45,520	5%	Melanoma of the skin	40,160	4%
Non-Hodgkin lymphoma	42,380	5%	Non-Hodgkin lymphoma	34,860	4%
Oral cavity & pharynx	38,380	4%	Kidney & renal pelvis	28,230	3%
Leukemia	35,470	4%	Pancreas	27,200	3%
Pancreas	30,400	3%	Leukemia	25,060	3%
All sites	**893,660**		**All sites**	**912,930**	

Estimated Deaths

Male			Female		
Lung & bronchus	72,500	23%	Lung & bronchus	63,220	22%
Prostate	33,330	10%	Breast	42,170	15%
Colon & rectum	28,630	9%	Colon & rectum	24,570	9%
Pancreas	24,640	8%	Pancreas	22,410	8%
Liver & intrahepatic bile duct	20,020	6%	Ovary	13,940	5%
Leukemia	13,420	4%	Uterine corpus	12,590	4%
Esophagus	13,100	4%	Liver & intrahepatic bile duct	10,140	4%
Urinary bladder	13,050	4%	Leukemia	9,680	3%
Non-Hodgkin lymphoma	11,460	4%	Non-Hodgkin lymphoma	8,480	3%
Brain & other nervous system	10,190	3%	Brain & other nervous system	7,830	3%
All sites	**321,160**		**All sites**	**285,360**	

Estimates are rounded to the nearest 10, and cases exclude basal cell and squamous cell skin cancers and in situ carcinoma except urinary bladder. Estimates do not include Puerto Rico or other US territories. Ranking is based on modeled projections and may differ from the most recent observed data.

What Do Prostate Cancer Patients Die Of?



2. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. *N Engl J Med*. 2014;370(10):932–942.

Current Tools to Evaluate Life Expectancy

Tool	Year	Variables	Number of Inputs
Cowen et al.[3]	2006	Age, Charlson Comorbidity Index, angina, systolic blood pressure, BMI, smoking, marital status, PSA, Gleason sum, clinical stage, treatment	27
Walz et al.[4]	2007	Age, Charlson comorbidity index, treatment type (radical prostatectomy vs external beam radiotherapy)	19
Hoffman et al.[5]	2015	Age, race, and patient-reported overall health	3
Daskivich et al.[6]	2015	Age, race, treatment, PSA, Gleason Score, cancer stage, and the Prostate Cancer-specific Comorbidity Index	30
Cho/Hawkens et al.[7]	2017	Age, race, Charlson Comorbidity Index	19

Only 23% of urologists report using life expectancy tools in practice.[8]

Enthusiasm for Big Data

Front Oncol. 2016; 6: 149.
Published online 2016 Jun 14. doi: 10.3389/fonc.2016.00149
PMCID: PMC4905980
PMID: 27379211

Big Data Analytics for Prostate Radiotherapy

James Coates,[1,*] Luis Souhami,[2] and Issam El Naqa[3]

Author information Article no

J Am Med Inform Assoc. 2015 Nov; 22(6): 1114.
Published online 2015 Nov 9. doi: 10.1093/jamia/ocv136
PMCID: PMC5009910
PMID: 26555016

The NIH Big Data to Knowledge (BD2K) initiative

Philip E Bourne,[1,*] Vivien Bonazzi,[1] Michelle Dunn,[1] Eric D Green,[2] Mark Guyer,[1] George Komatsoulis,[3] Jennie Larkin,[1] and Beth Russell[1]

Author information Copyright and License information Disclaimer

N C Med J. Author manuscript;
Published in final edited form a
N C Med J. 2014 Jul-Aug; 75

Big Data for Population-Based Cancer Research

The Integrated Cancer Information and Surveillance System

Anne-Marie Meyer, PhD, research assistant professor, Andrew F. Laura Green, MBA, project manager, Adrian Meyer, MS, director PhD, MPH, assistant professor, Ethan Basch, MD, MSc, director, director

JAMA. 2018 Apr 3;319(13):1317-1318. doi: 10.1001/jama.2017.18391.

Big Data and Machine Learning in Health Care.

Beam AL[1], Kohane IS[1].

Author information

PMID: 29532063 DOI: 10.1001/jama.2017.18391

Aim of the Study

To identify the most influential variables for predicting other cause mortality for men newly diagnosed with prostate cancer.

Methods

- Study Sample: SEER-CAHPS data
  - Men 65 years and older diagnosed with prostate cancer from 2004 to 2013.
- Primary Outcome: Defined as expired from causes other than prostate cancer.
- Potential Predictive Variables (76 total): Included patient demographics (7), cancer information (4), claims-based measures (60), and patient-reported health measures (5).

LASSO Regression

- Applied LASSO regression to identify the core set of variables from **76 potential inputs** that minimize prediction error for other-cause mortality.
- LASSO regression is often used in machine learning
  - Uses a shrinkage and variable selection method to identify **subset of predictive variables** that **minimize prediction error**

Passing Grade = y interecept + [slope1 x **Study Time**] + [slope2 x **Sleep**] + [slope3 x **Sign**] + [slope4 x **Mile Pace**]

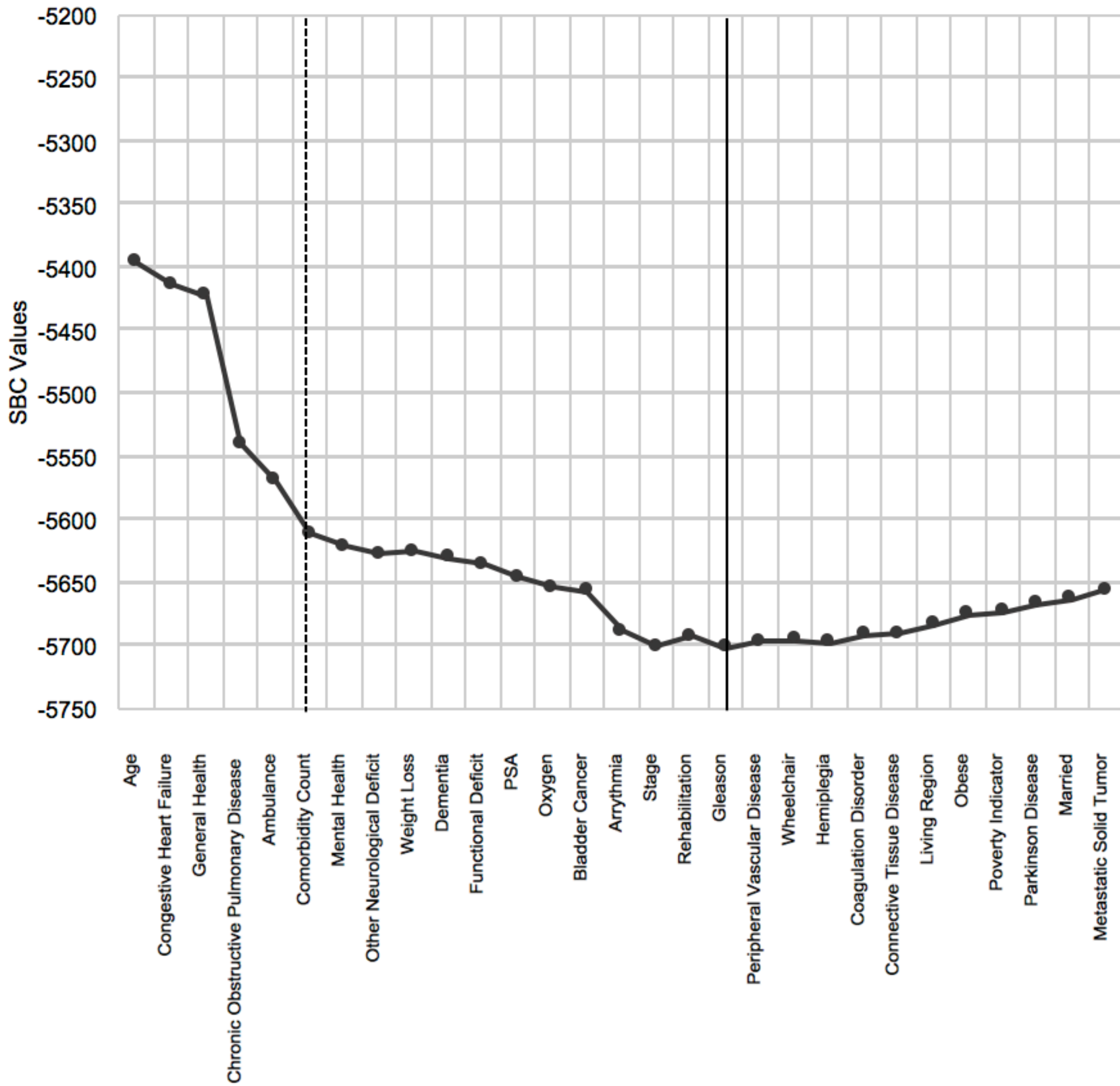
If slope3 + λ(slope3) = 0 *and* slope4 + λ(slope4) = 0

Passing Grade = y interecept + [slope1 x **Study Time**] + [slope2 x **Sleep**] + ~~[0 x **Sign**] + [0 x **Mile Pace**~~]

Results

- Among 3,240 men diagnosed with prostate cancer, 246 (7.62%) died of prostate cancer and 631 (19.48%) died of other causes.
- LASSO regression identified an 18-variable model:
  - 1 demographic variable
  - 3 cancer variables
  - 10 claims-based variables
  - 4 patient-reported variables

Figure 1. Fit statistics for the top 25 variables selected by the lasso regression model.



Top 6 Variables:
- Age
- CHF
- General health
- COPD
- Ambulance
- Comorbidity count

	10-yr AUC
6-variable	0.758
18-variable	0.783
Other models:	
-Hoffman	0.738
-Cho	0.746
-Daskivich	0.783

Conclusions

- Estimating other-cause mortality in men with prostate cancer can be accurately accomplished by using relatively few data inputs.

- Incorporating different types of data in combination with novel machine learning techniques may produce more parsimonious tools that facilitate usability.

References

1. American Cancer Society. Cancer Facts & Figures 2020. Atlanta: American Cancer Society; 2020.
2. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. *N Engl J Med*. 2014;370(10):932–942.
3. Cowen ME, Halasyamani LK, Kattan MW. Predicting life expectancy in men with clinically localized prostate cancer. J Urol. 2006;175:99–103.
4. Walz J, Gallina A, Saad F, et al. A nomogram predicting 10-year life expectancy in candidates for radical prostatectomy or radiotherapy for prostate cancer. J Clin Oncol. 2007;25:3576–81.
5. Hoffman RM. Self-reported health status predicts other-cause mortality in men with localized prostate cancer: Results from the prostate cancer outcomes study. J Gen Intern Med. 2015;30(7):924-934.
6. Daskivich TJ, Kwan L, Dash A, et al. An age adjusted comorbidity index to predict long-term, other cause mortality in men with prostate cancer. J Urol. 2015;194:73-8.
7. Hawken SR, Auffenberg GB, Miller DC, et al. Calculating life expectancy to inform prostate cancer screening and treatment decisions. BJU Int. 2017;120:9-11.
8. Kim SP, Karnes RJ, Nguyen PL, et al. Clinical implementation of quality of life instruments and prediction tools for localized prostate cancer: results from a national survey of radiation oncologists and urologists. J Urol. 2013; 189(6):2092–2098.