JU Insight

THE JOURNAL of UROLOGY®
www.auajournals.org/journal/juro

Development and Validation of Models to Predict Pathological Outcomes of Radical Prostatectomy in Regional and National Cohorts

Erkin Ötleş, Brian T. Denton, Bo Qu et al.

Correspondence: Karandeep Singh (email: kdpsingh@umich.edu).

Full-length article available at www.auajournals.org/10.1097/JU.0000000000002230.

Study Need and Importance: Prediction models are recommended by National Comprehensive Cancer Network® guidelines to support clinical decision making in prostate cancer; however, existing models to predict pathological outcomes of radical prostatectomy have been developed using data from tertiary care centers and may not generalize well to other settings. Data from a regional cohort (Michigan Urological Surgery Improvement Collaborative [MUSIC]) were used to develop models to predict extraprostatic extension, seminal vesicle invasion, lymph node invasion (LNI), and nonorgan-confined disease in patients undergoing radical prostatectomy. The MUSIC models were compared against the widely used Memorial Sloan Kettering models, Partin tables, and Briganti nomogram using data from a national cohort (Surveillance, Epidemiology, and End Results [SEER]).

What We Found: We identified 7,491 eligible patients in the SEER registry. The MUSIC model had good discrimination and was well calibrated. While the Memorial Sloan Kettering models had similar discrimination to the MUSIC models, in terms of calibration they overestimated the risk of extraprostatic extension, LNI, and nonorgan-confined disease. The Partin tables had inferior discrimination as compared to other models. When evaluated on the LNI outcome, the Briganti nomogram displayed good discrimination but overestimated risk.

Limitations: Neither the SEER registry nor the MUSIC registry have centralized pathology, so there may be variations in the quality of pathological reporting. The SEER registry population may not necessarily be nationally representative, as it oversamples western states as compared to the rest of the U.S. However, the SEER and MUSIC cohort results were largely concordant, which supports the notion that patient sampling did not play a large role in the findings.

Interpretation for Patient Care: Our study is the first to evaluate the most widely adopted models predicting prostate cancer pathology in national registry data. New models developed using the MUSIC registry outperform existing models in both national and regional registries and thus should be considered as replacements.

0022-5347/22/2072-0358/0
THE JOURNAL OF UROLOGY®

https://doi.org/10.1097/JU.0000000000002230
Vol. 207, 358-366, February 2022
Printed in U.S.A.

THE JOURNAL of UROLOGY®
www.auajournals.org/journal/juro

Development and Validation of Models to Predict Pathological Outcomes of Radical Prostatectomy in Regional and National Cohorts

Erkin Ötleş,[1,2] Brian T. Denton,[1,3] Bo Qu,[1] Adharsh Murali,[4,*] Selin Merdan,[1] Gregory B. Auffenberg,[5] Spencer C. Hiller,[3] Brian R. Lane,[6] Arvin K. George[3] and Karandeep Singh[3,4,7,8,†] for the Michigan Urological Surgery Improvement Collaborative

[1]Department of Industrial & Operations Engineering, University of Michigan College of Engineering, Ann Arbor, Michigan
[2]Medical Scientist Training Program, University of Michigan Medical School, Ann Arbor, Michigan
[3]Department of Urology, University of Michigan Medical School, Ann Arbor, Michigan
[4]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, Michigan
[5]Department of Urology, Northwestern University Feinberg School of Medicine, Chicago, Illinois
[6]Division of Urology, Spectrum Health, Grand Rapids, Michigan
[7]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan
[8]School of Information, University of Michigan, Ann Arbor, Michigan

Purpose: Prediction models are recommended by national guidelines to support clinical decision making in prostate cancer. Existing models to predict pathological outcomes of radical prostatectomy (RP)—the Memorial Sloan Kettering (MSK) models, Partin tables, and the Briganti nomogram—have been developed using data from tertiary care centers and may not generalize well to other settings.

Materials and Methods: Data from a regional cohort (Michigan Urological Surgery Improvement Collaborative [MUSIC]) were used to develop models to predict extraprostatic extension (EPE), seminal vesicle invasion (SVI), lymph node invasion (LNI), and nonorgan-confined disease (NOCD) in patients undergoing RP. The MUSIC models were compared against the MSK models, Partin tables, and Briganti nomogram (for LNI) using data from a national cohort (Surveillance, Epidemiology, and End Results [SEER] registry).

Results: We identified 7,491 eligible patients in the SEER registry. The MUSIC model had good discrimination (SEER AUC EPE: 0.77; SVI: 0.80; LNI: 0.83; NOCD: 0.77) and was well calibrated. While the MSK models had similar discrimination to the MUSIC models (SEER AUC EPE: 0.76; SVI: 0.80; LNI: 0.84; NOCD: 0.76), they overestimated the risk of EPE, LNI, and NOCD. The Partin tables had inferior discrimination (SEER AUC EPE: 0.67; SVI: 0.76; LNI:

Abbreviations and Acronyms

AUC = area under the receiver operating characteristic curve
EPE = extraprostatic extension
LNI = lymph node invasion
MSK = Memorial Sloan Kettering
MUSIC = Michigan Urological Surgery Improvement Collaborative
NOCD = nonorgan-confined disease
PCa = prostate cancer
PSA = prostate specific antigen
RP = radical prostatectomy
SEER = Surveillance, Epidemiology, and End Results Program
SVI = seminal vesicle invasion

Accepted for publication August 18, 2021.
Funding: EÖ was supported by NIH grant T32GM007863. BD and BQ were supported by the National Science Foundation, Grant No. CMMI-1536444. The Michigan Urological Surgery Improvement Collaborative is funded by the Blue Cross Blue Shield of Michigan.
Author Contributions: Erkin Ötleş: Conceptualization, validation, draft writing, critical revisions. Brian Denton: Conceptualization, supervision, critical revisions. Bo Qu: Conceptualization, data analysis, critical revisions. Adharsh Murali: Data curation, data analysis, critical revisions. Selin Merdan: Conceptualization, methodology, supervision, draft writing. Gregory Auffenberg: Conceptualization, supervision, critical revisions. Spencer Hiller: Conceptualization, critical revisions. Brian R. Lane: Conceptualization, critical revisions. Arvin K. George: Conceptualization, supervision, critical revisions. Karandeep Singh: Conceptualization, data curation, data analysis, draft writing, critical revisions, methodology, project administration, supervision.
Each MUSIC practice obtained an exemption or approval for collaborative participation from a local institutional review board.
* Deceased.
† Correspondence: Department of Learning Health Sciences, University of Michigan Medical School, 1161H NIB, 300 N. Ingalls St., Ann Arbor, Michigan 48109 (telephone: 734-936-1649; FAX: 734-647-3914; email: kdpsingh@umich.edu).

Editor's Note: This article is the third of 5 published in this issue for which category 1 CME credits can be earned. Instructions for obtaining credits are given with the questions on pages 481 and 482.

0022-5347/22/2072-0359/0
THE JOURNAL OF UROLOGY®

https://doi.org/10.1097/JU.0000000000002230
Vol. 207, 358-366, February 2022
Printed in U.S.A.

0.69; NOCD: 0.72) as compared to other models. The Briganti LNI nomogram had an AUC of 0.81 in SEER but overestimated the risk.

Conclusions: New models developed using the MUSIC registry outperformed existing models and should be considered as potential replacements for the prediction of pathological outcomes in prostate cancer.

Key Words: prostatic neoplasms, prostatectomy, clinical decision rules

PREDICTION of pathological outcomes plays an important role in preoperative counseling for men with prostate cancer (PCa) considering radical prostatectomy (RP). In patients with a very low risk of lymph node invasion (LNI), active surveillance may be preferred over RP, and even if RP is determined to be necessary, pelvic lymph node dissection may not be needed, thereby avoiding potential morbidity. On the other hand, patients at high risk of extraprostatic extension (EPE) or seminal vesicle invasion (SVI) may be advised that nerve-sparing RP may not be possible or advisable. Nerve-sparing surgery is associated with better postoperative sexual function,[1] and inability to nerve-spare may have implications on the patients' quality of life that will need to be addressed through preoperative counseling and shared decision making.

The use of prediction models to support clinical decision making is ubiquitous in PCa due to national guidelines recommending their use[2,3] and the availability of well-validated models. Three sets of models that are widely adopted to predict pathological outcomes are the Memorial Sloan Kettering (MSK)[4] models, the Johns Hopkins University Partin table,[5] and the Briganti nomogram (for LNI).[6] While the Briganti nomogram is limited to LNI, the others consist of 4 separate models to predict EPE, SVI, LNI, and the presence of any of the above, which indicates that the PCa has extended beyond the prostate and constitutes nonorgan-confined disease (NOCD). All 3 sets of models were developed using data from tertiary care centers whose populations may not be representative of other prostatectomy populations. Although MSK, Johns Hopkins, and Università Vita-Salute San Raffaele (for the Briganti nomogram) serve demographically diverse communities, these hospitals also care for some of the most complex cases. As a result, the average risk profile of patients at these institutions is likely much higher than the average patient evaluated for RP in the U.S. Models developed from patient cohorts with higher case complexity or acuity are known to overestimate the risk of adverse outcomes in settings with lower complexity, a phenomenon known as model miscalibration.[7] Thus, broad adoption of existing models may lead to patients undergoing unnecessary pelvic lymph node dissection due to potential miscalibration in nontertiary-care settings.

Although miscalibration of these models in a national cohort is a concern, prior external validation efforts using these models have largely ignored calibration[8] or focused on miscalibration at only a single center.[9] However, because urologists rely on the absolute risk estimates to make clinical decisions (eg whether to perform a pelvic lymph node dissection), the need for a well-calibrated model is critical. Concerned that existing models may be miscalibrated in national cohorts, we developed new models in our regional cohort of 50 urology practices participating in the Michigan Urological Surgery Improvement Collaborative (MUSIC) and compared these against existing models in the national Surveillance, Epidemiology, and End Results (SEER) registry.

MATERIALS AND METHODS

Data Sources

Two data sources were used for this study: the regional MUSIC registry and the national SEER registry. Additional details are provided in the supplementary methods (https://www.jurology.com).

Study Cohorts

From the MUSIC and SEER registries, we established 3 study cohorts: the MUSIC derivation cohort, the MUSIC validation cohort, and the SEER validation cohort (fig. 1). The 2 MUSIC cohorts were established using 2:1 random sampling stratified by practice, with two-thirds of patients assigned to the derivation cohort and a third to the validation cohort. After internally validating the MUSIC models on this validation cohort, we evaluated the MUSIC models on the SEER registry data. As a comparator, we also evaluated the performance of the MSK models, Partin tables, and the Briganti nomogram (for LNI) in both validation cohorts.

In both registries, we included patients in whom the prostate specific antigen (PSA), clinical T-stage, and biopsy information were available, including Gleason score and number of positive and negative biopsy cores. Patients without pathological outcomes data available were excluded. A small subset of MUSIC patients in whom the date of surgery was missing were also excluded.

Outcomes

We evaluated the models' ability to predict each of the 3 pathological outcomes both individually (EPE, SVI, and LNI) and as a group (NOCD). The MSK models and Partin tables have separate models for all 4 outcomes, and the Briganti nomogram is limited to LNI. For the newly developed MUSIC models, separate models were fit to each of these outcomes in a comparable fashion.

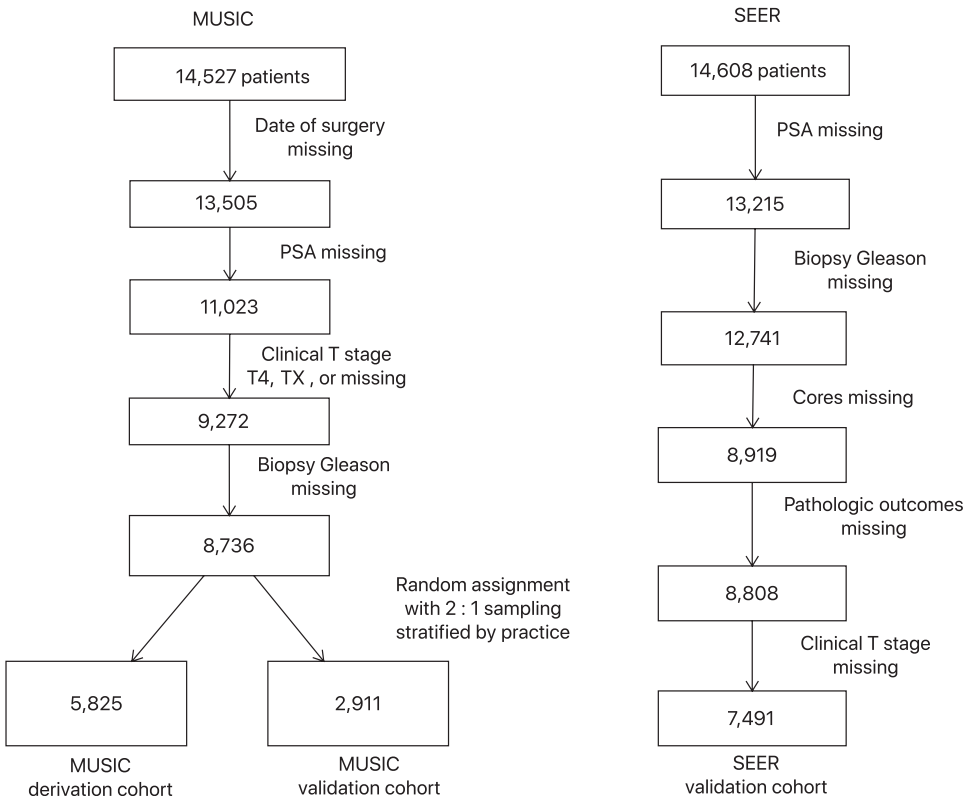

Figure 1. Flowchart of patient inclusion/exclusion criteria.

MUSIC Model Development

We hypothesized that new models fit using our regional cohort, which includes a diverse set of urology practices, would be better calibrated on a national sample (SEER) than existing models. Logistic regression models for each of the pathological outcomes (EPE, SVI, LNI, NOCD) were fit using the MUSIC derivation cohort with the following predictors: age, PSA, clinical T-stage, grade group, and the number of positive and negative cores.

MSK Models, Partin Tables and the Briganti Nomogram

The MSK models were originally developed and validated using data from MSK Cancer Center.[10] Because the models have previously been shown to become miscalibrated over time,[9] they are dynamically updated.[11] The MSK models predict pathological outcomes using PSA, clinical T-stage, biopsy Gleason score (primary and secondary), and number of positive and negative cores. Although separate MSK models are available for patients who lack data about biopsy cores, we focused our evaluation on the models that included core data. The last published evaluation of the MSK models (specifically, one focused on LNI) was in 2011.[9] We used the coefficients from the 2018 version of the model.[4]

The Partin tables were originally developed and validated using data from Johns Hopkins University. Although the 2007 version of the Partin tables have been evaluated in SEER data,[8] this evaluation did not consider calibration and does not reflect contemporary practice. The Partin tables were most recently updated using patient data from 2010 to 2015 and predict pathological outcomes using PSA, clinical T-stage, and Gleason grade group, and this latest version was used in our evaluation.[5] The Partin tables do not include the number of positive or negative cores as predictors.

The Briganti nomogram refers to a set of several models developed by a research group at the Università Vita-Salute San Raffaele between 2006 and 2019.[6,12–14] For this evaluation, we selected the primary model (ie "Model 1") from the Briganti 2017 nomogram[6] because this was the latest version of the model that did not require magnetic resonance imaging data, which was not available in our cohorts.

Model Validation

We evaluated all models in the MUSIC validation cohort (for internal validation) and in the SEER validation cohort (for external validation). Performance of these models was characterized in terms of both discrimination and calibration. Discriminative performance was measured using the area under the receiver operating characteristic curve (AUC), and calibration was assessed visually by comparing deciles of predicted risk with observed risk. Bootstrapped 95% confidence intervals were created for the AUC by resampling (unstratified with replacement) the cohort populations 1,000 times. Patients with clinical T3 disease were excluded from evaluations of the Partin tables due to the absence of T3 disease in the Partin table.[5] The Briganti nomogram requires percentage of positive cores with highest-grade and lower-grade disease as predictors. Because these were not

available in either cohort, we imputed the median values of 29.4 and 27.7, respectively, from the original publication describing the Briganti nomogram's development.[6]

Missing Data

After excluding patients with missing PSA, clinical T-stage, and biopsy information (fig. 1), the remaining variables were nonparametrically imputed with bagged trees for the MUSIC models only. Imputation was not required for the MSK models or Partin tables due to complete availability of predictors.

Net Benefit

Decision curves were used to calculate the net benefit of all models in the SEER validation cohort. The analysis was focused only on LNI because of previously published threshold ranges of 0%–20% risk.[6] The potential clinical impact of the models was examined by comparing the number of patients who would be recommended to undergo lymph node dissection (in the threshold range of 0%–20%) against those who actually had LNI.

Software

We used R 3.6.0 for all analyses (R Foundation for Statistical Computing, Vienna, Austria). The model code is available on GitHub,[15] and the MUSIC model is available as an interactive web calculator.[16] The model coefficients are available in supplementary tables 1–4 (https://www.jurology.com).

RESULTS

We identified 8,736 eligible patients in the MUSIC registry and 7,491 eligible patients in the SEER registry for our study cohorts. In the MUSIC registry, 5,825 (67%) were randomly assigned to a derivation cohort and 2,911 (33%) were randomly assigned to a validation cohort (fig. 1). The SEER validation cohort had a higher proportion of grade group 1 PCa (24% versus 18%), a lower number of positive cores (median 5 positive cores in SEER versus 7 and 8 in the MUSIC derivation and validation cohorts, respectively), and a higher proportion of patients with a cT1 disease (77% in SEER versus 71% and 72% in the MUSIC derivation and validation cohorts, respectively; table 1). While proportions of EPE, SVI, and LNI were fairly similar between the cohorts, fewer patients in SEER had NOCD (37% in SEER versus 45% and 43% in the MUSIC derivation and validation cohorts, respectively) due to more overlap among the individual outcomes (table 2).

Internal Validation (MUSIC Registry)

In the MUSIC validation cohort, the MUSIC models had better discrimination than the other models for EPE and NOCD and similar performance to other models for SVI and LNI (table 3). The MSK models overestimated risk of EPE, LNI, and NOCD, and the Briganti nomogram overestimated the risk of LNI, while both the MUSIC models and Partin tables were generally well calibrated (supplementary fig. 1, https://www.jurology.com).

Table 1. *Patient characteristics by cohort*

Characteristic	MUSIC Derivation	MUSIC Validation	SEER Validation
No. pts	5,825	2,911	7,491
Median yrs age (IQR)	63 (58, 68)	64 (58, 68)	62 (57, 67)
Median ng/mL PSA (IQR)	6.0 (4.5, 8.7)	6.0 (4.6, 8.9)	6.5 (4.9, 9.6)
No. primary Gleason (%):			
2	1 (<0.1)	0 (0)	1 (<0.1)
3	3,518 (60)	1,785 (61)	4,658 (62)
4	2,207 (38)	1,081 (37)	2,695 (36)
5	99 (1.7)	45 (1.5)	137 (1.8)
No. secondary Gleason (%):			
2	1 (<0.1)	0 (0)	0 (0)
3	2,265 (39)	1,098 (38)	3,203 (43)
4	3,136 (54)	1,613 (55)	3,672 (49)
5	423 (7.3)	200 (6.9)	616 (8.2)
No. International Society of Urological Pathology Grade Group (Gleason score) (%):			
1 (≤6)	1,047 (18)	515 (18)	1,814 (24)
2 (3+4)	2,434 (42)	1,253 (43)	2,750 (37)
3 (4+3)	1,211 (21)	577 (20)	1,370 (18)
4 (8)	683 (12)	354 (12)	953 (13)
5 (9–10)	450 (7.7)	212 (7.3)	604 (8.1)
No. pos cores (IQR)/No. missing	7 (5, 10)/13	8 (5, 10)/6	5.0 (3, 7)/0
No. neg cores (IQR)/No. missing	7 (5, 10)/13	8.0 (5, 10)/7	8.0 (5, 10)/0
No. clinical T stage (%):			
1	4,154 (71)	2,087 (72)	5,771 (77)
2a	847 (15)	449 (15)	673 (9.0)
2b	455 (7.8)	190 (6.5)	286 (3.8)
2c	304 (5.2)	157 (5.4)	449 (6.0)
3	65 (1.1)	28 (1.0)	312 (4.2)

External Validation (SEER Registry)

In the SEER validation cohort, the MUSIC models had similar discriminative performance as compared to the MSK models and the Briganti LNI nomogram, and all outperformed the Partin tables (table 3). However, the MSK models again overestimated the risk of EPE, LNI, and NOCD and the Briganti nomogram overestimated the risk of LNI, whereas the MUSIC models and Partin tables remained well calibrated (fig. 2).

Net Benefit

In the threshold range of 0%–20% risk (to perform a lymph node dissection), the MUSIC model achieved the highest net benefit across this range in the SEER validation cohort, although the difference was modest as compared to the MSK model (fig. 3). The Partin

Table 2. *Prevalence of patient outcomes by cohort*

Outcome	MUSIC Derivation	MUSIC Validation	SEER Validation
No. pts	5,825	2,911	7,491
No. EPE (%)	2,028 (35)	978 (34)	2,697 (36)
No. SVI (%)/No. missing	674 (12)/57	316 (11)/33	949 (13)/5
No. LNI (%)/No. missing	268 (5.9)/1,245	102 (4.4)/590	458 (6.1)/25
No. NOCD (%)/No. missing	2142 (45)/1,027	1026 (43)/503	2757 (37)/19

Percentages represent proportion of population with outcome out of patients with available findings for specified outcome; these percentages are not mutually exclusive.

Table 3. *Model performance with bootstrapped 95% confidence intervals*

Models Outcomes	MUSIC AUC (95% CI)	MSK AUC (95% CI)	Partin AUC (95% CI)	Briganti AUC (95% CI)
		MUSIC validation cohort (internal validation)		
EPE	0.77 (0.75–0.79)	0.70 (0.68–0.72)	0.66 (0.64–0.68)	—
SVI	0.82 (0.79–0.84)	0.81 (0.78–0.83)	0.77 (0.75–0.80)	—
LNI	0.82 (0.78–0.87)	0.81 (0.78–0.86)	0.78 (0.73–0.83)	0.81 (0.77–0.85)
NOCD	0.74 (0.72–0.76)	0.68 (0.65–0.70)	0.69 (0.67–0.71)	—
		SEER validation cohort (external validation)		
EPE	0.77 (0.76–0.78)	0.76 (0.75–0.77)	0.67 (0.66–0.69)	—
SVI	0.80 (0.79–0.82)	0.80 (0.78–0.81)	0.76 (0.74–0.78)	—
LNI	0.83 (0.81–0.85)	0.84 (0.82–0.85)	0.69 (0.66–0.72)	0.81 (0.79–0.83)
NOCD	0.77 (0.76–0.79)	0.76 (0.75–0.77)	0.72 (0.71–0.73)	—

table for LNI is not directly comparable to the others because exclusion of T3 disease (as described in the Methods) leads to a lower prevalence of LNI. A comparison of the potential clinical impact of using the MUSIC and MSK models is provided in supplementary figures 2 and 3 (https://www.jurology.com).

DISCUSSION

In this study, we found that newly developed MUSIC models outperformed existing models in the prediction of pathological outcomes following RP. While the MUSIC models had relatively similar AUCs to the MSK models and the Briganti

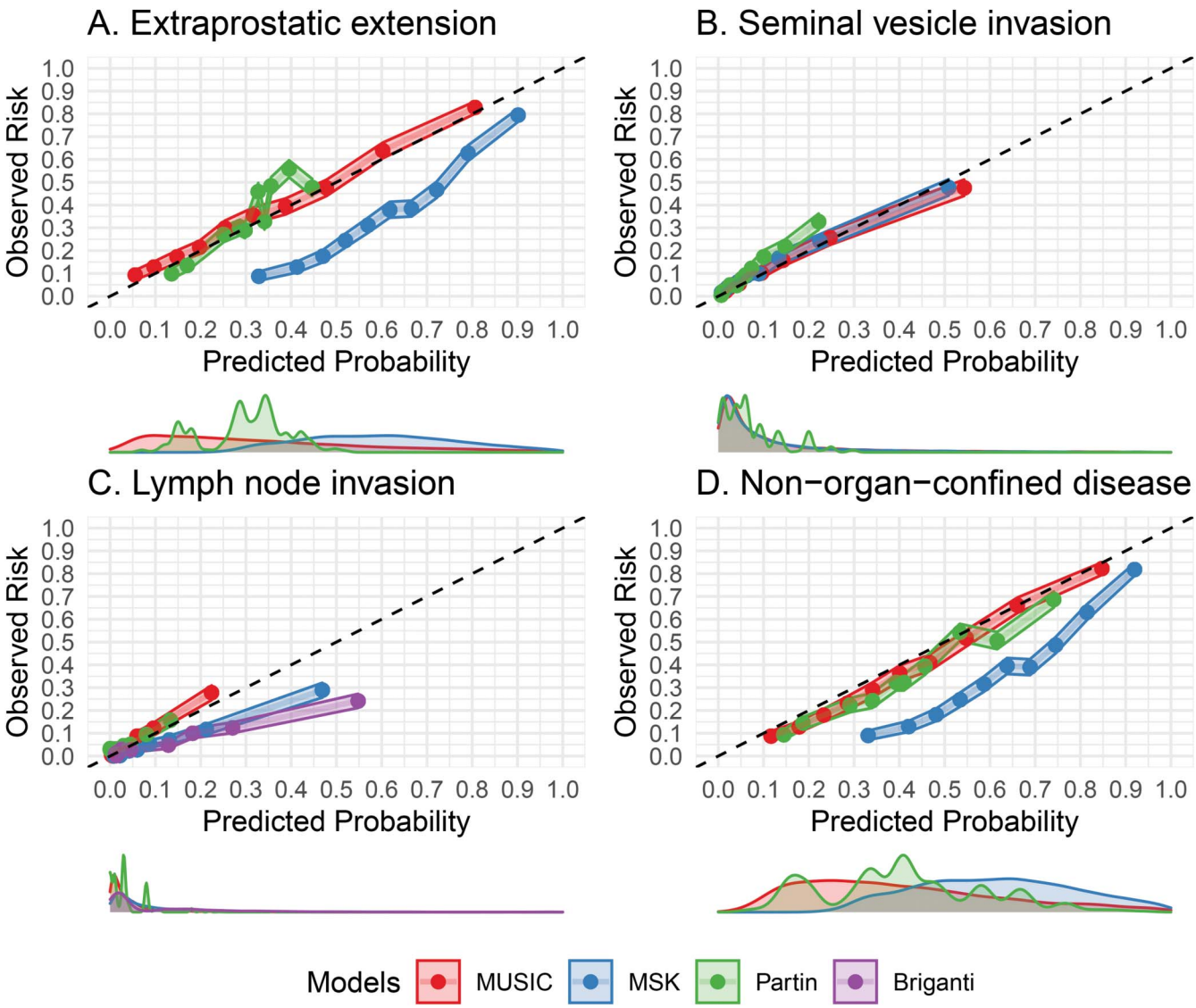

Figure 2. SEER validation cohort: calibration plot and distributions of MUSIC, MSK, Partin and Briganti models, with shaded 95% confidence intervals, with respect to EPE (*A*), SVI (*B*), LNI (*C*), and NOCD (*D*).

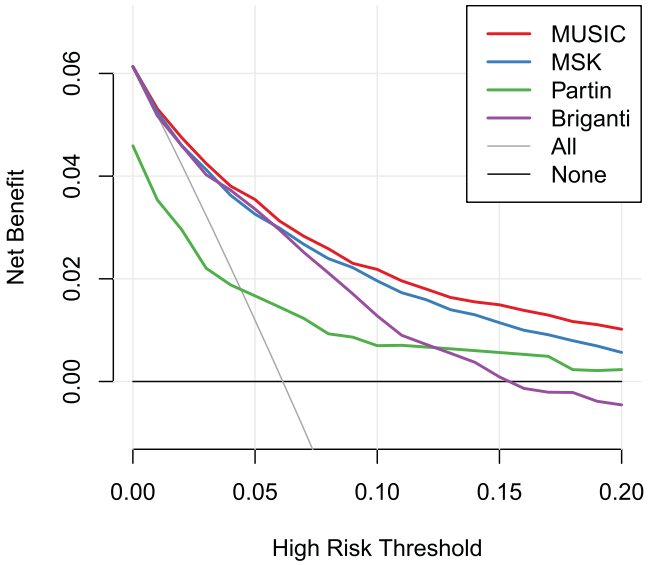

Figure 3. SEER validation cohort: decision curve analysis for outcome of LNI comparing MUSIC, MSK, Partin and Briganti models.

nomogram (for LNI), the MSK models overestimated the risk of EPE, LNI, and NOCD, and the Briganti nomogram overestimated the risk of LNI in both our internal validation and external validation cohorts. In contrast, the Partin tables were generally well calibrated but had inferior discrimination in both cohorts. The MUSIC models had the highest net benefit among all LNI models, though the difference between the MUSIC and MSK model was quite small.

Our findings for model discrimination are consistent with prior evaluations of the MSK models and Partin tables but not consistent with a prior published evaluation of the 2017 Briganti nomogram. An evaluation of the 2007 Partin tables (a prior version) on the 2005 SEER registry found AUCs of 0.62, 0.74, 0.77, and 0.68 for EPE, SVI, LNI, and NOCD, respectively.[8] Our evaluation of the updated Partin tables on 2015 SEER registry data found slight improvements in AUC for EPE (0.67) and NOCD (0.72), a similar AUC for SVI (0.76), and lower AUC for LNI (0.69). A prior meta-analysis evaluating LNI models based on 10,028 patients for MSK models and 69,681 patients for the Partin tables found pooled AUCs of 0.78 for both models.[17] Our evaluation of the MSK models and Partin tables found similar AUCs in the MUSIC validation cohort (0.81 and 0.78, respectively), although the Partin tables performed worse in the SEER cohort (AUC 0.69).

The original paper describing the Briganti paper found an AUC of 0.91, whereas our evaluation found an AUC of 0.81 in both validation cohorts. This difference could be due to both overestimation of model performance in the original publication (due to a small development cohort, 681 patients, and reuse of the same population for model validation), and underestimation in our evaluation (due to median imputation of percent highest-grade and lower-grade cores). However, the overestimation of risk from the Briganti nomogram was not caused by imputation because the overall percentage of positive cores in the original publication (which we used for imputation) is actually lower than the overall percentage of positive cores in the SEER registry (33.3%[6] vs 41.2%). Model miscalibration is a known problem in the setting of PCa more broadly.[18] Our finding that several models were miscalibrated on national registry data is important because national guidelines need to consider models on their impact broadly, and not only at the academic medical centers where the models were developed.

Our study has several limitations. Neither the SEER registry nor the MUSIC registry have centralized pathology, so there may be variations in the quality of pathological reporting. While the SEER registry contains a national sample of patients with PCa, this population may not necessarily be nationally representative. One known limitation of the SEER registry is that it oversamples western states as compared to the rest of the U.S. However, the fact that our results in the SEER registry were largely concordant with the MUSIC validation cohort—which includes 90% of urology practices in Michigan—supports the notion that patient sampling did not play a large role in the findings. On the other hand, our stringent inclusion criteria based on missingness of crucial information (such as PSA or biopsy Gleason grade) could have impacted our results if this missingness was informative because of significant reductions in cohort sizes. Particularly within MUSIC, where data are collected by trained abstractors with direct access to the urologists, we expect the missingness would have been noninformative.

Our study has national implications because the MSK models, Partin tables, and Briganti nomogram are widely adopted, and because several national guidelines recommend the use of risk stratification as part of preoperative counseling.[2,3] More broadly, our findings also have implications for other models developed using data from tertiary care referral centers. Even if the data are of high quality, selection bias may lead to nonrepresentative estimates of disease risk.

CONCLUSIONS

Our study provides the first external validation of recent MSK models, Partin tables, and the Briganti nomogram in national registry data. Finding the MSK models and Briganti nomogram to be miscalibrated, and the Partin tables to have lower discrimination than other models, our study offers an alternative in the form of newly developed MUSIC models. These models should be considered as potential replacements for the prediction of pathological outcomes in PCa.

ACKNOWLEDGMENTS

The corresponding author thanks all the support staff from the Michigan Urological Surgery Improvement Collaborative. Since the initial submission of this manuscript, our beloved co-author Adharsh Murali unexpectedly passed away. We give our love to Adharsh's family and friends.

REFERENCES

1. Agochukwu NQ, Wittmann D, Boileau NR et al: Validity of the patient-reported outcome measurement information system (PROMIS) sexual interest and satisfaction measures in men following radical prostatectomy. J Clin Oncol 2019; **37:** 2017.
2. 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.
3. National Comprehensive Cancer Network: NCCN guidelines prostate cancer version 4.2019. Available at https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf. Accessed August 11, 2021.
4. Memorial Sloan Kettering Cancer Center: MSKCC pre-radical prostatectomy nomogram. Available at https://www.mskcc.org/nomograms/prostate/pre_op. Accessed January 31, 2020.
5. Brady Urological Institute: Partin Tables Prostate Cancer Risk Assessment Tool. 2021. Available at https://www.hopkinsmedicine.org/brady-urology-institute/conditions_and_treatments/prostate_cancer/risk_assessment_tools/partin-tables.html. Accessed April 4, 2021.
6. Gandaglia G, Fossati N, Zaffuto E et al: Development and internal validation of a novel model to identify the candidates for extended pelvic lymph node dissection in prostate cancer. Eur Urol 2017; **72:** 632.
7. Van Calster B, McLernon DJ, van Smeden M et al: Calibration: the Achilles heel of predictive analytics. BMC Med 2019; **17:** 230.
8. Yu JB, Makarov DV, Sharma R et al: Validation of the partin nomogram for prostate cancer in a national sample. J Urol 2010; **183:** 105.
9. Godoy G, Chong KT, Cronin A et al: Extent of pelvic lymph node dissection and the impact of standard template dissection on nomogram prediction of lymph node involvement. Eur Urol 2011; **60:** 195.
10. Memorial Sloan Kettering Cancer Center Dynamic prostate cancer nomogram: coefficients. Available at https://www.mskcc.org/nomograms/prostate/pre_op/coefficients. Accessed April 4, 2021.
11. Vickers AJ, Kent M and Scardino PT: Implementation of dynamically updated prediction models at the point of care at a major cancer center: making nomograms more like Netflix. Urology 2017; **102:** 1.
12. Briganti A, Chun FKH, Salonia A et al: Validation of a nomogram predicting the probability of lymph node invasion based on the extent of pelvic lymphadenectomy in patients with clinically localized prostate cancer. BJU Int 2006; **98:** 788.
13. Briganti A, Larcher A, Abdollah F et al: Updated nomogram predicting lymph node invasion in patients with prostate cancer undergoing extended pelvic lymph node dissection: the essential importance of percentage of positive cores. Eur Urol 2012; **61:** 480.
14. Gandaglia G, Ploussard G, Valerio M et al: A novel nomogram to identify candidates for extended pelvic lymph node dissection among patients with clinically localized prostate cancer diagnosed with magnetic resonance imaging-targeted and systematic biopsies. Eur Urol 2019; **75:** 506.
15. Singh K: GitHub AskMUSIC source code. Available at https://github.com/ML4LHS/askmusic. Accessed April 4, 2021.
16. Michigan Urological Surgery Improvement Collaborative: Ask MUSIC radical prostatectomy pathologic outcomes app. Available at https://shiny.med.umich.edu/apps/kdpsingh/askmusic_prostate_path_outcomes/. Accessed April 4, 2021.
17. Cimino S, Reale G, Castelli T et al: Comparison between Briganti, Partin and MSKCC tools in predicting positive lymph nodes in prostate cancer: a systematic review and meta-analysis. Scand J Urol 2017; **51:** 345.
18. Auffenberg GB, Merdan S, Miller DC et al: Evaluation of prostate cancer risk calculators for shared decision making across diverse urology practices in Michigan. Urology 2017; **104:** 137.

EDITORIAL COMMENTS

In this manuscript, Ötleş and colleagues externally validated models predicting adverse pathological features at radical prostatectomy, which is important for guiding preoperative treatment planning. Both discrimination (separating out those with and without adverse pathological features) and calibration (how closely the predicted probability of adverse pathology approximates the observed probability in the population) were the focus.

The Partin Tables had inferior discrimination, likely due to not using data on number of positive/negative cores, as discrimination is mainly a property of the included variables.[1] The MUSIC model had good discrimination and calibration, while the MSK and Briganti nomograms overestimated risk. What are potential explanations for differences in calibration? The MSK and Briganti models were developed using tertiary referral center populations that are inherently different from the broader U.S./SEER population. As such, the MUSIC model, developed from a regional U.S. cohort, may have an advantage in generalizability. However, other factors may complicate the issue. For example, central tertiary center pathology review might be more reliable. In EORTC-22911, central review of radical prostatectomy pathology was more strongly predictive of biochemical progression compared to local assessments.[2] Furthermore, more extensive lymphadenectomies were performed in these tertiary center cohorts (reference 14 in article). As such, it is difficult to determine whether the MSK and Briganti models overestimate the risk of nodal

involvement or are occult nodal metastases relatively underdetected in the SEER cohort due to more limited lymphadenectomy and heterogenous pathology reporting.

So, which model to use? The answer is not straightforward. Nonetheless, with contemporary preoperative imaging practices, the bigger issue may be how best to integrate magnetic resonance imaging findings into these models (as has been put forth by the Briganti group; reference 14 in article). The MUSIC Collaborative may also have the infrastructure and data to develop and validate a similar model, so we should all look forward to further exciting work from this group.

Bimal Bhindi[1]
[1]*Section of Urology, Department of Surgery, University of Calgary Calgary, Alberta, Canada*

REFERENCES

1. Assel M, Sjoberg D, Elders A et al: Guidelines for reporting of statistics for clinical research in urology. J Urol 2019; **201:** 595.

2. van der Kwast TH, Collette L, Van Poppel H et al: Impact of pathology review of stage and margin status of radical prostatectomy specimens (EORTC trial 22911). Virchows Arch 2006; **449:** 428.

Risk prediction models form an important component of the care of patients with prostate cancer from the decision to undergo prostate biopsy to pretreatment counseling decisions, and to the nuances of treatment. For patients opting for radical prostatectomy, these nomograms have been used to estimate the risk of adverse pathological findings. These data may have important consequences for surgical planning and the extent of surgical extirpation and nerve sparing. However, existing nomograms have been derived at single referral centers, including Johns Hopkins where the Partin tables were derived, Memorial Sloan Kettering, and San Raffaele. Thus, while these nomograms are widely used and are recommended in guidelines, they may not appropriately capture risk in patients in the wider community. Typically, due to the phenomenon of model miscalibration, these models will overestimate the risk of adverse events.

In the present manuscript, the authors undertook 2 related but parallel exercises. First, they used the population-based cohort of patients in the Michigan Urological Surgery Improvement Collaborative (MUSIC) to derive a new risk model. They then compared this model to existing nomograms using the SEER cohort. As may be anticipated, their nomogram derived based on a population-based cohort outperformed those derived at single centers. Notably, the Partin tables had inferior discrimination and both MSK and Briganti nomograms overestimated the risks of adverse events.

While this nomogram has relatively favorable performance characteristics and is likely more suitable than existing nomograms for use by most urologists, many patients with prostate cancer will have undergone multi-parametric magnetic resonance imaging. Thus, nomograms such as the Briganti 2019 nomogram (reference 14 in article) which incorporate these data may allow for incrementally greater risk prediction. Future work that provides both the generalizability allowed from the MUSIC cohort and the additional prognostic information from multi-parametric magnetic resonance imaging will help to further advance the care of our patients.

Christopher J. D. Wallis[1] **and Zachary Klaassen**[2]
[1]*University of Toronto Toronto, Ontario, Canada*

[2]*Medical College of Georgia Augusta, Georgia*