Factors Influencing Selection of Active Surveillance for Localized Prostate Cancer

Jianyu Liu, Paul R. Womble, Selin Merdan, David C. Miller, James E. Montie, Brian T. Denton on behalf of the Michigan Urological Surgery Improvement Collaborative

OBJECTIVE To determine how well demographic and clinical factors predict the initiation of Active Surveillance (AS).

METHODS AS has been suggested as a way to reduce overtreatment of men who have prostate cancer; however, factors associated with the decision to choose AS are poorly quantified. Using the Michigan Urological Surgery Improvement Collaborative registry, we identified 2977 men with prostate cancer who made treatment decisions from January 1, 2012, through December 31, 2013. We used chi-square and Wilcoxon tests to examine the association between factors and initiation of AS. Logistic regression models were fit for D'Amico risk categories. Measures of model discrimination and calibration were estimated, including area under the curve (AUC) and Brier score (BS).

RESULTS Patient age, Gleason score, clinical T-stage, urology practice, and tumor volume (greatest percent of a core involved with cancer and proportion of positive cores) were associated with the decision to choose AS in the intermediate-risk cohort (AUC = 0.875, BS = 0.07) and the complete cohort (AUC = 0.89, BS = 0.10). Patient age, urology practice, and tumor volume were significant in the low-risk cohort (AUC = 0.71, BS = 0.22). The addition of urology practice increased AUC in the low-risk cohort from 0.71 to 0.76 and reduced BS from 0.22 to 0.21.

CONCLUSION The urology practice at which a patient is seen is an important predictor for whether patients will initiate AS. Predictions were least accurate for low-risk patients, suggesting that factors such as patient preference play a role in treatment decisions. UROLOGY 86: 901–905, 2015.

Active Surveillance (AS) is an expectant management strategy that intends to delay and possibly avoid curative therapy and its potential adverse effects on urinary and sexual function. AS is based on the use of regular monitoring via clinical exams, prostate-specific antigen (PSA) testing, biopsies, and possibly imaging to determine when change in the risk from the disease becomes evident. This process is continued until a patient decides to proceed with definitive treatment, progresses to a less-intensive "watchful waiting" approach, or dies from another cause. However, there are risks and benefits that make the decision to pursue AS challenging. Moreover, there is substantial variability in proposed criteria for patient selection, monitoring strategies, and thresholds for recommending intervention with curative-intent therapy.[1-3] In addition, data on selection criteria and outcomes have emerged mainly from single-site, tertiary care institutions.[4-6]

Understanding the factors that influence selection of AS at a population level can help quantify the causes of observed variation. Evidence suggests that the selection of AS as a treatment modality may be influenced by a variety of clinical factors, including patient age, comorbidity, race, PSA, Gleason score (GS), clinical T-stage, PSA density, burden of tumor within the prostate, and potentially imaging and biomarkers.[7,8] Beyond such clinical variables, individual patient preferences and recommendations from the treatment provider are also likely to influence the use of AS.[9]

In this context, we sought to identify which factors predict the decision to choose AS or curative therapy. To examine the association between clinical variables and the receipt of initial AS, we used data for men newly

Financial Disclosure: *David C. Miller had contract support from Blue Cross Blue Shield of Michigan for serving as Director of the Michigan Urological Surgery Improvement Collaborative and grant support from the National Cancer Institute. James E. Montie receives salary support from Blue Cross Blue Shield of Michigan (through the University of Michigan) for serving as Co-director of the Michigan Urological Surgery Improvement Collaborative. Brian T. Denton had grant support from the National Science Foundation. The remaining authors declare that they have no relevant financial interests.*

From the Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI; and the Department of Urology, University of Michigan, Ann Arbor, MI

Address correspondence to: Brian T. Denton, Ph.D., Department of Industrial and Operations Engineering, University of Michigan, 2893 IOE Building, 1205 Beal Ave, Ann Arbor, MI 48109-2117. E-mail: btdenton@umich.edu

Submitted: May 17, 2015, accepted (with revisions): August 28, 2015

http://dx.doi.org/10.1016/j.urology.2015.08.024
0090-4295

diagnosed with prostate cancer from the Michigan Urological Surgery Improvement Collaborative (MUSIC). We fit multivariable logistic regression (LR) models to identify the factors associated with the choice of AS vs curative therapy. We then used bootstrapping to assess the calibration and discrimination ability of these models to predict the initiation of AS for patients in different prostate cancer risk strata.

METHODS

Study Population

With support from Blue Cross Blue Shield of Michigan, MUSIC was established in 2011 as a statewide physician-led collaborative aiming to improve the quality and cost-effectiveness of prostate cancer care in Michigan. The collaborative now comprises a diverse group of 42 academic and community practices, including more than 80% of urologists in the state. Each practice involved in MUSIC obtained an exemption or approval for participation from a local institutional review board. The participating practices have trained data abstractors who review medical records and enter standardized data elements into a web-based clinical registry for men undergoing prostate biopsy or diagnosed with prostate cancer. The data include patient age, PSA, GS, proportion of positive over total number of biopsy cores, clinical T-stage, treatment decision, comorbidity, race, and PSA density. The analysis included 2977 patients with newly diagnosed, localized prostate cancer from 21 practices that entered at least 30 patients with newly diagnosed prostate cancer in the registry from January 1, 2012 through December 31, 2013.

Primary Outcome

The outcome of interest was initial treatment choice as documented in the MUSIC registry. These data are entered into the registry by the data abstractor only when it is explicitly written in the patient's chart. To ensure quality and accuracy of data collected, MUSIC employs standard operating procedures with specific variable definitions, ongoing abstractor education, and annual data audits performed by the coordinating center.[10,11] Additionally, with claims-based treatment as the reference, the Cohen kappa statistic has been used to assess the accuracy of the treatment assignment in the MUSIC registry.[9] With a random 21% sample of patients, excellent concordance ($\kappa = 0.93$) was observed between claims-based treatment and the MUSIC registry, providing external validation of the MUSIC registry.

Statistical Analyses

We compared clinical and pathological characteristics of patients who chose AS with those who received curative therapy with surgery or radiation therapy (defined by a binary indicator) for each of the D'Amico risk groups individually and for the complete cohort. The D'Amico stratification classifies patients into either low-risk (stage T1c, T2a, and PSA level ≤10 ng/mL and GS ≤6), intermediate-risk (stage T2b or GS of 7 or PSA level >10 and ≤20 ng/mL), or high-risk (stage T2c or PSA level >20 ng/mL or GS ≤8) groups.[12] Differences between these 3 groups of patients in medians for quantitative variables, and differences in distributions for categorical variables were compared using Mann-Whitney *U* test (also known as Wilcoxon test) and chi-square test, respectively.

We fit multivariate LR models to examine the independent association between the AS indicator and clinical factors, including practice group, patient age, PSA (PSA was transformed to ln[PSA + 1] to scale), GS, clinical T-stage, highest percent of a core involved with cancer (GPC), and the percentage of positive biopsy cores (number of cores containing cancer divided by total number of cores sampled). The practice group is a unique identifier that defines each urology practice in MUSIC as a group of urologists practicing in the same offices represented by a single clinical champion. Life expectancy was calculated using the Roswell park calculator, based on patient age, comorbidity index, and GS. Factors selected for multivariate LR models were based on univariate analysis. Stepwise LR was then used to examine factors not otherwise excluded, to finalize each multivariate LR model.

The finalized multivariate LR models were validated using bootstrapping to evaluate measures of model discrimination and calibration. For bootstrapping, random samples were drawn with replacement from the target cohort to create 200 replicate (validation) cohorts. Each cohort was used to fit an LR model with the selected factors. Performance measures included area under the curve (AUC), Brier score, and calibration slope. Brier score varies from 0 (perfect prediction) to 0.25 (no predictive value), whereas perfect calibration is indicated by a value of 1, indicating concordance between observed and model estimated probabilities. All statistical testing was two sided with a significance level of 0.05 and was performed using SAS 9.3.

RESULTS

Study Population

Among the 2977 men, 609 men with newly diagnosed prostate cancer initiated AS. Among patients who initiated AS, more than two-thirds had D'Amico low-risk cancers. On average, patients who underwent AS were older, had a lower tumor grade (GS and clinical T-stage), and had a smaller tumor volume (defined by GPC and portion of positive cores) than those who did not choose AS. Table 1 presents median and mean values of these characteristics for the complete cohort and stratified risk cohorts by D'Amico risk category.[12] All factors presented were significant for the complete cohort and the intermediate-risk cohort. However, only patient age, number of cores taken, tumor volume, and urology practice group were significant in the low-risk cohort.

Statistical Analyses

Univariate analyses using the chi-square test were used to assess the association between clinical factors and AS indicator value for the D'Amico low-risk cohort and the

Table 1. Patient characteristics

Variables	Complete Cohort Without AS	Complete Cohort With AS	*P* Value	Low-risk Without AS	Low-risk With AS	*P* Value
Number of patients	2368	609		428	411	
Age at diagnosis (years)			.0039			<.0001
Mean (median)	64.5 (64)	65.4 (66)		61.7 (62.5)	64.2 (65)	
Range	39-95	39-87		41-82	39-83	
Clinical T-stage, No. (%)			<.0001			.1910
T1	1599 (67.5%)	538 (88.3%)		373 (87.2%)	370 (90.0%)	
T2	701 (29.6%)	71 (11.7%)		55 (12.9%)	41 (10.0%)	
T3-T4	68 (2.9%)	0 (0%)		0 (0%)	0 (0%)	
PSA, ng/mL			<.0001			.0573
Mean (median)	22.0 (5.7)	6.6 (5.4)		5.1 (4.8)	5.3 (5.2)	
Range	0.1-6873.4	0.2-170.1		0.4-10.0	0.3-9.9	
PSA, ng/mL, No.			<.0001			.2703
<4 (%)	437 (18.5%)	132 (21.7%)		113 (26.4%)	95 (23.1%)	
4-10 (%)	1444 (61.0%)	401 (65.9%)		315 (73.6%)	316 (76.9%)	
>10 (%)	487 (20.6%)	76 (12.5%)		0 (0%)	0 (0%)	
Biopsy Gleason score, No.			<.0001			All ≤6
≤3 + 3 (%)	551 (23.3%)	487 (80.0%)		428 (100%)	411 (100%)	
3 + 4 (%)	904 (38.2%)	95 (15.6%)		0 (0%)	0 (0%)	
≥4 + 3 (%)	913 (38.6%)	27 (4.4%)		0 (0%)	0 (0%)	
Biopsy cores taken, No.			.0030			.0341
Mean (median)	12.7 (12)	12.2 (12)		12.1 (12)	12.5 (12)	
Range	1-77	1-78		1-26	1-70	
Positive cores, No.			<.0001			<.0001
Mean (median)	4.7 (4)	1.9 (1)		2.7 (2)	1.7 (1)	
Range	1-39	1-20		1-14	1-10	
Positive cores, %			<.0001			<.0001
Mean (median)	39.5 (33.3)	16.1 (10.0)		23.3 (16.7)	14.3 (8.3)	
Range	3.3-100	2.8-100		4.5-100	3.8-100	
GPC, %			<.0001			<.0001
Mean (median)	48.9 (50.0)	18.0 (10.0)		24.9 (17.0)	14.9 (9.0)	
Range	0-100	0-100		0-100	0-95	

AS, Active Surveillance; GPC, greatest positive core percent; PSA, prostate-specific antigen.

Table 2. Multivariable LR models for AS initiation for the complete cohort

Complete Cohort Without Practice Group			Complete Cohort With Practice Group		
Factors	OR (95% CI)	*P* Value	Factors	OR (95% CI)	*P* Value
Age at diagnosis	1.056 (1.041-1.071)	<.0001	Age at diagnosis	1.056 (1.041-1.072)	<.0001
Clinical T-stage		(.0018)	Clinical T-stage		(.0131)
T2-T4	Reference		T2-T4	Reference	
T1	1.656 (1.206-2.274)	.0018	T1	1.516 (1.092-2.107)	.0131
Biopsy Gleason sum		(<.0001)	Biopsy Gleason sum		(<.0001)
≥4 + 3	Reference		≥4 + 3	Reference	
3 + 4	2.276 (1.439-3.598)	.0004	3 + 4	2.441 (1.532-3.890)	.0002
≤3 + 3	11.931 (7.689-18.514)	<.0001	≤3 + 3	14.597 (9.258-23.013)	<.0001
Positive cores, %	0.023 (0.009-0.058)	<.0001	Positive cores, %	0.026 (0.010-0.066)	<.0001
GPC, %	0.986 (0.981-0.992)	<.0001	GPC, %	0.987 (0.981-0.992)	<.0001
			Practice group	–	(<.0001)

CI, confidence interval; Cohort, complete cohort; OR, odds ratio.

complete cohort. Some factors (including PSA density and race) were not significant in univariate analysis. There were also factors (including comorbidity) that were significant in univariate models but were not significant in multivariate models because of the correlation with other factors. Multivariate analyses for the complete cohort and the low-risk cohort are presented in Tables 2 and 3. Patient age, GS, clinical T-stage, urology practice, GPC, and proportion of positive cores (P <.05) were associated with the decision to initiate AS in the complete cohort. In the low-risk cohort, only patient age, GPC, portion of positive cores, and urology practice group were significant (P <.05). Life expectancy and comorbidity did not improve predictive performance; thus, we did not incorporate life expectancy or comorbidity in our final models. The Pearson correlation coefficient between the proportion of patients initiating AS

Table 3. Multivariable LR models for the initiation of AS for the low-risk cohort

Low-risk Cohort Without Practice Group			Low-risk Cohort With Practice Group		
Factors	OR (95% CI)	*P* Value	Factors	OR (95% CI)	*P* Value
Age at diagnosis	1.056 (1.041-1.071)	<.0001	Age at diagnosis	1.054 (1.032-1.076)	<.0001
Positive cores, %	0.023 (0.009-0.058)	<.0001	Positive cores, %	0.019 (0.005-0.076)	<.0001
GPC, %	0.986 (0.981-0.992)	<.0001	GPC, %	0.985 (0.977-0.994)	.0008
			Practice group	–	(.0026)

Table 4. Predictive performance metrics of multivariable LR models

		Base Model			Bootstrapping*		
Model Specification	N	AUC	Brier Score	Calibration Slope	AUC	Brier Score	Calibration Slope
Complete cohort with practice group	2958	0.885	0.104	1	0.889	0.101	0.944
Complete cohort without practice group	2958	0.875	0.108	1	0.876	0.105	0.990
Low-risk cohort with practice group	831	0.751	0.203	1	0.764	0.207	0.746
Low-risk cohort without practice group	831	0.706	0.217	1	0.708	0.219	0.935
Intermediate-risk cohort	1315	0.864	0.076	1	0.875	0.070	0.773
High-risk cohort	694	0.881	0.024	1	0.886	0.022	0.864

* The mean value of 200 replicates for each performance measure is presented for results of bootstrapping.

at each urology practice group and the number of patients seen at each practice group was 0.35 ($P = .11$).

The results for AUC, Brier score, and calibration are presented in Table 4. The predictive performance was excellent for intermediate- and high-risk cohorts. In the low-risk cohort, only patient age, GPC, portion of positive cores, and urology practice group were significant (AUC = 0.751, Brier score = 0.22). The addition of urology practice in particular significantly increased AUC in the low-risk cohort from 0.71 to 0.76 and reduced the Brier score slightly from 0.22 to 0.21.

DISCUSSION

For low-risk patients, patient age, GPC, portion of positive cores, and practice group were associated with initiation of AS. These results are consistent with heterogeneity in intermediate-risk patients such that GS and clinical T-stage become important. In the low-risk cohort where GS and clinical T-stage are favorable, more subtle differences between the burden of disease in the prostate (GPC and portion of positive cores) are significant factors. The best LR model including life expectancy and comorbidity did not demonstrate better predictive performance than our proposed models.

It has been shown that urology practice is associated with initiation of AS.[9] Our study expands on this by examining measures of predictive performance for model discrimination and calibration. Our study also includes a larger sample of patients of all risk categories compared with the sample in Womble et al's study. We found that the predictive ability of the models is better for the intermediate- and high-risk cohorts compared with the low-risk cohort. However, the predictive ability for the low-risk cohort is still very good. Moreover, the addition of urology practice improved the predictive performance for the low-risk cohort significantly. There was no statistically significant correlation between the proportion of patients initiating AS and the number of patients seen at each practice group. These observations suggest that clinical factors and urologists' preferences are important drivers of patients' decisions to initiate AS.

The results from this population-based sample, which has a diverse representation from academic, community, solo practices, and both large and small groups, are reassuringly consistent with selection criteria recommended by single-site, tertiary care centers that have been at the vanguard of AS use. The finding that the prediction was less accurate in the low-risk strata implies that other unmeasured factors, such as patient preference or provider concern regarding inadequate assessment of the true nature of the cancer in the prostate, are commonly a driving force in the selection process. There is enthusiasm that the opportunity for shared decision making between the patient and provider, ideally supplemented with prostate cancer-specific decision aids, will help patients understand better the benefits and risks of AS by more precisely presenting a realistic picture of the risk from the untreated disease and lead to less anxiety in the choice of AS. There is also optimism that emerging technologies, such as improved magnetic resonance imaging or gene expression biomarkers, will better identify patients who are less well suited for initial AS and thus provide reassurance for the patient and the provider who do choose AS.

Our study has several limitations. First, this study focuses only on patients from the 21 practices in Michigan providing data during the time period of the study; practices joined MUSIC over this time period, but there was no systematic recruiting practice. Second, there is the possibility for geographic variability in the factors impacting AS

decisions. Third, there are other practice patterns, such as the use of magnetic resonance imaging and prognostic biomarkers, which were not captured in this study, and which could potentially explain some of the influence of practice group on decisions to initiate AS. Finally, because of the small number of patients, the analysis was not feasible at the individual provider level and thus there could be variability of use of AS within a practice group. Factors such as practice group impact the decision on whether to use AS, but a more thorough analysis of the patient decision-making process, including individual patient-physician interaction could further elucidate the root causes of variation in patient decisions to initiate AS.

These limitations notwithstanding, our findings show that clinical factors and practice group are good predictors of whether patients will initiate AS in a real-world setting, and confirm the acceptance of AS as a viable strategy for patients with localized prostate cancer by the urology community in Michigan using selection factors consistent with published data and guidelines. Additional research is needed to address qualitative factors influencing decision making and to introduce across the population improved educational tools for patients and providers alike.

CONCLUSION

Patient age, intraprostatic tumor volume, and urology practice group are the most significant factors impacting the choice of treatment, independent of risk category. Predictions based on these factors alone are less accurate for low-risk patients than for the rest of the cohort, suggesting the importance of identifying other, perhaps behavioral, factors associated with patient preferences that influence the decisions to initiate AS.

Acknowledgment. The authors acknowledge the significant contribution of the clinical champions, urologists, administrators, and data abstractors in each participating MUSIC practice, as well as members of the MUSIC Coordinating Center at the University of Michigan (http://www.musicurology.com). In addition, the authors acknowledge the support provided by David Share, M.D., Tom Leyden, M.B.A., Rozanne Darland, B.S.B.A., Sarah Lanivich, M.B.A., and other collaborators in the Value Partnerships program at Blue Cross Blue Shield of Michigan.

This material is also based in part upon work supported by the National Science Foundation (NSF) under Grant Number CMMI 1536444. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

1. Loblaw A, Zhang L, Lam A, et al. Comparing prostate specific antigen triggers for intervention in men with stable prostate cancer on active surveillance. *J Urol*. 2010;184:1942-1946. doi:10.1016/j.juro.2010.06.101.
2. Bul M, Zhu X, Valdagni R, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. *Eur Urol*. 2013;63:597-603. doi:10.1016/j.eururo.2012.11.005.
3. Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol*. 2010;28:126-131. doi:10.1200/JCO.2009.24.2180.
4. Suardi N, Capitanio U, Chun FK, et al. Currently used criteria for active surveillance in men with low-risk prostate cancer: an analysis of pathologic features. *Cancer*. 2008;113:2068-2072. doi:10.1002/cncr.23827.
5. Soloway MS, Soloway CT, Williams S, Ayyathurai R, Kava B, Manoharan M. Active surveillance; a reasonable management alternative for patients with prostate cancer: the Miami experience. *BJU Int*. 2008;101:165-169. doi:10.1111/j.1464-410X.2007.07190.x.
6. Dall'Era MA, Konety BR, Cowan JE, et al. Active surveillance for the management of prostate cancer in a contemporary cohort. *Cancer*. 2008;112:2664-2670. doi:10.1002/cncr.23502.
7. Soloway MS, Soloway CT, Eldefrawy A, Acosta K, Kava B, Manoharan M. Careful selection and close monitoring of low-risk prostate cancer patients on active surveillance minimizes the need for treatment. *Eur Urol*. 2010;58:831-835. doi:10.1016/j.eururo.2010.08.027.
8. Dall'Era MA, Albertsen PC, Bangma C, et al. Active surveillance for prostate cancer: a systematic review of the literature. *Eur Urol*. 2012;62:976-983. doi:10.1016/j.eururo.2012.05.072.
9. Womble PR, Montie JE, Ye Z, et al. Contemporary use of initial active surveillance among men in Michigan with low-risk prostate cancer. *Eur Urol*. 2015;67:44-50. doi:10.1016/j.eururo.2014.08.024.
10. Riedinger CB, Womble PR, Linsell SM, et al. Variation in prostate cancer detection rates in a statewide quality improvement collaborative. *J Urol*. 2014;192:373-378. doi:10.1016/j.juro.2014.02.088.
11. Womble PR, Dixon MW, Linsell SM, et al. Infection related hospitalizations after prostate biopsy in a statewide quality improvement collaborative. *J Urol*. 2014;191:1787-1792. doi:10.1016/j.juro.2013.12.026.
12. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA*. 1998;280:969-974. doi:10.1001/jama.280.11.969.