Understanding the performance of active surveillance selection criteria in diverse urology practices

Scott R. Hawken BS*[,1], Paul R. Womble MD*[,1], Lindsey A. Herrel MD[1], Zaojun Ye MS[1], Susan M. Linsell MHSA[1], Patrick M. Hurley MD[2], James E. Montie MD[1], David C. Miller MD, MPH[1]
for the Michigan Urological Surgery Improvement Collaborative

Department of Urology, University of Michigan, Ann Arbor, MI[1]
Providence Hospital, Southfield, MI[2]
*Equal contribution from authors

Abstract word count: 228
Word count: 2,495
Tables: 3; **Figures:** 1

Key words: prostate cancer, active surveillance, guidelines, quality improvement

Corresponding Author:
David C. Miller, MD, MPH
Associate Professor
Department of Urology
University of Michigan
NCRC Building 16, Room 108E
2800 Plymouth Road
Ann Arbor, MI 48109-2800

Phone: (734) 936-0054
Fax: (734) 232-2400
Email: dcmiller@med.umich.edu

Abstract

Purpose: We used data from the Michigan Urological Surgery Improvement Collaborative (MUSIC) to evaluate the performance of published selection criteria for active surveillance (AS) in diverse urology practice settings.

Materials and Methods: For several AS guidelines, we calculated the proportion of men meeting each set of selection criteria that actually entered AS (defined as the sensitivity of the guideline). After identifying the most sensitive guideline for the entire cohort, we compared demographic and tumor characteristics between patients meeting this guideline who entered AS and those who received initial definitive therapy.

Results: Among 4,882 men with newly diagnosed prostate cancer, 18% underwent AS. When applied to the entire cohort, the sensitivity of published guidelines ranged from 49% (Toronto) to 62% (Johns Hopkins, JH). At a practice-level, the sensitivity of the JH criteria varied widely from 27% to 84% ($p<0.001$). Compared with men undergoing AS, patients meeting JH criteria that received definitive therapy were younger ($p<0.001$), more likely to have a positive family history ($p=0.003$), a lower PSA ($p<0.001$), greater number of positive cores (2 vs 1) on biopsy ($p<0.001$), and higher cancer volume in the positive core(s)($p=0.002$).

Conclusions: The sensitivity of published AS selection criteria varies widely across diverse urology practices. Among patients meeting the most stringent criteria, those who received initial definitive therapy had characteristics suggesting greater cancer risk, underscoring the nuanced clinical factors that influence treatment decisions.

Introduction

Owing to concerns about overtreatment of men with lower-risk prostate cancer,[1–4] initial active surveillance (AS) is being used more frequently for patients with early-stage disease.[5] The potential benefits of AS include avoidance of treatment-related side effects (e.g., urinary incontinence, erectile dysfunction) by delaying or not pursuing definitive therapy; however, these benefits must be weighed against the potential risk for cancer progression. Given the growing acceptance of AS among both patients and physicians, there are now many published guidelines describing optimal selection criteria for patients to enter initial AS.[6–13]

Although most guidelines for entry into initial AS are widely recognized,[6–13] their application and performance in clinical practice is not well characterized. One measure of interest is the sensitivity of such guidelines in the "real-world" setting—that is, how many men that meet these selection criteria actually proceed with initial AS. While it makes sense that the most restrictive criteria would also be the most sensitive, the actual proportion of patients undergoing surveillance according to these different guidelines is not well defined for diverse urology practice settings. In addition, little is known about the factors that lead men who meet established criteria for entering surveillance to nonetheless proceed with initial local therapy.

In this context, we examined the sensitivity of several published guidelines for identifying men that selected initial AS in the community and academic practices comprising the Michigan Urological Surgery Improvement Collaborative (MUSIC). In addition to comparing the sensitivity of these guidelines across MUSIC practices, we also evaluated differences in demographic characteristics, comorbidity, and cancer severity among patients meeting the most sensitive selection criteria who entered AS and those who underwent initial definitive therapy.

Materials and Methods

Michigan Urologic Surgery Improvement Collaborative

MUSIC was established in 2011 with the aim of improving the quality and cost-efficiency of prostate cancer care in the state of Michigan.[5,14–16] The collaborative now includes 42 urology practices comprising more than 90% of urologists in the state. MUSIC receives financial support from Blue Cross Blue Shield of Michigan (BCBSM), and each participating practice obtained an exemption or approval for participation from a local institutional review board.

Study population and data elements

For all men undergoing prostate biopsy and/or with a new prostate cancer diagnosis seen in participating practices, trained abstractors enter a standardized set of data elements into a web-based registry including, patient age, race/ethnicity, Charlson comorbidity index score (CCI), serial PSA results, clinical stage, biopsy Gleason score, number of positive cores, cancer-directed treatments, and follow-up laboratory and pathology results. Quality assurance steps for MUSIC data have been described previously.[5,14–16] The population for this analysis includes 4,883 men with newly-diagnosed prostate cancer entered into the MUSIC registry from March 2012 through June 2014.

Identification of primary treatment

To ensure complete and accurate data, abstractors wait 3 months from the date of prostate cancer diagnosis before entering information about cancer treatment into the MUSIC registry. In addition, explicit documentation in the medical record is required to assign a specific treatment (e.g., AS, radical prostatectomy, external beam radiation therapy, etc.). For instance, the definition of AS in MUSIC (as defined by participating urologists and provided to the data abstractors) is as follows: “Active Surveillance is a slightly more structured and aggressive form of watchful waiting. It is recommended that a patient undergo a DRE and a PSA test every 3 or 6 months, depending on the patient’s precise history and clinical condition, and to re-biopsy the patient yearly or every two years. The physician monitors the patient aggressively and will regularly discuss disease status with the patient so that joint decisions are

made about the need for actual treatment." As described by Womble et al, treatment assignment in MUSIC has been externally validated with claims data with excellent concordance.[5]

Identification of AS selection criteria

For this analysis, we identified several of the most prominent guidelines for selecting patients for initial AS, including those from the following institutions and organizations: Johns Hopkins (JH),[6] National Comprehensive Cancer Network (NCCN): Low Risk and Very Low Risk,[7] Memorial Sloan Kettering Cancer Center (MSKCC),[8] University of California San Francisco (UCSF),[9,10] and University of Toronto.[11,12] The selection criteria for each of these guidelines are based on routinely available clinical data, but vary according to their specific elements and definitions.

Statistical Analyses

We first compared clinical and pathologic characteristics of patients that entered AS and those that received other initial management. Next, we calculated the proportion of all men meeting each set of selection criteria that entered AS, as well as the associated standard Wald asymptotic 95% confidence intervals. We refer to this proportion throughout the manuscript as the sensitivity of real-world practice patterns for each guideline. The numerator is comprised of men that met each selection criteria and chose AS, and the denominator is all men that met the selection criteria.

After identifying the guideline with the most sensitive selection criteria for the entire study population (i.e., for the collaborative as a whole), we then examined the variation in the sensitivity of this guideline at a practice-level for sites with greater than 10 patients meeting these criteria. We chose to examine the most sensitive guideline in order to identify the cohort of patients that the greatest number of urologists would likely agree are candidates for surveillance. Next, we used chi-square testing to examine differences in the sensitivity of this guideline across MUSIC practices.

Finally, for men meeting selection criteria for the most sensitive guideline, we examined differences between patients who entered AS and those who received definitive therapy. Because we were only interested in men considered eligible to receive local therapy, we excluded from this analysis 47 patients managed with watchful waiting and 4 patients managed with androgen deprivation therapy. We analyzed differences in variables that are not included explicitly in the guideline selection criteria (e.g., age, race, comorbidity, practice size), as well as those that are included in the criteria but still maintain a range of clinically meaningful values after selection, such as PSA level, PSA density, number of positive biopsy cores, and greatest percentage of a core positive for cancer (GPC). Using the same variables, we also compared men meeting the most sensitive AS guidelines who received radiation versus those who underwent surgery. For each comparison, we used a two-sided t-test, Wilcoxon rank-sum test, or chi square test, where appropriate. All statistical testing was performed at the 5% significance level using SAS v9.3 (SAS Institute Inc., Cary, NC, USA).

Results

Table 1 presents clinical characteristics for the 4,883 men with newly diagnosed prostate cancer. Among this group, 18% (901 men) entered initial AS, 42% (2052 men) underwent surgery, 21% (1033 men) received radiation therapy, 5% (241 men) were treated with androgen deprivation therapy, 4% (188 men) selected watchful waiting, 1% (46 men) received other treatments (e.g., cryosurgical ablation), and the initial therapy was unknown for 9% (425 men) of these cases. The 9% of men with missing data for treatment were younger (mean age 63 vs 65 years, $p<0.001$), more likely to be African American (22% vs 15%, $p<0.001$), have lower clinical stage tumors (78% T1, 20% T2 vs 72% T1, 26% T2, $p=0.03$), and from a practice with >10 urologists (51% vs 39%, $p<0.001$), but were otherwise similar to those with treatment documented in the MUSIC registry.

Overall, men entering initial AS were older, have clinical stage T1 tumors (vs T2 or T3/4), a lower median PSA, a lower biopsy Gleason Score, fewer positive biopsy cores on biopsy, and a lower GPC (Table 1).

The AS selection criteria examined in this analysis are comprised of specific combinations of routinely available clinical data, including Gleason score, PSA, PSA density, clinical T-stage, number of positive biopsy cores, and GPC. Table 2 presents a summary of the selection criteria for several published guidelines, as well as the sensitivity of each set of criteria for identifying men entering AS when applied to the entire cohort. Across all patients, the sensitivity of these criteria for identifying patients that actually entered AS ranged from a low of 48.9% (95% CI: 46.2-51.6%) for the Toronto criteria to a high of 62.4% (95% CI: 58.0-66.7%) for the Johns Hopkins criteria. The Toronto criteria identified the greatest absolute number of men entering initial AS (647 men); comparatively the NCCN: Very Low Risk criteria identified the fewest (290 men). Fourteen MUSIC practices had at least 10 patients meeting the JH selection criteria; across these sites, the sensitivity of the JH guideline for identification of patients actually receiving surveillance ranged from 27% to 84% ($p<0.001$, Figure 1). Among the men missing treatment data, only 10% (44 men) met the JH selection criteria.

For the cohort of men meeting the JH selection criteria for initial AS and eligible for treatment (i.e., those not undergoing watchful waiting or receiving primary ADT), we compared demographics and cancer characteristics between patients that actually entered AS and those that received definitive therapy (Table 3). Compared with men undergoing initial AS, those receiving initial local therapy were younger ($p<0.001$), had lower PSA values ($p<0.001$), and were more likely to have a positive family history ($p=0.003$). Treated men also had evidence of greater tumor volume, as indicated by 2 (vs 1) positive cores on biopsy ($p<0.001$), and higher percentage of a core involved with cancer (GPC) (median 10% vs 7%, $p=0.003$). There were no statistically significant differences between these two groups for any of the other variables analyzed, including race and comorbidity.

Among the patients undergoing definitive therapy despite meeting the JH guideline, 68% and 32% underwent surgery and radiation respectively. Those that received radiation rather than surgery were older (mean age 66 vs 59 years, p<0.001), had more comorbid conditions (33% with CCI of 1, 17% with CCI ≥2 vs 11% with CCI of 1, 12% with CCI ≥2, p=0.004), and were less likely to have a family history of prostate cancer (26% vs 52%, p=0.006).

Discussion:

We examined data from a large number of community and academic urology practices to understand the degree to which various published AS selection criteria actually identify men who enter surveillance in real-world practice. Not unexpectedly, the sensitivity of these guidelines is variable, with those containing more stringent criteria generally capturing a greater proportion of all men meeting the criteria that actually entered AS. Conversely, more liberal criteria identified a greater absolute number of men entering AS. The most sensitive guideline overall (from Johns Hopkins) still showed significant variation in its performance at a practice-level. Among the cohort of patients meeting the JH criteria and considered eligible for local therapy, those who received definitive treatment had higher tumor volume on biopsy, as well as other characteristics favoring treatment (e.g., younger age, positive family history), that are not explicit components of current AS selection criteria.

In addition, our findings suggest that when urologists counsel men about AS versus definitive treatment, they consider age and family history—two relevant factors not routinely captured by AS guidelines—as well small differences in the volume of cancer. Previous work exploring the implications of such differences in tumor volume has been mixed, with others demonstrating that a GPC>10%, but not the presence of one vs two positive cores at biopsy, increases the risk of adverse surgical pathology for patients with Gleason 6 prostate cancer.[17] While our findings certainly seem reassuring with respect to current practice patterns, the clinical significance of these differences in tumor volume is admittedly debatable.

Our overall findings are consistent with previous investigations showing that AS guidelines vary in the number of men with newly diagnosed prostate cancer that meet such selection criteria.[13] Moreover, the observed practice-level variation in sensitivity of AS guidelines is consistent with our own prior reports on AS utilization among patients with low-risk cancers[5], as well as prior studies demonstrating variation in primary treatment of localized prostate cancer by practice site and provider.[3,18] While our study identified statistically significant differences in measures of cancer risk between patients meeting JH criteria and entering surveillance versus those receiving initial local therapy, the clinical significance of these differences is modest at best. As such, it seems likely that variation in the use of AS among men meeting the stringent JH selection criteria may reflect primarily differences in patient preferences or provider perceptions and beliefs about AS that were not measured in this analysis.

The optimal selection criteria for AS remain uncertain. Whereas utilizing more stringent criteria would likely identify patients with a lower risk of disease progression, it may also exclude some men that are good candidates for surveillance. Our findings also indicate that urologists in Michigan have not coalesced around a single set of selection criteria for this important initial treatment decision.

Our analysis has several limitations. First, our cohort included only patients and practices in Michigan, and our findings may therefore not be generalizable to a broader population. Second, treatment data is missing for a small number of men in the cohort. Although such missing data raises concerns about selection bias, even if all 44 men missing treatment data did not receive AS, this would only decrease the sensitivity of the JH criteria to 57%; conversely, if all of these men entered initial AS, the sensitivity of the JH criteria would increase to 65%. Third, we measured neither patient preferences nor physician beliefs and perceptions about the risks of prostate cancer, treatment side effects, and comorbidities. These unmeasured factors undoubtedly play a role treatment decisions, including the selection of patients for AS. Finally, although our study examines the initial entry into AS, the effectiveness of active surveillance ultimately depends on long-term utilization with selected intervention for men with disease progression. Therefore, further study is needed to better define the outcomes of AS in this cohort.

Despite these limitations, our findings have important implications for both patients and providers. For patients, our findings suggest opportunities to expand the use of AS, particularly among men meeting selection criteria who prioritize preservation of urinary and/or sexual function. For providers, these data suggest that, even within a single state, there appear to be substantial differences in beliefs and perceptions around prostate cancer risk (among other factors) that affect treatment recommendations for patients with low-risk tumors. As suggested by others, differing interpretations of the evidence-base supporting the safety of AS may explain the disconnect between providers that routinely put patients on AS and those that do not.[19] It is also possible that differences in other unmeasured factors, such as financial considerations, practice setting (e.g., urban vs rural), and/or the presence and strength of any academic affiliation, are impacting treatment decisions.

Moving forward, a better understanding of the entire decision making process is needed. This includes characterization of provider perceptions of risk and thresholds for treatment, how cancer risk is communicated with patients, shared decision making between patients and providers, and the degree to which differences in patient preferences drive the variation observed in this analysis. Once available, such data will provide essential context for understanding the implications of existing variation in the use of active surveillance for men with low-risk prostate cancer.

Acknowledgment Statement: The authors acknowledge the significant contribution of the clinical champions, urologists, administrators and data abstractors in each participating Michigan Urological Surgery Improvement Collaborative (MUSIC) practice (details around specific participating urologists and practices can be found at www.musicurology.com), as well as members of the MUSIC Coordinating Center at the University of Michigan. In addition, we would like to acknowledge the support provided by David Share, MD, Tom Leyden, MBA, Rozanne Darland, BSBA, Sarah Lanivich, MBA, and the Value Partnerships program at Blue Cross Blue Shield of Michigan.

References:

1. Daskivich TJ, Chamie K, Kwan L, et al: Overtreatment of men with low-risk prostate cancer and significant comorbidity. Cancer 2011; **117**: 2058–2066.

2. Miller DC, Gruber SB, Hollenbeck BK, et al: Incidence of initial local therapy among men with lower-risk prostate cancer in the United States. J. Natl. Cancer Inst. 2006; **98**: 1134–1141.

3. Cooperberg MR, Broering JM and Carroll PR: Time trends and local variation in primary treatment of localized prostate cancer. J. Clin. Oncol. 2010; **28**: 1117–1123.

4. Jacobs BL, Zhang Y, Skolarus TA., et al: Growth of high-cost intensity-modulated radiotherapy for prostate cancer raises concerns about overuse. Health Aff. 2012; **31**: 750–759.

5. 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.

6. Tosoian JJ, Trock BJ, Landis P, et al: Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J. Clin. Oncol. 2011; **29**: 2185–90.

7. National Comprehensive Cancer Institute (NCCN): Prostate Cancer (1.2015). Available at http://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf. Accessed October 2014.

8. Berglund RK, Masterson TA, Vora KC, et al: Pathological upgrading and up staging with immediate repeat biopsy in patients eligible for active surveillance. J. Urol. 2008; **180**: 1964–7.

9. Welty CJ, Cowan JE, Nguyen H, et al: Extended Followup and Risk Factors for Disease Reclassification in a Large Active Surveillance Cohort for Localized Prostate Cancer. J. Urol. 2014: 1–5.

10. 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.

11. Klotz L, Zhang L, Lam A, et al: Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. J. Clin. Oncol. 2010; **28**: 126–31.

12. Klotz L: Active surveillance for prostate cancer: for whom? J. Clin. Oncol. 2005; **23**: 8165–9.

13. Dall'Era M a, Albertsen PC, Bangma C, et al: Active surveillance for prostate cancer: a systematic review of the literature. Eur. Urol. 2012; **62**: 976–83.

14. 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.

15. 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.

16. Merdan S, Womble PR, Miller DC, et al: Toward Better Use of Bone Scans Among Men With Early-stage Prostate Cancer. Urology 2014; **84**: 793–8.

17. Katz MH, Shikanov S, Sun M, et al: Gleason 6 prostate cancer in one or two biopsy cores can harbor more aggressive disease. J. Endourol. 2011; **25**: 699–703.

18. Hoffman KE, Niu J, Shen Y, et al: Physician Variation in Management of Low-Risk Prostate Cancer. JAMA Intern. Med. 2014; **174**: 1450–1459.

19. Kim SP, Gross CP, Nguyen PL, et al: Perceptions of Active Surveillance and Treatment Recommendations for Low-risk Prostate Cancer: Results from a National Survey of Radiation Oncologists and Urologists. Med. Care 2014; **52**: 579–85.

Table 1. Characteristics of patients undergoing initial AS versus other management strategies

Variable		Patients entering initial AS	Patients receiving other treatments	p-value
Age				
	Mean (median)	65 (66)	65 (65)	0.02
	Range	39-87	38-95	
Race, n (%)				
	Caucasian	670 (84)	2589 (82)	
	African American	98 (12)	460 (15)	0.1
	Other	29 (4)	88 (3)	
Charlson Comorbidity Index score, n (%)				
	0	595 (68)	2282 (65)	
	1	145 (16)	697 (20)	0.05
	≥2	145 (16)	526 (15)	
Clinical Stage, n (%)				
	T1	791 (88)	2404 (68)	
	T2	104 (12)	1050 (30)	<0.001
	T3/4	0 (0)	88 (2)	
PSA level, ng/mL				
	Median	5.3	5.9	<0.001
Biopsy Gleason Score sum, n (%)				
	≤6	724 (81)	796 (23)	
	3 + 4	135 (15)	1375 (39)	<0.001
	4 + 3	21 (2)	585 (17)	
	8 - 10	10 (1)	732 (21)	
Positive cores, n				
	Mean (median)	1.9 (1)	4.6 (4)	<0.001
	Range	0-11	1-24	
GPC, %				
	Mean (median)	17.3 (10)	48.1 (46)	<0.001
	Range	0-100	0-100	

GPC: Greatest Percentage Positive of a Biopsy Core.

Table 2. Sensitivity of published Active Surveillance selection criteria among men in Michigan with newly-diagnosed prostate cancer

Guideline	Selection Criteria						Patients Meeting Selection Criteria (n)	Sensitivity (95% CI)
	Gleason Score	PSA	PSA Density	T-Stage	Positive Cores	GPC (%)		
JH	≤ 6	--	< 0.15	cT1c	≤ 2	≤ 50	486	62.4 (58.0-66.7)
NCCN: Very Low Risk	≤ 6	< 10	< 0.15	cT1c	≤ 2	≤ 50	466	62.2 (57.8-66.6)
MSKCC	≤ 6	< 10	--	≤ cT2a	≤ 3	< 50	984	56.3 (53.2-59.4)
UCSF	≤ 6	≤ 10	--	≤ cT2	≤ 33%	≤ 50	1083	54.3 (51.3-57.3)
NCCN: Low Risk	≤ 6	< 10	--	≤ cT2a	--	--	1271	49.7 (46.9-52.4)
Toronto	≤ 6	< 10	--	--	--	--	1323	48.9 (46.2-51.6)

GPC: Greatest Percentage Positive of a Biopsy Core, JH: Johns Hopkins, NCCN: National Comprehensive Cancer Institute, MSKCC: Memorial Sloan Kettering Cancer Center, UCSF: University of California San Francisco.

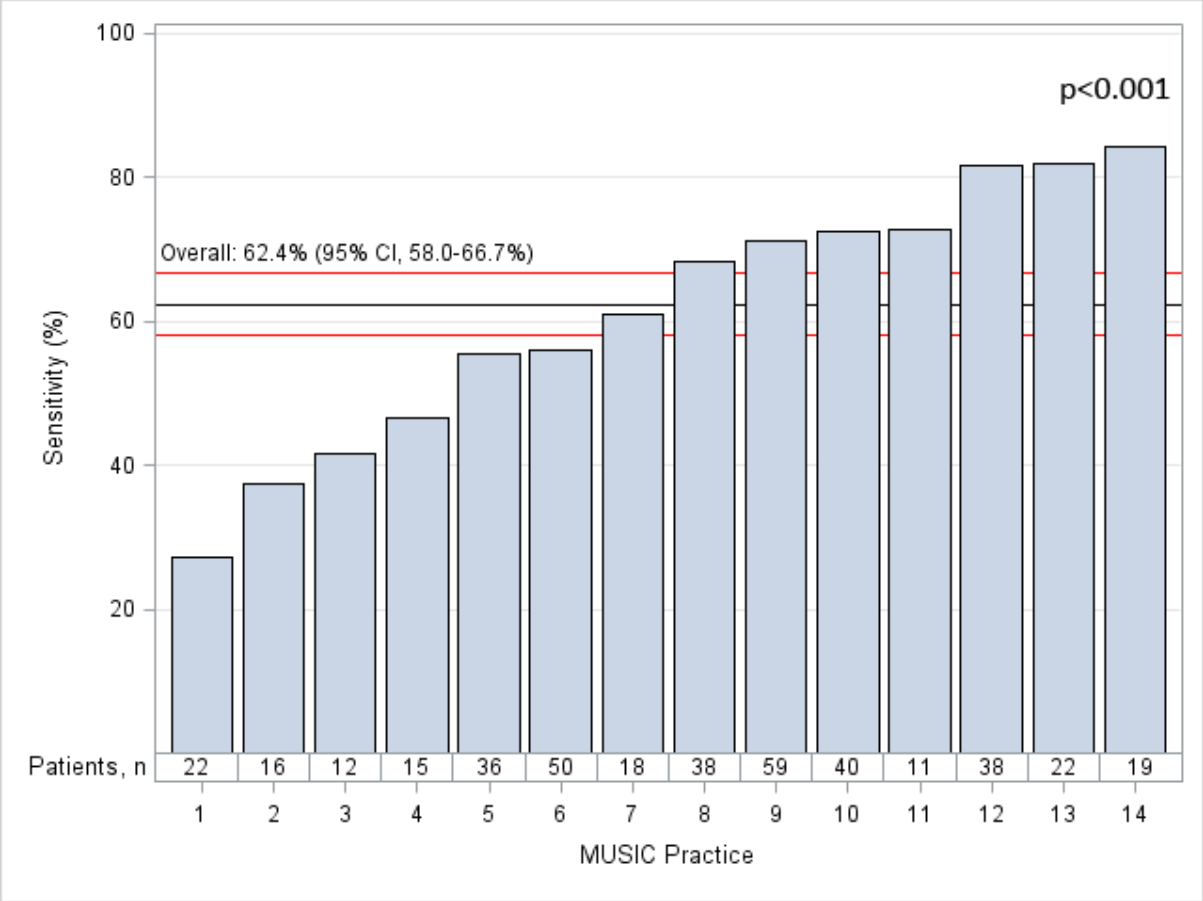
Table 3. Comparison of patients meeting JH Criteria that received initial AS versus definitive local therapy

Characteristic		Entry Criteria Met / Entered AS	Entry Criteria Met / Local therapy	p-Value
Total		303	132	
Age				
	Mean (median)	64 (65)	61 (62)	<0.001
	Range	41-83	41-77	
PSA level, ng/mL				
	Median	5	4.5	<0.001
# of Positive Cores				
	1	226 (75%)	76 (58%)	<0.001
	2	77 (25%)	56 (42%)	
GPC				
	Median (range)	7 (1-50)	10 (1-50)	0.002
Family History				
	Positive	82 (28%)	53 (44%)	0.003

Figure 1. Variation in sensitivity of the Johns Hopkins selection criteria for identifying men on active surveillance across MUSIC practices.

Figure Legend

Figure 1. Sensitivity of the Johns Hopkins (JH) guideline in practices with greater than 10 patients meeting the selection criteria. The overall sensitivity of the guideline was 62.4%. The variability among practices was statistically significant ($p < 0.001$).



Key of Definitions for Abbreviations:

AS: Active Surveillance

BCBSM: Blue Cross Blue Shield of Michigan

CCI: Charlson comorbidity index score

GPC: Greatest percentage of a biopsy core involved with cancer

JH: John Hopkins

MSKCC: Memorial Sloan Kettering Cancer Center

MUSIC: Michigan Urological Surgery Improvement Collaborative

NCCN: National Comprehensive Cancer Network

UCSF: University of California San Francisco