Original Article

Optimizing Active Surveillance Strategies to Balance the Competing Goals of Early Detection of Grade Progression and Minimizing Harm From Biopsies

Christine L. Barnett, PhD [1]; Gregory B. Auffenberg, MD, MS [1]; Zian Cheng[1]; Fan Yang[1]; Jiachen Wang[1]; John T. Wei, MD[1]; David C. Miller, MD, MPH[1]; James E. Montie, MD[1]; Mufaddal Mamawala, MBBS, MPH[2]; and Brian T. Denton, PhD [1]

BACKGROUND: Active surveillance (AS) for prostate cancer includes follow-up with serial prostate biopsies. The optimal biopsy frequency during follow-up has not been determined. The goal of this investigation was to use longitudinal AS biopsy data to assess whether the frequency of biopsy could be reduced without substantially prolonging the time to the detection of disease with a Gleason score $\geq$ 7. **METHODS:** With data from 1375 men with low-risk prostate cancer enrolled in AS at Johns Hopkins, a hidden Markov model was developed to estimate the probability of undersampling at diagnosis, the annual probability of grade progression, and the 10-year cumulative probability of reclassification or progression to Gleason score $\geq$ 7. It simulated 1024 potential AS biopsy strategies for the 10 years after diagnosis. For each of these strategies, the model predicted the mean delay in the detection of disease with a Gleason score $\geq$ 7. **RESULTS:** The model estimated the 10-year cumulative probability of reclassification from a Gleason score of 6 to a Gleason score $\geq$ 7 to be 40.0%. The probability of undersampling at diagnosis was 9.8%, and the annual progression probability for men with a Gleason score of 6 was 4.0%. On the basis of these estimates, a simulation of an annual biopsy strategy estimated the mean time to the detection of disease with a Gleason score $\geq$ 7 to be 14.1 months; however, several strategies eliminated biopsies with only small delays (<12 months) in detecting grade progression. **CONCLUSIONS:** Although annual biopsy for low-risk men on AS is associated with the shortest time to the detection of disease with a Gleason score $\geq$ 7, several alternative strategies may allow less frequent biopsying without sizable delays in detecting grade progression. ***Cancer* 2017;000:000–000.** © *2017 American Cancer Society.*

KEYWORDS: active surveillance, biopsy, Markov model, prostate cancer, reclassification.

INTRODUCTION

Although prostate cancers often demonstrate indolent clinical behavior,[1] many men with low-risk tumors still receive surgery or radiation therapy, both of which are associated with potentially serious complications, including incontinence, impotence, and other side effects.[2] These complications are particularly distressing because evidence shows that these men may not survive longer with surgery or radiation than they do with expectant management approaches. Active surveillance (AS) is a form of expectant management that involves monitoring patients through regular clinical examinations, biomarker tests, radiologic imaging, and biopsies. Because of the concern that many men who are diagnosed with prostate cancer are overtreated, AS has been promoted as a way for low-risk men to delay and possibly avoid surgery or radiation treatment. However, many approaches to implementing AS have been recommended, and the best approach is unclear.[3]

Because of a lack of evidence in support of a single optimal AS strategy, it is left to individual urologists and patients to decide how frequently to conduct follow-up biopsies. No previous study has made a link between different AS follow-up strategies and the delay in the detection of progression to high-grade cancer. The risk of progression is one of the most important considerations when one is weighing long-term risk for patients on AS. The ideal strategy to minimize the risk of delaying the detection of high-grade cancer progression is to biopsy patients frequently (eg, annually, as suggested by Tosoian et al[4]). However, this risk competes with the harm of frequent biopsies, which result in pain and anxiety for patients, and the potential for complications such as infections. The severe infection rate for biopsy is approximately 1% to 2%[5]; however, recent studies suggest that infection rates for patients undergoing AS increase as a function of the number of biopsies that they have undergone.[6] Studies have observed discontinuation of AS by patients without signs of progression,[7] and some have suggested that reducing surveillance biopsies may encourage compliance with AS.[8]

Corresponding author: Brian T. Denton, PhD, University of Michigan, 1205 Beal Avenue, 2893 IOE Building, Ann Arbor, MI 48109-2117; btdenton@umich.edu

[1]University of Michigan, Ann Arbor, Michigan; [2]Johns Hopkins University, Baltimore, Maryland

We thank H. Ballentine Carter for providing the data from the Johns Hopkins active surveillance study.

DOI: 10.1002/cncr.31101, **Received:** April 21, 2017; **Revised:** September 26, 2017; **Accepted:** October 2, 2017, **Published online** Month 00, 2017 in Wiley Online Library (wileyonlinelibrary.com)

In this context and with data from a large longitudinal AS cohort, we set out to determine whether the number of biopsies received over 10 years of AS could be reduced from an annual biopsy schedule without substantially increasing the time to detecting grade progression in cases where it occurred. To do this, we used longitudinal data from the Johns Hopkins AS study to conduct a hidden Markov model (HMM) analysis to estimate the initial biopsy sampling error, biopsy accuracy, and the rate of progression from low- to intermediate- or high-grade prostate cancer over time. We further conducted model validation and sensitivity analysis. Finally, we used the model to evaluate all possible follow-up surveillance strategies as well as previously proposed strategies for AS found in the literature on the basis of the mean delay time to the detection of grade progression and the planned number of biopsies over the first 10 years after the initiation of AS.

MATERIALS AND METHODS

Data

The Johns Hopkins AS study data included men enrolled in AS from 1992 to 2015. The study enrolled men with favorable-risk prostate cancer: clinical stage $\leq$ T1c, prostate-specific antigen (PSA) density ≤ 0.15, Gleason score ≤ 6, total positive cores ≤ 2, and single core positivity $\leq 50\%$. Because of patient preference, older men with low-risk disease (ie, clinical stage $\leq$ T2a, PSA level < 10 ng/mL, and Gleason score ≤ 6) were also enrolled in the study. The cohort was predominantly composed of men with low-volume prostate cancer (<3 cores and $<50\%$ core involvement) with a Gleason score of 6. The collected data included the PSA levels, age, and annual biopsy results (eg, Gleason score, number of positive cores, and maximum percentage of core involvement). The Johns Hopkins protocol includes semiannual PSA and digital rectal examinations and annual prostate biopsy. If a patient's biopsy results no longer meet the inclusion criteria, he is recommended for curative treatment. The data set used was anonymized with respect to patient identifiers, and the approval of the University of Michigan institutional review board was obtained before the initiation of the study.

HMM for Prostate Cancer Grade Progression

The specific type of model that we used was an HMM in which a patient's progress through health states was defined by his prognostic grade groups, which were based on the Gleason score, the most important clinical factor for assessing the risk of prostate cancer mortality. The term *hidden* refers to the fact that the exact health state of the patient was unknown in the absence of prostatectomy. The probability of progression to a higher prognostic grade group was determined by transition probabilities. We use the term *progression* to refer collectively to reclassification based on the undersampling of existing higher grade lesions, the true evolution of Gleason pattern 3 into Gleason pattern 4, or the occurrence of a de novo higher grade cancer. We based the model on 1-year time periods between state transitions to be consistent with the highest proposed frequency of biopsies and because that was the planned frequency of biopsies in the Johns Hopkins study.

We used the Baum-Welch algorithm to compute maximum likelihood estimates for the HMM parameters.[9] The Baum-Welch algorithm is a special case of the general expectation-maximization algorithm,[10] an iterative algorithm that combines forward and backward passes on a longitudinal observation sequence to find the choice of transition probabilities, observation probabilities, and initial distribution of patients that maximizes the likelihood of observing the collection of observed sequences. To initiate the Baum-Welch algorithm, we needed initial estimates of the model parameters: the annual progression rate from a Gleason score ≤ 6 to a Gleason score ≥ 7, the sensitivity and specificity of biopsy to disease with a Gleason score ≥ 7, and the initial proportion of patients undergoing AS with a Gleason score ≤ 6. These estimates are not directly observable in the data set because biopsies are imperfect; thus, we used estimates from the literature to initialize the algorithm. Alam et al[11] studied reclassification rates for men in the Johns Hopkins AS study and found that the majority of men were reclassified within the first 2 years, most likely because of initial biopsy misclassification. We estimated the annual rate of evolution from a Gleason score ≤ 6 to a Gleason score ≥ 7 or for the development of a de novo cancer with a Gleason score ≥ 7 to be 5% by calculating the rate of progression at a patient's 3rd through 13th biopsies. The estimates for the sensitivity and specificity of biopsy to disease with a Gleason score ≥ 7 were calculated to be 62.5% and 89.4%, respectively, on the basis of data reported by Epstein et al,[12] who compared biopsy results with the Gleason score at radical prostatectomy. Finally, using data reported by Epstein et al, we estimated that 74.9% of the patients diagnosed with Gleason score ≤ 6 on biopsy had Gleason score ≤ 6 at radical prostatectomy, whereas 25.1% had Gleason score ≥ 7 at radical prostatectomy. We used these estimates as the starting points for the Baum-Welch algorithm, and we ran the algorithm with stopping criteria defined by a tolerance of 10^{-6} on the

difference between the log likelihoods for consecutive iterations.

We used the resulting model to evaluate a large number of AS biopsy schedules, including schedules proposed in the literature. The University of California San Francisco recommends biopsy 1 year after diagnosis and then every 1 to 2 years. We modeled 2 versions of this policy: in version 1, biopsy is performed after 1 year and then every 1.5 years, and in version 2, biopsy is performed after 1 year and then every 2 years.[13] According to the Prostate Cancer Research International: Active Surveillance (PRIAS)/University of Toronto (UT) schedule, biopsy is performed after 1, 4, 7, and 10 years.[14]

Simulation Model

We used the HMM parameter estimates to simulate the mean delay time in detecting progression among patients who progress to high-grade cancer over the 10-year period after the diagnosis of prostate cancer for all 2^{10} possible AS biopsy strategies. The delay time depends on the HMM parameter estimates, which include the initial probability that a patient has a Gleason score ≤ 6 or a Gleason score ≥ 7 at the time of diagnosis, the annual transition probability from a Gleason score ≤ 6 to a Gleason score ≥ 7, and the sensitivity and specificity of biopsy to disease with a Gleason score ≥ 7. Together with the AS biopsy schedule, these parameters collectively govern the time to reach the high-grade cancer state and the subsequent detection of grade progression. We defined the biopsy schedule as a vector of binary decision variables indicating whether a biopsy is planned at a particular time period or not. We then simulated all 2^{10} possible AS strategies and evaluated the mean time to detect cancer with a Gleason score ≥ 7. Finally, we identified those strategies that were nondominated, that is, those strategies for which no other strategy simultaneously recommended fewer biopsies and had a lower mean time to detect high-grade cancer.

Model Validation and Sensitivity Analysis

To validate the results obtained, we used the base-case estimates of our model to simulate the detection rate on the basis of 10,000 samples under the assumption of annual biopsy as planned in the Johns Hopkins AS study protocol, and we compared the results with the observed detection rates in the Johns Hopkins data.

Next, we conducted experiments based on a hypothetical HMM for which we knew the true values for model parameters, and we tested our implementation of the Baum-Welch algorithm on sampled results for 1375 simulated patient observation sequences; this was the number of patients in the study who underwent their first surveillance biopsy. Because there were missing data in the Johns Hopkins study resulting from patients who discontinued AS in the absence of grade progression, we sought to test the assumption that the missing data were not informative. Therefore, we censored the data for simulated observation sequences according to the observed mean rate of patients discontinuing AS without grade progression. We then ran the Baum-Welch algorithm on the simulated data and compared the parameter estimates with the true parameters used to generate the simulated data.

To validate that the resulting parameter estimates were not sensitive to the initial parameter estimates, we varied our initial estimate for each parameter with a range of ± 0.1 and an upper limit of 0.99. We then ran the Baum-Welch algorithm on each new set of initial estimates and compared the resulting parameter estimates. We further performed a bootstrapping analysis for which we randomly sampled 1375 patients with replacement from the Johns Hopkins data set. We generated 30

TABLE 1. Patient Characteristics at the Time of Diagnosis

Characteristic	AS Cohort (n = 1499)
Age at diagnosis, No. (%)	
≤49 y	18 (1.2)
50-59 y	208 (13.9)
60-69 y	911 (60.8)
70-79 y	352 (23.5)
≥80 y	10 (0.7)
Race, No. (%)	
White	1314 (87.7)
Black	115 (7.7)
Other	60 (4.0)
NA	10 (0.7)
PSA at diagnosis, No. (%)	
0-2.5 ng/mL	162 (10.8)
2.5-4 ng/mL	249 (16.6)
4-6 ng/mL	558 (37.2)
6-10 ng/mL	322 (21.5)
>10 ng/mL	85 (5.7)
NA	123 (8.2)
PSA density at diagnosis, No. (%)	
0-0.05	166 (11.1)
0.05-0.10	538 (35.9)
0.10-0.15	428 (28.6)
0.15-0.20	134 (8.9)
>0.20	114 (7.6)
NA	119 (7.9)
Gleason score at diagnosis, No. (%)	
≤6	1488 (99.3)
3 + 4	5 (0.3)
4 + 3	1 (0.1)
NA	5 (0.3)

Abbreviations: AS, active surveillance; NA, not available; PSA, prostate-specific antigen.

TABLE 2. Biopsy Characteristics at Diagnosis and During Surveillance

Characteristic	Biopsy							
	Diagnosis	First	Second	Third	Fourth	Fifth	Sixth	Seventh
No. of patients	1493	1370	922	644	447	298	187	122
Age at biopsy, mean (SD), y	66 (6.0)	67 (6.1)	67 (6.0)	68 (5.5)	68 (5.3)	69 (5.1)	70 (5.1)	71 (4.3)
Time since last biopsy, mean (SD), mo	0 (0.0)	13 (8.2)	15 (7.5)	15 (7.1)	16 (8.6)	15 (6.9)	16 (7.6)	14 (3.9)
Most recent PSA level, mean (SD), ng/mL	5.3 (2.9)	5.4 (3.4)	5.5 (4.1)	5.5 (4.0)	5.9 (4.9)	5.9 (4.2)	6.4 (5.1)	6.0 (4.7)
Most recent PSA density, mean (SD)	0.12 (0.07)	0.11 (0.07)	0.11 (0.08)	0.10 (0.07)	0.10 (0.09)	0.10 (0.07)	0.10 (0.08)	0.09 (0.07)
No. of biopsy cores, median (range)	12 (6-58)	12 (4-31)	12 (6-60)	12 (6-28)	12 (8-18)	12 (6-16)	14 (6-24)	14 (6-15)
Cores positive for cancer, No. (%)								
0%	0 (0.0)	568 (41.5)	435 (47.2)	336 (52.2)	237 (53.0)	154 (51.7)	97 (51.9)	57 (46.7)
>0% and <34%	808 (54.1)	624 (45.5)	415 (45.0)	271 (42.1)	188 (42.1)	129 (43.3)	79 (42.2)	58 (47.5)
≥34%	10 (0.7)	58 (4.2)	20 (2.2)	7 (1.1)	9 (2.0)	5 (1.7)	5 (2.7)	2 (1.6)
NA	675 (45.2)	120 (8.8)	52 (5.6)	30 (4.7)	13 (2.9)	10 (3.4)	6 (3.2)	5 (4.1)
Gleason score, No. (%)								
No cancer	0 (0.0)	568 (41.5)	435 (47.2)	336 (52.2)	237 (53.0)	154 (51.7)	97 (51.9)	57 (46.7)
≤6	1488 (99.7)	670 (48.9)	413 (44.8)	275 (42.7)	181 (40.5)	130 (43.6)	78 (41.7)	54 (44.3)
7 (3 + 4)	0 (0.0)	78 (5.7)	49 (5.3)	16 (2.5)	18 (4.0)	9 (3.0)	10 (5.3)	5 (4.1)
7 (4 + 3)	0 (0.0)	30 (2.2)	14 (1.5)	12 (1.9)	6 (1.3)	4 (1.3)	1 (0.5)	3 (2.5)
≥8	0 (0.0)	18 (1.3)	7 (0.8)	1 (0.2)	2 (0.4)	1 (0.3)	0 (0.0)	1 (0.8)
NA	5 (0.3)	6 (0.4)	4 (0.4)	4 (0.6)	3 (0.7)	0 (0.0)	1 (0.5)	2 (1.6)
Outcome, No. (%)								
Progression	0 (0.0)	126 (9.2)	70 (7.6)	29 (4.5)	26 (5.8)	14 (4.7)	11 (5.9)	9 (7.4)
No progression	1493 (100.0)	1244 (90.8)	852 (92.4)	615 (95.5)	421 (94.2)	284 (95.3)	176 (94.1)	113 (92.6)

Abbreviations: NA, not available; PSA, prostate-specific antigen; SD, standard deviation.

different bootstrap samples and ran the Baum-Welch algorithm on each sample, and we compared the resulting parameter estimates.

Finally, we performed an analysis to assess the potential error in HMM parameter estimates due to missing data for patients who left the study for reasons other than grade progression. We generated 100 different simulated data sets with a hypothetical model based on parameters from our base-case analysis. For each data set, we sampled 1375 patients with patient dropout based on point estimates of the dropout rate in the cohort. We then ran the Baum-Welch algorithm for each of the 100 simulated data sets and compared the resulting parameter estimates with the base-case results.

RESULTS

Data

There were 1521 patients in the data set, and we removed 22 patients from the data set because of missing diagnostic biopsy information. Table 1 describes the patient characteristics at diagnosis for the remaining 1499 patients. Among the men who discontinued AS and received treatment, 50.9% received surgery, and 46.2% received radiation therapy. The mean and variance of the time between biopsies were 14.2 and 60.1 months, respectively. The median number of biopsies per patient, including diagnosis biopsy, was 3, and the range was 1 to 14. Table 2 shows the biopsy characteristics; we have defined *progression* to be a transition from a Gleason score ≤ 6 to a Gleason score ≥ 7 on biopsy. Because of this definition, we excluded 6 additional patients initially diagnosed with disease with a Gleason score of 7 from the analysis in Table 2. Among the remaining patients, there were 1375 patients who had at least 1 surveillance biopsy after the diagnosis biopsy. The median time between biopsies was 1 year, and the biopsy compliance rate for the first 2 years was 90%. The 6-year compliance rate was 80% for men < 75 years old and 50% for men ≥ 75 years old.

HMM Analysis

Using the Baum-Welch algorithm, we estimated the annual progression rate from a Gleason score ≤ 6 to a Gleason score ≥ 7 to be 4.0%; the sensitivity and specificity of biopsy for disease with a Gleason score ≥ 7 to be 61.0% and 98.6%, respectively; and the initial proportion of patients undersampled at the time of diagnosis with Gleason score ≥ 7 to be 9.8%.

Simulation Model of AS Strategies

Our simulation model found that 40% of patients progress to higher grade cancer in 10 years, and a strategy that performs annual biopsies (the Johns Hopkins strategy) takes a mean of 14.1 months to detect progression. The strategies minimizing the mean delay time for each choice of planned number of biopsies over 10 years are plotted in

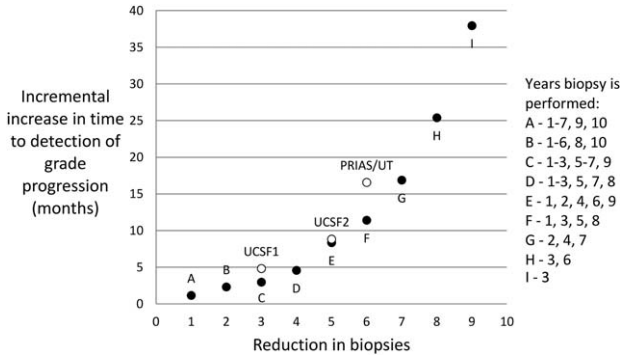

Figure 1. Simulation results for nondominated active-surveillance strategies from 2^{10} simulated strategies and published strategies based on the estimated hidden Markov model parameters. The incremental time to detection and the reduction in biopsies are with respect to an annual biopsy strategy. The mean time to the detection of grade progression for an annual biopsy plan is 14.1 months. PRIAS indicates Prostate Cancer Research International: Active Surveillance; UCSF, University of California San Francisco; UT, University of Toronto.

Figure 1, which shows the incremental time to detection and the reduction in biopsies with respect to a strategy that performs annual biopsies. Figure 1 shows that the University of California San Francisco strategy performs well in comparison with the optimal strategy, whereas the PRIAS/UT strategy increases the mean time to detection by 5.2 months in comparison with an optimal strategy that performs the same number of biopsies at years 1, 3, 5, and 8. The optimal strategy performs biopsies earlier than the PRIAS/UT strategy.

The simulation model found that under an annual biopsy strategy 13.5%, 4.9%, and 1.8% of patients would be detected more than 12, 24, and 36 months after grade progression, respectively. Table 3 presents the increased risk of various biopsy schedules with respect to the annual schedule. For example, strategy F eliminates 6 biopsies in the first 10 years of AS but only increases a patient's risk of detection with grade progression after more than 24 months by 9.8% in comparison with an annual biopsy strategy. Table 3 also shows that our optimal strategies perform better than the published strategies with the same number of biopsies. For example, strategies F and the PRIAS strategy both perform 4 biopsies in the first 10 years of AS; however, strategy F decreases a patient's risk of >24 months to detect grade progression by 3.6 percentage points in comparison with the PRIAS strategy.

HMM Validation and Sensitivity Analysis

To validate the obtained results, we used the base-case estimates of our model to simulate the detection rate under the assumption of annual biopsy as planned in the Johns Hopkins AS study protocol. Figure 2 shows that the difference between the model-based results and the observed results were small. Model-predicted results were based on 10,000 samples.

Modifications to the definition of a Gleason score of 6 versus a Gleason score of 7 by the International Society of Urological Pathology in 2005 may have caused changes in the rate of grade reclassification. For this reason, we also performed a second analysis of only those patients diagnosed during or after 2005 (n = 995). The results for annual progression were slightly higher (4.6% vs 4%), and the misclassification rate was slightly lower (8.7% vs 9.8%), as would be expected from more stringent criteria.

The results for the hypothetical HMM for which the true values for model parameters are known are presented in Table 4, which shows the true model parameters from the hypothetical model and the 95% confidence intervals for our model parameter estimates based on the Baum-Welch algorithm applied to 1000 sets of simulated biopsy data with 1375 sequences (ie, the number of patients who underwent surveillance biopsies in the Johns Hopkins cohort) per set.

After varying the starting points of model parameters, we found that the resulting parameter estimates varied by less than 0.5% from the values calculated with our original starting points, and this suggests that the starting points did not significantly affect our parameter estimates. The 95% confidence intervals based on bootstrapping are presented in Table 5, with the sensitivity of biopsy to prostate cancer with a Gleason score ≥ 7 having the most variation.

In the final sensitivity analysis, we fit HMMs to 100 simulated data sets that emulated the characteristics of missing data in the cohort as described in the Materials and Methods section. We used the results of the base-case analysis to define the true model parameters and simulated observations in the context of missing data due to patient dropout. The means for progression, biopsy sensitivity and specificity, and misclassification due to undersampling were 4.00% (95% confidence interval, 3.98%-4.02%), 62.5% (95% confidence interval, 59.8%-64.77%), 98.76% (95% confidence interval, 98.48%-99.04%), and 8.96% (95% confidence interval, 8.16%-9.76%), respectively. These simulated results compare very favorably with the base-case results, and this suggests that it is reasonable to assume that missing data due to patient drop out are not informative.

DISCUSSION

AS for prostate cancer includes follow-up with serial prostate biopsies; however, the optimal biopsy frequency

TABLE 3. Changes in the Risk of Delay of Detection of Higher Grade Disease for >1, > 2, and >3 Years in Comparison With an Annual Biopsy Strategy

Strategy Label	No. of Biopsies in 10 y	Increased Risk of > 12 mo to Detect Grade Progression, %[a]	Increased Risk of > 24 mo to Detect Grade Progression, %[a]	Increased Risk of > 36 mo to Detect Grade Progression, %[a]
A	9	1.5	1.2	0.5
B	8	3.0	1.9	1.3
C	7	3.4	2.7	1.4
D	6	5.2	3.5	2.4
E	5	7.1	6.6	5.6
F	4	9.1	9.8	6.4
G	3	17.0	11.6	9.6
H	2	19.3	21.5	12.2
I	1	21.3	25.3	17.8
PRIAS	4	11.0	13.4	10.3
UCSF1	7	4.9	6.1	2.9
UCSF2	5	7.4	8.2	5.1

Abbreviations: PRIAS, Prostate Cancer Research International: Active Surveillance; UCSF, University of California San Francisco.
An annual biopsy strategy resulted in a 13.5% risk of > 12 months, a 4.9% risk of > 24 months, and a 1.8% risk of > 36 months to detect grade progression.
[a] In comparison with an annual biopsy strategy.

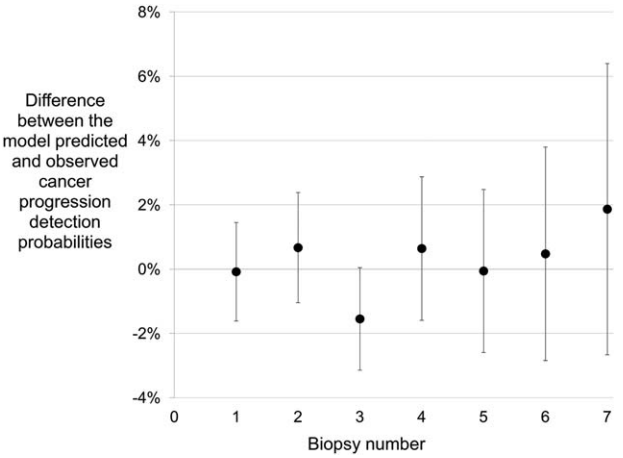

Figure 2. Difference between the biopsy detection rate predicted by the simulation model and the observed rate in the Johns Hopkins study. The model-predicted results were based on 10,000 samples. The confidence intervals for the observed results are shown; the confidence intervals for the model-predicted results are too small to see in the figure.

TABLE 4. Results From a Comparison of HMM Parameter Estimates From the Baum-Welch Algorithm With the True Model Parameter Estimates From a Known Model

Model Parameter	True Value	HMM 95% CI Estimate
Proportion of patients with GS ≤ 6 at diagnosis	0.866	0.846-0.889
Annual progression rate	0.040	0.033-0.039
Sensitivity of biopsy to prostate cancer with GS ≥ 7	0.610	0.582-0.688
Specificity of biopsy to prostate cancer with GS ≥ 7	0.986	0.980-0.988

Abbreviations: CI, confidence interval; GS, Gleason score; HMM, hidden Markov model.
The 95% confidence interval is based on 1000 sets of simulated biopsy data with 1375 sequences per set.

during follow-up has not been determined. The goal of this investigation was to use longitudinal AS biopsy data to assess whether the frequency of biopsy could be reduced without substantially prolonging the time to the detection of disease with a Gleason score ≥ 7. Using longitudinal data from 1375 men with favorable-risk prostate cancer enrolled in AS at Johns Hopkins who underwent at least 1 surveillance biopsy, we developed an HMM to estimate the probability of undersampling at diagnosis, the annual probability of grade progression, and the 10-year cumulative probability of reclassification or progression to a Gleason score ≥ 7. We simulated 1024 potential AS biopsy strategies for the 10 years after the diagnosis and predicted the mean delay in the detection of disease with a Gleason score ≥ 7. Although annual biopsy for low-risk men on AS is associated with the shortest time to the detection of disease with a Gleason score ≥ 7, several alternative strategies may allow less frequent biopsying without sizable delays in detecting grade progression.

Many experts have called for the use of AS to address overtreatment concerns for men with low-risk prostate cancer. AS delays and possibly prevents immediate treatment via surgery or radiation therapy until and unless there is evidence that the disease has progressed; however, it comes with a burden to patients because of the need to conduct follow-up clinical examinations, tests, and surveillance biopsies. The intensiveness of follow-up determines

TABLE 5. Bootstrapping Results Based on 30 Different Bootstrap Samples of 1375 Patients

Model Parameter	95% CI
Proportion of patients with GS $\leq$ 6 at diagnosis	0.857-0.876
Annual progression rate	0.032-0.040
Sensitivity of biopsy to prostate cancer with GS $\geq$ 7	0.587-0.635
Specificity of biopsy to prostate cancer with GS $\geq$ 7	0.980-0.988

Abbreviations: CI, confidence interval; GS, Gleason score.

the frequency of clinical examinations, tests, and biopsies. In the absence of randomized trials comparing AS pathways, there is no consensus among urologists about the best way to trade off the burden of surveillance with the benefits of avoiding cancer progression.[15] We provide a new model for AS that quantifies the trade-off between benefits and harms of various AS strategies. These decisions must trade off between the potential long-term benefits of detecting disease progression and the burden of surveillance, including the potential harms and side effects of biopsies (eg, pain, anxiety, and hospitalization for infection in 2%-3% of cases).

There are multiple definitions of progression for prostate cancer, including definitions based on increases in PSA, PSA velocity and density, and tumor volume. Grade progression, which refers to a change in the Gleason score, is a definitive form of progression recognized by all published AS guidelines. However, currently, there is debate about whether grade progression is possible or the occurrence of higher grade cancer on biopsy occurs because of biopsy sampling error. Some studies suggest that a combination of sampling error, true progression, and development of de novo cancer are responsible for increased grade detection over time.[14,16] Our findings lend additional evidence to these studies and suggest that a combination of mechanisms are responsible for the detection of higher grade cancers in the future. This suggests that there may be benefits from more frequent biopsies after diagnosis and less frequent biopsies in later years. This is supported by our simulation results. If grade progression does not occur, as some believe, then the incremental time to detection reported in Figure 2 would be even lower.

A chief concern about AS is the possibility that prostate cancer progresses in the interval of time between biopsies or that progression is missed because of the imperfect sensitivity of biopsies. The potential for undetected progression raises questions about health outcomes for patients on AS who progress and receive treatment. Studies comparing radical prostatectomy outcomes for patients initially on AS and patients undergoing radical prostatectomy immediately after diagnosis have shown that low-risk men who undergo an annual biopsy on AS do not have worse surgery outcomes.[4] In addition, Klotz et al[17] reported that patients undergoing AS with biopsies every 3 to 4 years had mortality rates consistent with those of patients who received initial definitive treatment. Assuming a uniform distribution of progression times during the 3- to 4-year intervals would suggest that delays of approximately 18 to 24 months in detecting grade progression may not have a clinically significant impact. Our results suggest that reducing the number of biopsies by half (in comparison with an annual biopsy strategy) would result in an incremental increase in the time to the detection of grade progression that is well below these estimates.

Our results pertain to patients with disease with a Gleason score of 6, and they are not informative for patients with disease with a Gleason score of 7 (particularly 3 + 4), some of whom may be good candidates for AS. However, there is no clear consensus on an acceptable safe cancer volume limit for patients with a Gleason score of 3 + 4. Therefore, this remains an important question for future study.[18] Our findings are not applicable to patients enrolling in AS with disease with a Gleason score of 7. Nevertheless, Liu et al[19] reported that 80% of patients enrolling in AS are diagnosed with Gleason score 6, so our results would be applicable to the majority of patients enrolling in AS.

Our study has some notable limitations. First, our results apply to patients with favorable-risk prostate cancer (ie, clinical stage $\leq$ T1c, PSA density $\leq$ 0.15, Gleason score $\leq$ 6, total positive cores $\leq$ 2, and single core positivity $\leq$ 50%) and older men with low-risk disease (ie, clinical stage $\leq$ T2a, PSA level $<$ 10 ng/mL, and Gleason score $\leq$ 6) because these were the patients who were enrolled in the Johns Hopkins AS study, and thus there is a need to validate our findings in other AS studies; however, this initial study lays the groundwork for such future validation work. Our results provide the trade-off between the number of biopsies and the mean delay time to the detection of progression; however, the amount of time that is considered safe to delay detection is not known. Nevertheless, data from the literature provide evidence that short delay times do not have a significant clinical impact. Metastasis is a better endpoint, but the data needed to fit an HMM with this endpoint do not yet exist.

Although compliance with annual biopsies in this cohort was high and this made it an ideal source for the

HMM analysis that we presented, it was not perfect. Moreover, there was patient dropout for reasons other than biopsy-confirmed progression. Thus, there is the potential for a selection bias to influence the results. Although this cannot be eliminated, we mitigated this risk in part by conducting simulation analyses, which showed that the HMM analysis was not sensitive to missing data. Finally, our model assumes that progression rates do not vary over time and that PSA is not used to initiate treatment. We do not believe that this is a strong assumption because PSA kinetics are not used to trigger intervention in the Johns Hopkins study. These limitations notwithstanding, we believe that this study provides important evidence about the trade-off between various AS strategies and the optimal timing of biopsies during AS.

Although annual biopsy for low-risk men on AS is associated with the shortest time to the detection of disease with a Gleason score ≥ 7, several alternative strategies may allow less frequent biopsying without clinically significant increases in the time to detecting grade progression. For instance, on the basis of the model results, performing biopsies in years 1, 3, 5, and 8 would increase the time to detection by less than a year in comparison with an annual biopsy schedule over a 10-year period. In addition, the optimal model-based biopsy schedules tend to perform more biopsies in the beginning because of the risk of undersampling. External validation of the results with other AS studies is warranted to assess the degree to which surveillance biopsies can be safely attenuated.

FUNDING SUPPORT

This material is based on work supported in part by the National Science Foundation (through grant CMMI 0844511 to Brian T. Denton and grant DGE 1256260 to Christine L. Barnett). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CONFLICT OF INTEREST DISCLOSURES

David C. Miller reports a contract for serving as director of the Michigan Urological Surgery Improvement Collaborative from Blue Cross Blue Shield of Michigan outside the submitted work.

AUTHOR CONTRIBUTIONS

Christine L. Barnett: Conceptualization, methodology, software, validation, formal analysis, investigation, writing, visualization, and project administration. **Gregory B. Auffenberg**: Conceptualization and writing. **Zian Cheng**: Software, validation, and formal analysis. **Fan Yang**: Software, validation, and formal analysis. **Jiachen Wang**: Software, validation, and formal analysis. **John T. Wei**: Conceptualization and writing. **David C. Miller**: Conceptualization and writing. **James E. Montie**: Conceptualization and writing. **Mufaddal Mamawala**: Data curation. **Brian T. Denton**: Conceptualization, methodology, resources, writing, visualization, supervision, project administration, and funding acquisition.

REFERENCES

1. Miller DC, Gruber SB, Hollenbeck BK, Montie JE, Wei JT. Incidence of initial local therapy among men with lower-risk prostate cancer in the United States. *J Natl Cancer Inst.* 2006;98: 1134-1141.
2. Anandadas CN, Clarke NW, Davidson SE, et al. Early prostate cancer—which treatment do men prefer and why? *BJU Int.* 2011;107: 1762-1768.
3. 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.
4. 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;29:2185-2190.
5. Gonzalez C, Averch T, Boyd L. AUA/SUNA white paper on the incidence, prevention and treatment of complications related to prostate needle biopsy. https://www.suna.org/resources/pnbWhitePaper.pdf. Accessed October 26, 2017.
6. Ehdaie B, Vertosick E, Spaliviero M, et al. The impact of repeat biopsies on infectious complications in men with prostate cancer on active surveillance. *J Urol.* 2014;191:660-664.
7. Loeb S, Folkvaljon Y, Makarov DV, Bratt O, Bill-Axelson A, Stattin P. Five-year nationwide follow-up study of active surveillance for prostate cancer. *Eur Urol.* 2015;67:233-238.
8. Al Otaibi M, Ross P, Fahmy N, et al. Role of repeated biopsy of the prostate in predicting disease progression in patients with prostate cancer on active surveillance. *Cancer.* 2008;113:286-292.
9. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE.* 1989;77:257-286.
10. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B Stat Methodol.* 1977;39:1-38.
11. Alam R, Carter HB, Landis P, Epstein JI, Mamawala M. Conditional probability of reclassification in an active surveillance program for prostate cancer. *J Urol.* 2015;193:1950-1955.
12. Epstein JI, Feng Z, Trock BJ, Pierorazio PM. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. *Eur Urol.* 2012;61:1019-1024.
13. 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.
14. van den Bergh RC, Roemeling S, Roobol MJ, Roobol W, Schroder FH, Bangma CH. Prospective validation of active surveillance in prostate cancer: the PRIAS study. *Eur Urol.* 2007;52:1560-1563.
15. Lawrentschuk N, Klotz L. Active surveillance for low-risk prostate cancer: an update. *Nat Rev Urol.* 2011;8:312-320.
16. Inoue LY, Trock BJ, Partin AW, Carter HB, Etzioni R. Modeling grade progression in an active surveillance study. *Stat Med.* 2014;33: 930-939.
17. Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. *J Clin Oncol.* 2014;33:272-277.
18. Patel HD, Tosoian JJ, Carter HB, Epstein JI. Adverse pathologic findings for men electing immediate radical prostatectomy: defining a favorable intermediate-risk group. *JAMA Oncol*. In press.
19. Liu J, Womble PR, Merdan S, Miller DC, Montie JE, Denton BT; Michigan Urological Surgery Improvement Collaborative. Factors influencing selection of active surveillance for localized prostate cancer. *Urology.* 2015;86:901-905.