Optimizing Active Surveillance Strategies to Balance Competing Goals of Early Detection of Grade Progression with Minimizing Harm from Biopsies

Running Title: Optimizing Active Surveillance for Prostate Cancer

Christine L. Barnett [a], Gregory B. Auffenberg [a], Zian Cheng [a], Fan Yang [a], Jiachen Wang [a], John T. Wei [a], David C. Miller [a], James E. Montie [a], Mufaddal Mamawala [b], and Brian T. Denton [a]

[a] University of Michigan, Ann Arbor, MI; [b] Johns Hopkins University, Baltimore, MD

Corresponding Author:

Brian T. Denton, PhD.

2893 IOE Building, 1205 Beal Ave

Ann Arbor, MI 48109-2117

Phone: 734-763-2060, Fax: 734-764-3451, Email: btdenton@umich.edu

Financial Support: This work was supported by the National Science Foundation (CMMI 0844511 to BTD, DGE 1256260 to CLB).

Conflicts of Interest: There are no conflict of interest disclosures from any authors.

Condensed Abstract: Active surveillance for low-risk prostate cancers can reduce the harms of overtreatment; however, the optimal timing of surveillance biopsies is unknown. It is possible to perform fewer biopsies than previously recommended without significant delays in detection of grade progression.

Keywords: Active surveillance; prostate cancer; Markov model; biopsy; reclassification

Author Contributions:

Barnett: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing, Visualization, Project administration

Auffenberg: Conceptualization, Writing

Cheng: Software, Validation, Formal analysis

Yang: Software, Validation, Formal analysis

Wang: Software, Validation, Formal analysis

Wei: Conceptualization, Writing

Miller: Conceptualization, Writing

Montie: Conceptualization, Writing

Mamawala: Data curation

Denton: Conceptualization, Methodology, Resources, Writing, Visualization, Supervision, Project administration, Funding acquisition

Acknowledgements

This material is based upon work supported in part by the National Science Foundation through Grant Number CMMI 0844511, and by the National Science Foundation Graduate Research Fellowship under Grant Number DGE 1256260. 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. The authors would like to thank H. Ballentine Carter for providing the data from the Johns Hopkins Active Surveillance Study.

ABSTRACT

Background: Active surveillance (AS) for prostate cancer (PCa) 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 if the frequency of biopsy could be reduced without substantially prolonging the time to detection of Gleason ≥7 disease.

Methods: Using data from 1375 men with low-risk PCa enrolled in AS at Johns Hopkins, we developed a hidden Markov model (HMM) to estimate the probability of under sampling at diagnosis, the annual probability of grade progression, and the 10-year cumulative probability of reclassification or progression to Gleason ≥7. We simulated 1024 potential AS biopsy strategies for the 10 years following diagnosis. For each of these strategies the model predicted the mean delay in detection of Gleason ≥7 disease.

Results: The model estimated 10-year cumulative probability of reclassification from Gleason 6 to Gleason ≥7 was 46.0%. The probability of under sampling at diagnosis was 9.8% and the annual progression probability for men with Gleason 6 was 4.0%. Based on these estimates, simulation of an annual biopsy strategy estimated the mean time to detection of Gleason ≥7 disease to be 14.1 months; however, several strategies eliminated biopsies with only small (<12 months) delays in detecting grade progression.

Conclusions: While annual biopsy for low-risk men on AS is associated with the shortest time to detection of Gleason ≥7 disease, several alternative strategies may allow for less frequent biopsy without sizable delays in detecting grade progression.

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 given that 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 by conducting regular clinical exams, biomarker tests, radiologic imaging, and biopsies. Due to concerns 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).

Due to 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. Risk of progression is one of the most important considerations when weighing long-term risk for patients on AS. The ideal strategy to minimize risk of delaying the detection of high-grade cancer progression is to biopsy patients frequently (e.g. annually as suggested by Tosoian et al. (4)). However, this risk competes with the harms of frequent biopsies resulting in pain and anxiety for patients, and the potential for complications, such as infection. Severe infection rates for biopsy are approximately 1-2% (5); however, recent studies suggest that infection rates for patients undergoing AS increases as a function of the number of biopsies

they have received (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).

In this context and using data from a large longitudinal AS cohort, we set out to determine if 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. In order to do this, we used longitudinal data from the Johns Hopkins AS study to conduct a hidden Markov model analysis to estimate 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 mean delay time to detect grade progression and the planned number of biopsies over the first 10 years following initiation of AS.

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 ≤T1c, PSA density ≤0.15, Gleason score ≤6, total positive cores ≤2, and single core positivity ≤50\%. Due to patient preference, older men with low-risk disease (i.e., clinical stage ≤T2a, PSA <10 ng/mL, and Gleason score ≤6) were also enrolled in the study. The cohort was predominantly comprised of men with low volume Gleason score 6 prostate cancer (< 3 cores and <50% core involvement). The data collected included PSA, age, and annual biopsy results (e.g., Gleason score, number of

positive cores, maximum percentage core involvement). The Johns Hopkins protocol includes semiannual PSA and digital rectal examination and annual prostate biopsy. If a patient's biopsy results no longer meet the inclusion criteria, they are recommended for curative treatment. The dataset used was anonymized with respect to patient identifiers and approval of the University of Michigan IRB was obtained prior to initiation of the study.

Hidden Markov Model for Prostate Cancer Grade Progression

The specific type of model we employ is a *hidden Markov model* (HMM) in which patient's progress through health states as defined by their *prognostic grade groups* based on Gleason score, the most important clinical factor for assessing risk of prostate cancer mortality. The term *hidden* refers to the fact that the exact health state of the patient is unknown, in the absence of prostatectomy. The probability of progression to a higher prognostic grade group is determined by *transition probabilities*. We use the term "progression" to refer collectively to reclassification based on under sampling of existing higher grade lesions, true evolution of Gleason pattern 3 into pattern 4, or the occurrence of a "de novo" higher grade cancer. We based the model on one-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 (EM) 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: annual progression rate from Gleason score ≤ 6 to Gleason score ≥ 7, the sensitivity and specificity of biopsy to Gleason score ≥ 7 disease, and the initial proportion of patients undergoing AS with Gleason score ≤ 6 disease. These estimates are not directly observable in the dataset, because biopsies are imperfect; thus, we used estimates from the literature to initialize the algorithm. Alam 2015 (11) studied reclassification rates for men in the Johns Hopkins AS study, and found that the majority of men are reclassified within the first two years, most likely due to initial biopsy misclassification. We estimated annual rate of evolution from Gleason score ≤ 6 to Gleason score ≥ 7 or for development of a de novo Gleason score ≥ 7 cancer to be 5% by calculating the rate of progression at patient's third through thirteenth biopsies. Estimates for the sensitivity and specificity of biopsy to Gleason score ≥ 7 disease were calculated to be 62.5% and 89.4%, respectively, based on data reported in Epstein et al. (12), which compared biopsy results to Gleason score at radical prostatectomy. Finally, using data reported in Epstein et al. (2012), we estimated that 74.9% of patients diagnosed with Gleason score ≤ 6 disease on biopsy have Gleason score ≤ 6 disease at radical prostatectomy, while 25.1% have Gleason score ≥ 7 disease at radical prostatectomy. We used these estimates as the starting points for the Baum-Welch algorithm, and we ran the algorithm using a 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. University of California, San Francisco (UCSF) recommends a biopsy 1 year after diagnosis, then every 1 to 2 years. We modeled two versions of this policy: UCSF1 performs a biopsy after 1 year, then every 1.5 years and UCSF2 performs

a biopsy after 1 year, then every 2 years (13). PRIAS/University of Toronto (UT) performs a biopsy after 1, 4, 7, and 10 years (14).

Simulation Model

We used the HMM parameter estimates to simulate mean delay time in detecting progression among patients who progress to high-grade cancer over the 10-year period following 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 a patient has Gleason score ≤ 6 or Gleason score ≥ 7 at the time of diagnosis, the annual transition probability from Gleason score ≤ 6 to Gleason score ≥ 7, and the sensitivity and specificity of biopsy to Gleason score ≥ 7 disease. Together with the AS biopsy schedule these parameters collectively govern the time to reach the high-grade cancer state and subsequent detection of grade progression. We defined the biopsy schedule as a vector of binary decision variables that indicated whether a biopsy is planned at a particular time period or not. We then simulated all 2^{10} possible AS strategies and evaluated mean time to detect Gleason score ≥ 7 cancer. Finally, we identified those strategies that were non-dominated, i.e., those strategies for which no other strategy simultaneously recommended fewer biopsies and had 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 based on 10,000 samples assuming annual biopsy as planned in the Johns Hopkins AS study protocol, and compared the results to 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, which is the number of patients in the study who received their first surveillance biopsy. Since 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 to 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 using a range of ±0.1 with 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 bootstrapping analysis for which we randomly sampled 1375 patients with replacement from the Johns Hopkins dataset. We generated 30 different bootstrap samples and ran the Baum-Welch algorithm on each sample, and compared the resulting parameter estimates.

Finally, we performed 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 datasets using a hypothetical model based on parameters from our base case analysis. For each dataset, we sample 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 datasets and compared the resulting parameter estimates to the base case results.

RESULTS

Data

There were 1521 patients in the dataset and we removed 22 patients from the dataset due to missing diagnostic biopsy information. Table 1 describes the patient characteristics at diagnosis of the remaining 1499 patients. Among 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 was 14.2 and 60.1 months, respectively. The median number of biopsies per patient, including diagnosis biopsy, was 3 and ranged from 1 to 14. Table 2 shows the biopsy characteristics, where we have defined progression to be transition from a Gleason score ≤ 6 to Gleason score ≥ 7 on biopsy. Due to this definition, we excluded six additional patients initially diagnosed with Gleason 7 disease from the analysis in Table 2. Among the remaining patients there were 1375 patients who had at least one surveillance biopsy following the diagnosis biopsy. The median time between biopsies was one year and the biopsy compliance for the first two years is 90%. The 6-year compliance rate for men < 75 years is 80% and 50% in men > 75 years.

Hidden Markov Model Analysis

Using the Baum-Welch algorithm, we estimated the annual progression rate from Gleason score ≤ 6 to Gleason score ≥ 7 to be 4.0%; the sensitivity and specificity of biopsy to Gleason score ≥ 7

disease to be 61.0% and 98.6%, respectively; and the initial proportion of patients under sampled at the time of diagnosis, with Gleason score ≥ 7 disease, 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 that a strategy that performs annual biopsies (the Johns Hopkins strategy) takes a mean of 14.1 months to detect progression. The strategy that minimized the mean delay time for each choice of planned number of biopsies over 10 years are plotted in Figure 1, which shows the incremental time to detection and the reduction in biopsies relative to a strategy that performs annual biopsies. Figure 1 shows that UCSF performs well compared to the optimal policy, while the PRIAS/UT increases mean time to detection by 5.2 months compared to an optimal policy that performs the same number of biopsies, at years 1, 3, 5, and 8. The optimal policy performs biopsies earlier than the PRIAS/UT policy.

The simulation model found that 13.5%, 4.9%, and 1.8% of patients would be detected more than 12 months, 24 months, and 36 months after grade progression, respectively. Table 3 presents the increased risk of varying biopsy schedules relative to the annual schedule. For example, policy F eliminates six biopsies in the first 10 years of active surveillance, while only increasing a patient's risk of being detection with grade progression after more than 24 months by 9.8% relative to an annual biopsy strategy. Table 3 also shows that our optimal policies perform better than the published policies using the same number of biopsies. For example, policies F and the PRIAS policy both perform four biopsies in the first 10 years of active surveillance; however, policy F decreases a patient's risk of >24 months to detect grade progression by 3.6 percentage points compared to the PRIAS policy.

HMM Validation and Sensitivity Analysis

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

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

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 interval for our model parameter estimates based on the Baum-Welch algorithm applied to 1000 sets of simulated biopsy data with 1375 sequences (i.e., the number of patients who received surveillance biopsies in the Johns Hopkins cohort) per set.

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

In the final sensitivity analysis, we fit HMMs to 100 simulated datasets that emulated the characteristics of missing data in the cohort as described in the 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 mean and 95% confidence intervals for progression, biopsy sensitivity and specificity, and misclassification due to under sampling were 4.00% [3.98%, 4.02%], 62.5% [59.8%, 64.77], 98.76 [98.48, 99.04], and 8.96 [8.16, 9.76]. These simulated results compare very favorably with the base case results suggesting it is reasonable to assume that missing data due to patient drop out is not informative.

DISCUSSION

AS for prostate cancer includes follow-up with serial prostate biopsies; however, 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 if the frequency of biopsy could be reduced without substantially prolonging the time to detection of Gleason ≥7 disease. Using longitudinal data from 1375 men with "favorable-risk" PCa enrolled in AS at Johns Hopkins who received at least one surveillance biopsy, we developed a HMM to estimate the probability of under sampling at diagnosis, the annual probability of grade progression, and the 10-year cumulative probability of reclassification or progression to Gleason ≥7. We simulated 1024 potential AS biopsy strategies for the 10 years following diagnosis and predicted the mean delay in detection of Gleason ≥7 disease. While annual biopsy for low-risk men on AS is associated with the shortest time to detection of Gleason ≥7 disease, several alternative strategies may allow for less frequent biopsy 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 avoids 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 due to the need to conduct follow-up clinical exams, tests, and surveillance biopsies. The intensiveness of follow-up determines the frequency of clinical exams, tests, and biopsies. In the absence of randomized trials comparing AS pathways there is no consensus among urologists about the best way to tradeoff the burden of surveillance with the benefits of avoiding cancer progression (15). We provide a new model for AS that quantifies the tradeoff between benefits and harms of various AS strategies. These decisions must tradeoff between the potential long-term benefits of detecting disease progression with the burden of surveillance, including the potential harms and side effects of biopsies (e.g. pain, anxiety, and hospitalization for infection in 2–3% of cases).

There are multiple definitions of progression for prostate cancer, including definitions based on increase in PSA, PSA velocity and density, and tumor volume. Grade progression, which refers to a change in 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 if the occurrence of higher grade cancer on biopsy occurs due to biopsy sampling error. Some studies suggest that a combination of sampling error, true progression, and development of a de novo cancer are responsible for increased grade detection over time (14,16). Our findings lend additional evidence to these studies, suggesting a combination of mechanisms are responsible for detection of higher grade cancers in the future. This suggests there may be benefits to more frequent biopsies following 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 due to 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 to patients receiving radical prostatectomy immediately following diagnosis have shown that low-risk men who receive annual biopsy on AS do not have worse surgery outcomes (4). Additionally, Klotz et al. reported that patients undergoing AS with biopsies every 3 to 4 years had mortality rates consistent with patients who received initial definitive treatment (17). Assuming a uniform distribution of progression times during the 3 to 4 years intervals would suggest 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 (compared to an annual biopsy strategy) would result in an incremental increase in time to detection of grade progression that is well below these estimates.

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

Our study has some notable limitations. First, our results apply to patients with "favorable risk" prostate cancer (i.e., clinical stage ≤T1c, PSA density ≤0.15, Gleason score ≤6, total positive cores ≤2, and single core positivity ≤50%) and older men with low-risk disease (i.e., clinical stage ≤T2a, PSA <10, and Gleason score ≤6), since these are the patients that were enrolled in the Johns Hopkins AS study and thus there is a need to validate our findings on other active surveillance studies; however, this initial study lays the groundwork for such future validation work. Our results provide the trade-off between number of biopsies and mean delay time to 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 significant clinical impact. Metastasis is a better endpoint, but the data needed to fit a hidden Markov model with this endpoint do not yet exist.

Although compliance to annual biopsies in this cohort was high, making it an ideal source for the hidden Markov model analysis we presented, it was not perfect. Moreover, there was patient dropout for reasons other than biopsy confirmed progression. Thus, there is the potential for selection bias to influence the results. While this cannot be eliminated, we mitigated this risk in part by conducting simulation analyses that showed the hidden Markov model 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 this is a strong assumption, because PSA kinetics are not used to trigger intervention in the Johns Hopkins study. These limitations notwithstanding, we believe this study provides important evidence about the trade-off between varying active surveillance strategies and the optimal timing of biopsies during active surveillance.

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

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(16):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(11):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(6):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(16):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. AUA White Paper. 2012.
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(3):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(2):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(2):286-292.
9. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989;77(2):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):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(6):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(5):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(12):2664-2670.
14. van den Bergh RC, Roemeling S, Roobol MJ, Roobol W, Schröder FH, Bangma CH. Prospective validation of active surveillance in prostate cancer: the PRIAS study. Eur Urol. 2007;52(6):1560-1563.
15. Lawrentschuk N, Klotz L. Active surveillance for low-risk prostate cancer: an update. Nat Rev Urol. 2011;8(6):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(6):930-9.

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 Dec 15;33(3):272-7.

18. Patel, H. D., Tosoian, J. J., Carter, H. B., & Epstein, J. I. (2017). Adverse Pathologic Findings for Men Electing Immediate Radical Prostatectomy: Defining a Favorable Intermediate-Risk Group. JAMA oncology.

19. Liu, J., Womble, P.R., Merdan, S., Miller, D.C., Montie, J.E., Denton, B.T., and Michigan Urological Surgery Improvement Collaborative. "Factors influencing selection of active surveillance for localized prostate cancer." *Urology* 86, no. 5 (2015): 901-905.

Figure Legends

Figure 1: Simulation results for non-dominated AS strategies from 2^{10} simulated strategies and published strategies based on the estimated HMM parameters. Incremental time to detection and the reduction in biopsies are relative to an annual biopsy strategy. (Note: mean time to detection of grade progression for annual biopsy plan = 14.1 months)

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

Table 1: Patient characteristics at time of diagnosis.

Characteristic	AS cohort (N = 1499)
Age at diagnosis, yr (%)	
≤ 49	18 (1.2)
50–59	208 (13.9)
60–69	911 (60.8)
70–79	352 (23.5)
≥ 80	10 (0.7)
Race, no. (%)	
White	1314 (87.7)
Black	115 (7.7)
Other	60 (4.0)
NA	10 (0.7)
PSA at diagnosis, ng/mL	
0–2.5	162 (10.8)
2.5–4	249 (16.6)
4–6	558 (37.2)
6–10	322 (21.5)
> 10	85 (5.7)
NA	123(8.2)
PSA density at diagnosis	
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)

Table 2: Biopsy characteristics at diagnosis and surveillance biopsies.

Characteristics	Biopsy							
	Diagnosis	**First**	**Second**	**Third**	**Fourth**	**Fifth**	**Sixth**	**Seventh**
Patients, *n*	1493	1370	922	644	447	298	187	122
Age at biopsy, yr, mean (SD)	66 (6.0)	67 (6.1)	67 (6.0)	68 (5.5)	68 (5.3)	69 (5.1)	70 (5.1)	71 (4.3)
Months since last biopsy, no, mean (SD)	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, ng/mL, mean (SD)	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)
Percentage of cores positive for cancer, *n* (%)								
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, *n* (%)								
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, *n* (%)								
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)

Table 3: The change in the risk of delay of detection of higher grade disease for >1, >2, and >3 years relative to an annual biopsy strategy.

Policy Label	Number of biopsies in 10 years	Increased risk of >12 months to detect grade progression* (%)	Increased risk of >24 months to detect grade progression* (%)	Increased risk of >36 months to detect grade progression* (%)
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

*relative to annual biopsy strategy

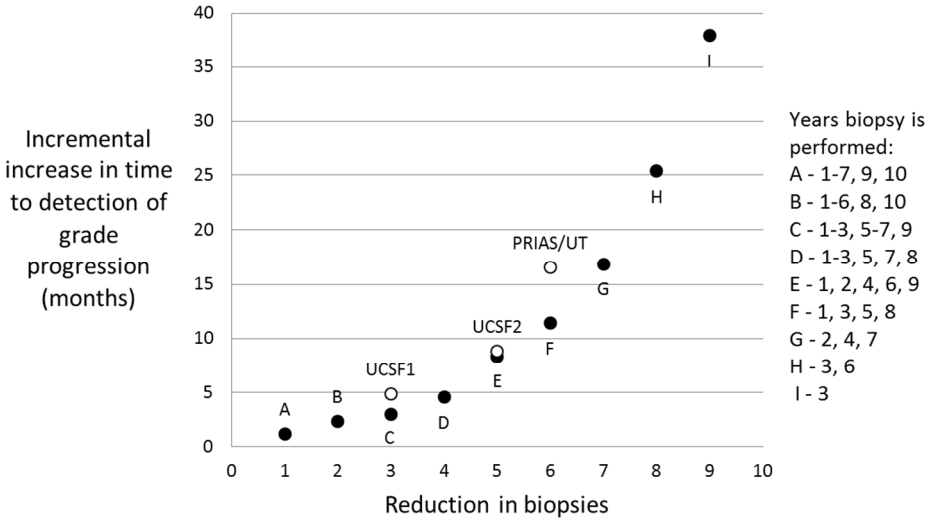
Note: annual biopsy strategy resulted in 13.5% risk of >12 months, 4.9% risk of >24 months, and 1.8% risk of >36 months to detect grade progression.

Table 4: Results comparing hidden Markov model (HMM) parameter estimates from the Baum-Welch algorithm to the true model parameter estimates from a known model. The 95% confidence interval is based on 1000 sets of simulated biopsy data with 1375 sequences per set. GS = Gleason score.

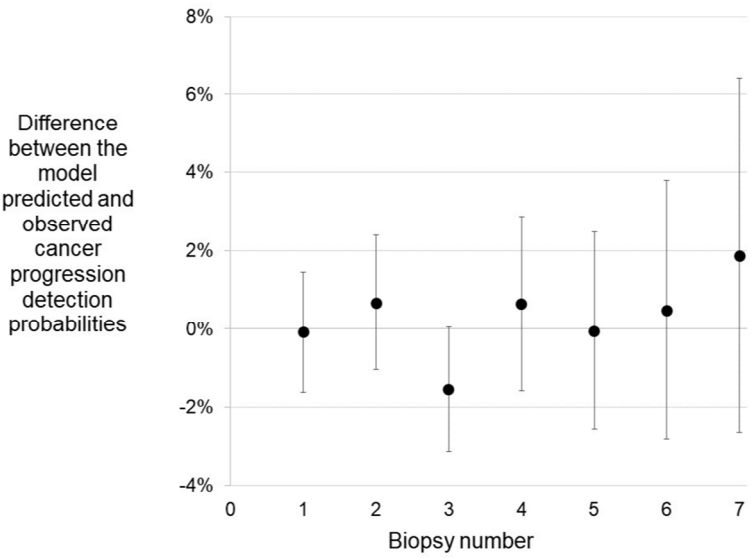
Model Parameter	True Value	HMM 95% confidence interval estimate
Proportion of patients with GS ≤ 6 at diagnosis	0.866	0.846 – 0.889
Annual progression rate	0.040	0.033 – 0.039
Sensitivity of biopsy to GS ≥ 7 prostate cancer	0.610	0.582 – 0.688
Specificity of biopsy to GS ≥ 7 prostate cancer	0.986	0.980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

Table 5: Bootstrapping results based on 30 different bootstrap samples of 1375 patients. GS = Gleason score.

Model Parameter	95% confidence interval
Proportion of patients with GS ≤ 6 at diagnosis	0.857 – 0.876
Annual progression rate	0.032 – 0.040
Sensitivity of biopsy to GS ≥ 7 prostate cancer	0.587 – 0.635
Specificity of biopsy to GS ≥ 7 prostate cancer	0.980 – 0.988



108x65mm (300 x 300 DPI)



103x80mm (300 x 300 DPI)