Gmail **Ani Rudh <anirudhtomer@gmail.com>**

Major revision - BJU-2020-0444

BJUI Office <onbehalfof@manuscriptcentral.com> 27 April 2020 at 10:36
Reply-To: editorial.office@bjui.info
To: anirudhtomer@gmail.com
Cc: Freddie.Hamdy@bjui.info

Dear Dr. Tomer

Re: Personalized Biopsy Schedules Based on Risk of Gleason Upgrading for Low-Risk Prostate Cancer Active Surveillance Patients

Thank you for sending your Manuscript to the BJU International, which we read with great interest. We are keen to consider the paper further, but do not feel it is suitable for publication in its current version. The reviewers raised a number of issues and we would like to invite you to revise the manuscript, taking into account the reviewers' comments and suggestions.

We would be pleased to look at a revised version. Please address the reviewer comments in your response, and indicate to the Editor all the changes made in the manuscript. We hope you will find that the constructive reviews helpful in revising your paper.

The current Author Guidelines are attached for reference if you have been asked to modify your article or abstract to comply with these. Please also review the attached Search Engine Optimisation tips to improve your article's discoverability.

We look forward to receiving your revised manuscript within the next 45 days. Please submit online through your author centre, and note that revisions do not guarantee subsequent acceptance for publication.

Given the current situation with COVID-19, we understand that your priorities may lie elsewhere. Please email editorial.office@bjui.info if you require an extension.

I look forward to hearing from you in due course and thank you again for sharing your work with the BJUI.

Yours sincerely,

Freddie Hamdy
BJUI Editor in Chief Elect

Section Editor
Comments to the Author:
(There are no comments.)

Referee: 1

This is a study to develop risk based biopsy schedules in AS, using longitudinal PSA measures, biopsy data and baseline age. The data were used to estimate timing to Gleason upgrading & a threshold risk to inform re-biopsy scheduling during AS follow-up. There was a development set (PRIAS, n=7813 with Gleason grade group 1 at baseline, of whom 1134 experienced Gleason upgrading), an independent validation set (GAP3 database of six cohorts, >20,000 men worldwide), & an online tool was developed to make the model available to clinicians for shared decision making. Instantaneous PSA velocity was a stronger predictor than baseline PSA. This work is quite novel in that it is exploring the use of risk prediction to create a biopsy schedule for each man on AS, to reduce unnecessary biopsies for those men where upgrading is unlikely but to allow shorter intervals to re-biopsy where upgrading risk is above a threshold. Loss to follow-up in PRIAS was only 4%. Statistical analysis was sophisticated and appropriate. The web application is very useful clinically as it gives patient-specific estimates for both personalized and fixed schedules. As there are no validated protocols for biopsy rescheduling, and the quality of this study in terms of power, follow-up, analysis & validation is high, this is is undoubtedly an important paper. My comments are:

1. A limitation that the authors recognize is that MRI is not included in the model. They argue that the model is potentially extendable to include MRI in the future. Thus it is still very useful to publish the current model they have developed so that it can be extended in cohorts where more comprehensive MRI data are available.

2. It will be important to know if this risk-based scheduling strategy will improve cancer-related outcomes for men in terms of metastases and mortality. Are there any data on these outcomes that could be investigated to show the ultimate clinical value of risk-based re-scheduling?
3. The model ROC & prediction error were moderate, so the value of the model to inform clinical practice is uncertain because false positives & negatives are highly likely. In particular, could the authors comment on how confident clinicians can be that they will not cause harm (eg by delaying biopsy in men wrongly classed as low risk) when using this risk-based approach? A simulation of how may biopsies could be wrongly delayed might be useful. Can the authors comment on whether clinicians should wait for more model development before using the online tool to make definitive decisions?

Referee: 2

This a very ambitious study, overall very well written and structured. The M&M section is appropriately detailed and comprehensive to such extent level that clinicians are not necessary able to follow. Here, it is much appreciated that the authors included intuitive figures for easier, clinical-focused understanding and including the underlying statistical coding that enables biostatistician review and reproducibility.
Moreover, a web-application is also highly appreciated by treating physicians and/or patients.

#1 With regard to specific developments such as MRI and the important major reclassification rate at confirmatory biopsy at one year (p7, “However, we recommend…), readers might be irritated that respective “anticipatory” mentioning of these aspects are within the M&M sections (p5 and p7, respectively). Accordingly, mpMRI should be relocated into the discussion section.
#2 However, upgrading rates at 1 year should be shortly specified even before p7 and maybe it is better the modify it into the true timepoint zero of this paper, since many segments, e.g. figure 4, heavily suggests so and it is already handled in such way within the study. In consequence, an according modification of the whole manuscript is recommended in this regard to improve intuitive reading and avoid any potential misunderstanding.
#3There is another potentially added benefit of using the confirmatory biopsy as timepoint zero (which if applied, should be discussed), namely a homogenization or convergence effect of those cohorts with regard to comparability. Specifically, external validation relies on cohorts with different AS entry criteria sets that have different stringency and result in different AS eligibility rates and likely in different upstaging and upgrading outcomes (either at confirmatory biopsy with regard to upgrading, and both, upgrading/upstaging if RP treatment was performed, for example).
#4In general, the effect of different AS entry definitions (that were likely applied to the respective GAP3 cohorts) on the external validation within this study and the effect on application of the proposed personalized schedules should be discussed. Alternatively, authors can consider to sub-select GAP3 patients that fulfill PRIAS entry criteria (if such can be derived). If this was already performed, this should be explicitly mentioned.
Please consider to shortly discuss following aspects, too:
#5First, the model relies on repeated PSA measurements. As such, a personalized schedule should also be accompanied by personalized schedule recommendations with regard to PSA measurements, in order to enable a minimum longitudinal granularity for model application. It is assumed that PSA measurement equals the examined time point within the model. If this is the case, it should be spelled out that PSA measurement must be performed at least at proposed intervals/visits. Conversely, what effect would interim PSA measurement (i.e. between scheduled visits) have on personalized schedules?
#6Second, additional information such as prostate volume (p5, l20) would enable inclusion of PSAD. This is indeed an important limitation since several AS entry criteria rely on that variable and several series have shown that PSAD contribute more information to upgrade/upstage prediction within AS eligible patients than PSA. Accordingly, this should be mentioned in the limitations section.

#7Minor comments:
Introduction: P 6: Please don’t use future time in the final segment, since it might be understood that future, follow-up papers `will address these subjects. In fact, a validation was performed in the study at hand.
P5, Line 31: Please relocate the sentence “However, the model we propose… into discussion section.”
P6, L15: please elaborate “within-patient correlation”
For segment M&M 2.1 and 2.2. please shortly elaborate examples of treatment and consider showing these in supplemental material.
Figure 4: There are no dashed lines as suggested in the figure description.

Referee: 3

In this study, the authors develop a model (Bayesian joint model) to predict risk of upgrading on AS from a large AS series (PRIAS study), to inform the development of personalized AS protocols for repeat biopsy, knowledge which would help both physicians and patients. The model is then validated in the largest cohorts included in the GAP3 database. The statistical methods seem rigorous and the Supplementary appendix is impressive. A lot of work has gone into this paper. The online application has nice graphics, is user-friendly and allows for personalization.

MAJOR
1. Biopsy intervals are created based on “using the model predicted upgrading-risks, we scheduled biopsies whenever a patient's upgrading-risk was above a certain threshold”. However, timing of detection of progression depends on at what intervals biopsy was done in the included cohorts. Furthermore, the cohorts differed in terms of inclusion criteria for AS, as well as the biopsy techniques and biopsy schemes that were used in the cohorts for detecting cancer. In the app, for Patient 1, the risk of upgrading between Aug 2010 and Aug 2014 varies, for example: 24% (Hopkins), 37% (MSK) and 40% (Toronto). Which cohort should the clinician choose when guiding decision making in clinical practice? Does this require that clinicians know more details and nuances about the cohorts to use this app?
2. PSAV was found to be a strong predictor of upgrading, however, PSAV is used to trigger further investigation with e.g., MRI and/or biopsy, at which upgrading is detected. Is this circular reasoning problematic in terms of using PSAV as a predictor of an outcome, which is contingent upon the predictor used for decision making?
3. The time-dependent AUC was moderate (0.6-0.7) in the validation cohorts. The authors acknowledge that this statistic compares better to AUCs from other models eg logistic regression. However, is the accuracy sufficient to make the model ready for clinical use for decision-making regarding timing of biopsy in clinical practice? The authors also note that calibration was not optimal in the validation cohorts.
4. The authors write “Our model was mis-calibrated for validation cohorts (Panel B, Figure 4). Recalibrating the baseline hazard of upgrading in validation cohorts resolved this issue.” What is the reason for this? Eligibility criteria for AS? (eg very low risk vs low risk) Additional testing and/or MRI and/or confirmatory biopsy at baseline in some cohorts but not others? 5. The online application allows choosing the time to predict cumulative risk of upgrading, for example for Patient 1 between Aug 2010 and Aug 2014. The cumulative risk of reclassification increases with the duration of follow-up. How should a clinician use this information to make biopsy recommendations for patients? Should one convert the date scale in one’s head and make inference about the next biopsy date? Or upload one’s own patient data in an Excel sheet?
6. How was the risk threshold of 9.5% chosen? The app says “The risk threshold with the least euclidean distance is chosen as the automatic risk threshold.” What is meant by this? Furthermore, how do clinicians and patient interpret this risk? What does this risk mean in terms of net benefit (reduced unnecessary biopsy and less harm from biopsy) vs net risk (missed/delayed progression)? How many biopsies does a physician need to do if increasing/decreasing the risk threshold? Would be nice to include an example table so the reader can compare strategies and see what happens with each decision.
7. If physicians choose their own risk thresholds (between x%-y%) and biopsy intervals thereafter (between x-y years), what would be the effect on net benefit (reducing biopsies while keeping the timely detection of upgrading within the window of curability) as compared to if physicians followed the regular one-size-fits-all AS schedule with biopsy every x years for all men on AS? Could this be simulated in a trial perhaps, with a distribution of risk scores chosen by physicians as compared to the cohort with regular intervals? Imagine for example patients and physicians are either risk takers or risk averse, would a risk-based strategy increase or decrease overall net benefit/net harm? For example impose delay if biopsying too infrequently for perceived low risk or adding on unnecessary biopsy if biopsying too frequently for high risk?
7. “The maximum follow-up time up to which predictions can be obtained depends on each cohort”. For some cohorts, the follow-up is 4-6 years. Some cohorts biopsied every 1,2 or 3 years. How many data points were used to model and predict progression?
8. The model uses a cumulative incidence approach, ignoring treatment based on multiple positive biopsies as competing risk. Did the authors consider a competing risk approach?

MINOR
1. Are co-authors from the 7 cohorts in GAP3 included in the author list?
2. Have the included cohorts agreed to publishing risk estimates for the individual cohorts based on patient examples in the online app?
3. The text under Figure 4 talks about model miscalibration comparing solid and dashed lines, however the reader can only see solid lines in different colors.
4. Figure 5B. Why is biopsy scheduled less (rather than more) frequently for higher risk? Looks like biopsy is at year 2, 3.5, 4.5. and 6 for the patient with 5% risk of upgrading but at year 2, 4.5. and 6 for the patient with 10%. Or are these just examples?
5. Figure 5C. What is the expected time delay in detecting upgrading with the different strategies being compared to?

Additional comments from reviewers about ethics:
Referee: 1
Ethics:

Referee: 2
Ethics: Ethical standards were appropriately adhered to.

Referee: 3
Ethics: N/A

---

2 attachments

*** BJUI-Author-Guidelines-March-2019.pdf**

687K

*** BJUI-search-engine-optimisation-tips-v2.pdf**
1099K