TRANSLATIONAL ANDROLOGY AND UROLOGY

Peer Review File

Article Information: Available at http://dx.doi.org/10.21037/tau-20-1082

Review Comments:

Using the GAP3 consortium data, the authors state their aim to assess heterogeneity in disease progression between centers by follow-up protocols. They report that progression rates varied between centers, citing a hazard ratio of 2.5, and that this persisted after adjustment for clinical factors. This is a large and powerful data source, but this report is severely hindered by lack of clarity and ultimately not appearing to achieve the stated goals. As detailed below, it is not even clear what the hazard ratio of 2.5 -- the only quantitative measure provided in the abstract -- actually represents. Please see specific comments below.

1. "Disease progression" is the main outcome of interest - please define it. This presumably includes clinical and path progression, but these are not sufficiently defined. What is clinical progression? How does it differ from pathologic progression? Does pathologic progression include increased volume or only upgrading on biopsy?

1. RE: Thank you for your question. We have now added in the information we have used previously to describe this definition (Eur Urol 2019): “With respect to indicators of disease progression, we have previously (1) described that there is heterogeneity in recording across centres. We therefore used the following coding for defining signs of disease progression: ‘Convert to watchful waiting’, ‘Clinical progression’, ‘Pathological progression’, ‘Clinical and Pathological progression’, ‘PSA progression (PSA-DT < 3 years)’, ‘Other PSA kinetics’, ‘Patient choice/Anxiety’, ‘Doctors Anxiety’, ‘Radiological progression’, ‘Died’, ‘Lost to FU’, ‘Other/Unknown’ or ‘Still on active surveillance’. If the reason for discontinuation of AS was classified as ‘other/unknown’ but the ‘pathological progression status’ was ‘Gleason Grade Group 3 or higher’ or the ‘clinical progression status’ was ‘cT3 or higher’ or ‘PSA progression status’ was ‘PSA>20’, the reason for discontinuation of AS was also categorised as signs of disease progression. Thus, similar to our previous publication (1), “disease progression” can refer to risk reclassification or disease progression as such and was used as the main outcome of interest in the current project. Conversion to active treatment, watchful waiting, or death were considered as competing risks.”

2. The table states "progression - treatment (event)" -- is this the "disease progression" outcome referenced above? Use the exact same wording and be crystal clear, as these are points of confusion.

2. RE: Based on the above definition of disease progression, we have now updated our Table to clearly revlect this.

3 . "The median hazard ratio can be interpreted as the median increase in risk when comparing two similar patients from a centre with low disease progression rate and from a centre with high disease progression rate."
So the centers are being grouped by progression rate? And the HR of 2.5 references low vs. high progression rates? That would mean you have set up for there to be a notable difference in progression rates (by virtue of grouping by low vs. high), and you are then simply providing a magnitude of risk difference by sharing the HR of 2.5? Are these progression rate groupings defined in the text? And progression rate is distinct from the surveillance of intensity (high, int, low) previously described? This is tremendously confusing.

3. RE: We apologise for the confusion and would like to clarify that centres are not being grouped according to their progression rates. We rather fitted a Cox frailty model in the GAP3 database including a log-normal centre-specific frailty term. This frailty term is a random effect for centre and the heterogeneity in upgrading risk between the different centres is quantified by the variance of the random effect term. This is, however, hard to interpret and we therefore calculated the median hazard ratio to aid interpretation of results.
The median hazard ratio describes the median hazard ratio obtained when comparing hazards of the occurrence of the event of interest of in individual from a randomly selected clusters with another individual with identical covariates, but randomly selected from a different cluster.

We have rephrased the methods section as:
“To further investigate heterogeneity in baseline upgrading risk between centres we estimated Cox frailty models. These frailty models include a random effect term for centre and the difference in upgrading risk between centres is parameterized using the variance of the frailty term. To aid the interpretation of the results we transformed the variance term to the median hazard ratio (MHR)(9). The MHR can be interpreted as the median hazard ratio of disease progression when moving a patient from a randomly chosen centre to a randomly chosen centre with a higher disease progression risk (Or comparing a patient with the same characteristics between two randomly chosen centres). We calculated the MHR using a frailty model not including any patient and tumour characteristics and investigated whether differences in included patients could partially explain the observed heterogeneity by adjusting for characteristics included in the previous analyses.”

One of our hypotheses was that the observed heterogeneity between different centres in upgrading risk could partly be explained by surveillance intensity (e.g. higher risk of upgrading due to more intense surveillance). We then refined our analysis and added intensity as a covariate to the model with frailty term to investigate of the between-centre differences would decrease.

4. Tables 1-3

It appears "progression - treatment" is a composite of the reasons listed below, not including WW/convert/death (competing) or other/unknown. This should be clear from the format of the table. That % can be shown as the sum of the others and then the outcomes not included in that composite variable should be listed so it is clear they are not contributing to that sum. Why is other/unknown not included in that composite variable? Also please add median(IQR) follow up time to the table.
4. RE: Thank you for your suggestions. Table 1 has been amended accordingly.

5. "In each group, there was considerable variability in the disease progression rate (median HR 2.5, Figure 2)" - this seems to be a central point of the paper but to the reader it is not even clear what this means. As asked above, what are the "groups"?
5. RE: Groups refer to centres with similar active surveillance follow-up protocols. We have clarified this as follows in the methods section: For the current analyses, we have grouped centres according to similarities in follow-up protocols (high, intermediate, and low) – the term “groups” in the results section hence refers to these different AS follow-up protocols. Please also see our response above where we have clarified the interpretation of the median HR.

6. Studies were grouped by progression rates? Or does this relate to the variability in disease progression based on intensity of surveillance - low vs. intermediate vs. high? HR would mean 2.5 times increased risk of the outcome during overall follow up? If so, the follow-up of each study and within each group needs to be clearly shared. That would mean the HR was linearized from low to intermediate and intermediate to high (assuming linearity across groups). Instead, would provide pairwise HRs - specific hazards between low and intermediate, intermediate and high, low and high.
6. RE: We apologies for the confusion, but have tried clarifying this better in our response above.

7. How does Figure 2 itself convey the point being made that there is considerable heterogeneity? In the adjusted plot, to me it looks like basically the vast majority of log hazards are between -0.2 and 2, with a few outliers below -1.0.

7. RE: The figure shows the difference in upgrading risk from the average centre in the GAP3 database (value of zero). A log-hazard value of 1 would mean a 2.7 fold increase in upgrading risk compared to the average centre in the GAP3 cohort. This is a substantially higher effect compared to the covariates included in the prediction model, indicating that differences in upgrading risks between patients are unexplained differences between centres rather than differences in patients and tumor characteristics.

8. Why are the publications from these various cohorts not referenced?
8. RE: Please note these cohorts are referenced as part of the cohort profile of GAP3, which we have published previously (Bruinsma et al., BJUI 2017).

9. Please be clear - were primary data provided by each institution for this analysis?
9. RE: Yes and this is also clarified in the first paragraph of the methods section: “By combining patient data from established AS cohorts worldwide, the global GAP3 database was created between 2014 and 2016. In addition to ethical approval for sharing digital patient data in a centralized global database, inclusion criteria required an active registry of AS patients over the last two years or more, including at least ~50 patients annually.”

10. In Discussion: "However, the relative effects of patient characteristics on risk of disease progression were fairly stable between centres." What data in the paper support this statement? I do not recall seeing hazard ratios of patient characteristics.
10. RE: Thank you for pointing this out. We agree that this was not a clear statement and have replaced it bey the following: “When combining worldwide cohorts of men on AS, we noted unexplained differences in disease progression rates between centres (median HR: 2.5). After adjustment for various clinical factors (age, year of diagnosis, Gleason grade group, number of positive cores and PSA), substantial heterogeneity in disease progression remained between centres.”

11. The paper concludes there is a need for local adjustments for differences in risk of progression. I was confused by the intent of this analysis. To me, the most useful question would have been whether or not risk of progression was associated with frequency of biopsy (low, intermediate, high). One would expect that looking more often results in finding more often, and the authors could then provide useful information as to whether i) that expectation holds true, and ii) how clinical factors impacted progression rates - i.e. were clinical factors also predictive of progression, or was detecting progression mostly dependent on how often biopsy was performed? Again, this may have been part of the analysis but the text is too unclear to follow (I do not follow what exactly the HR of 2.5 is referring to).
11. RE: Please see our response above where we aimed to explain the analyses, aims,

and interpretation more clearly.

Ultimately there are a lot of great data here. It is difficult to know whether parts of the analysis are reasonable because the text lacks clarity in what exactly is being presented. Please at least re-write the text to more clearly explain the analysis, and strongly consider performing the more informative analysis suggested above.

Minor:

1. No need to state age>18 in my opinion
1. RE: We agree that this is self-explanatory, but it has been a criteria clearly identified by the centres and hence it is included for completeness.

2. 4-4: would say "traditionally" performed bx on annual basis, acknowledging that this is no longer the case in most if not all centers.
2. RE: The data reflects the practice in the centres at the time of data collection and hence we cannot make statements about current practice.

3. Correct "Pargin" tables –
3. RE: Thank you for spotting this error. We have now corrected this.

4. Table 3 text states High Group but should be low.
4. RE: Thank you. We have corrected this error.