BMJ - Decision on
Manuscript ID
BMJ.2017.038395

Body: Dear Dr. Seibert

Manuscript ID BMJ.2017.038395 entitled "A genetic risk score to guide age-specific, personalized prostate cancer screening"

Thank you for sending us your paper." We sent it for external peer review and discussed it at our manuscript committee meeting. We recognise its potential importance and relevance to general medical readers, but I am afraid that we have not yet been able to reach a final decision on it because several important aspects of the work still need clarifying.

We hope very much that you will be willing and able to revise your paper as explained below in the report from the manuscript meeting, so that we will be in a better position to understand your study and decide whether the BMJ is the right journal for it. We are looking forward to reading the revised version and, we hope, reaching a decision.

Yours sincerely,

Rubin Minhas
Dr Rubin Minhas
BMJ Associate Editor
rm1000@live.com

*** PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. ***

https://mc.manuscriptcentral.com/bmj?URL_MASK=9c20aebaedc44a2fa6e3c04a05c300f0

Report from The BMJ's manuscript committee meeting

These comments are an attempt to summarise the discussions at the manuscript meeting. They are not an exact transcript.

ATTENDEES: John Fletcher (chair), Elizabeth Loder, Georg Roggla, Tiago Villaneuva, Rubin Minhas, Daoxin Yin, Rafael Perrera (statistician).

Potentially interesting, but the reviewers raise some fairly weighty issues. The paper will need a formal stats review and a conditional publication consideration subject to satisfactory responses from all reviewers prior to completion. It is recognised the authors have a significant undertaking to accept. A major concern here is the positive slant towards the subject and it remains to be seen whether the research group is able to produce a more objective and dispassionate account. The editorial committee will be guided by the complete satisfaction of the reviewers when the paper is re-reviewed.

In your response please provide, point by point, your replies to the comments made by the reviewers and the editors, explaining how you have dealt with them in the paper.

** Comments from the external peer reviewers**

Reviewer: 1

Recommendation:

Comments:
The manuscript entitled, "A genetic risk score to guide age-specific, personalized prostate cancer screening." The study presents data from a genetic study conducted in over 31,700 men and a separate validation dataset comprised over 6,400 men. The results of this manuscript are very impressive and the data provides novel insights into how genetic testing can influence the timing of PSA screening. In addition, it presents novel prostate cancer risk-associated SNPs that are associated with an earlier age of diagnosis for prostate cancer.
Overall, this manuscript is well written and provides new information. However, there are several concerns/clarifications that should be addressed prior to final acceptance of this manuscript:

The authors should include separate analyses on aggressive prostate cancer (defined as either Gleason >/=8, >/=pT3, node positive, biochemical recurrence or death). This will help clarify whether these SNPs predispose to an early age diagnosis of aggressive or indolent prostate cancer? Without such information, targeting men with higher inherited risk based on these SNPs may exacerbate over-screening, over-biopsy, and over-treatment of prostate cancer, major challenges of prostate cancer care we are facing now.

In addition, it is unclear whether the prostate cancer SNPs presented are novel or overlap with the ones that have previously been described (e.g. the >100 SNPs implicated in traditional GWAS studies). How many of these SNPs are the same? Which method performs better, their new methods/ or the polygenic risk score based on 100 established risk-associated SNPs that the authors have previously described?

Additional Questions:
Please enter your name: Brian Helfand

Job Title: Associate Professor

Institution: NorthShore University HealthSystem

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here: NONE

Reviewer: 2

Recommendation:

Comments:
BMJ 2017 038395
This is a highly impressive project, a large, collaborative venture addressing an important question.
My enthusiasm for this paper is weakened by what I feel is the use of unconventional metrics and a rather slanted presentation. Let me first state for the record my view that germline data most likely do have a role in prostate cancer screening. Indeed, I am co-PI of a large NIH grant examining exactly this issue. Hence my criticisms of this paper are not criticisms of the concept per se, but an attempt to better identify the role of germline genetics in prostate cancer screening.
Major comments
1. The metrics used in the paper are unusual and, in my view, give an unreasonably positive spin on the findings. The primary result is expressed as a shift in the incidence curve between the 1st and 99th centile. First off, most authors compare, say, the top and bottom quintile, or the top 10% vs. patients below the median; comparing 1st and 99th centile gives an exaggerated impression of effect size. Second, talking about a shift in the curve seems appropriate for an endpoint that everyone gets (e.g. death) but not for an event experienced by only a minority (such as getting cancer). Conventional metrics I have seen in papers, and used in papers of my own include difference in absolute risk (e.g. 4% of aggressive cancer by age 85 in patients with scores below median vs. risk of 22% in patients in the 10% of scores); C index or AUC (e.g. genetic risk score predicted aggressive cancer with a concordance index of 0.73); Lorenz curve (e.g. 40% of aggressive prostate cancer by age 75 occur in men in the top 10% of genetic risk).
2. A particular problem with using a comparison between 1st and 99th centile, is that the estimation at the low end of risk is highly problematic: in brief, it doesn't pass the smell test. In table 1, the authors say that delta age, their metric of risk, goes from -3 at the 20th centile, to -11 at the 99th centile to -35 at the 0.1st centile. In most epidemiological settings, risk at the low end is fairly flat: the risk of lung cancer in someone who has a 3 year pack history is pretty much identical to the risk of someone with a 1 year pack history of smoking or who never smoked; similarly, the risk of prostate cancer for someone with a PSA of 0.5 or 1 isn't much different. I'm struggling to think of any example in epidemiology where there are dramatic differences in risk at the low end. In this particular case, it is very hard to think through biologically what is happening. How exactly does the germline differ between patients at the 20th, 1st and 0.1st centile of risk to cause such dramatic shifts in risk? is it, say, that patients at the 1st centile are positive for 10 risk alleles and those at the 0.1st centile are positive for only 5, and that those 5 additional alleles make a massive difference to risk?
3. Another example of the rather biased presentation involving an inappropriate metric is the statement in the discussion that "PHS was a remarkably strong predictor of age at diagnosis … p=10E-53)". A p value gives the strength of evidence against the null hypothesis, not the strength of a predictor, the basic difference between clinical and statistical significance.
4. The use of SEER data to help calculate absolute risk is rather problematic. SEER data are in the context of a rather heavily screened population. It is not at all

clear that a patient's risk of cancer or aggressive cancer given a particular genetic background corresponds to a relevant multiplication of the risk from SEER.

5. It has been known for close to 25 years that PSA is an incredibly strong predictor of the long-term risk of aggressive prostate cancer. Indeed, a series of papers in the BMJ reported statistics such as an AUC of 0.8 – 0.9 for 25 year risk of metastasis, or that 40% of deaths occur in the top 10% of PSAs. These studies were recently replicated in a US population in J Clin Oncol. 2016 Aug 10;34(23):2705-11. The problem here is that none of these data were cited at all let alone introduced into the analyses. Some obvious questions might be: of the few patients with low PSA who are subsequently diagnosed with aggressive prostate cancer, does genetic risk predict outcome? Can genetic risk predict which of the patients with a higher PSA eventually get aggressive prostate cancer?

6. PSA only appears to come into the picture using the ProtecT data for biopsy outcome. However, the authors' approach is somewhat problematic because it treats PSA as a binary variable (i.e. what is the PPV for positive biopsy amongst men with a positive PSA test). The problem here is that what we really want to know the marginal value of genetics over and above PSA. For instance, if everyone with a certain mutation were at very high risk for cancer and had PSA in the 25 – 50 range, then the PPV of genetics would look fantastic. But there would be little if any clinical value because if a patient had a PSA of 25, you'd biopsy them immediately and not worry about genetics. The standard method for determine whether a marker such as PHS is of value is to look at the AUC of standard markers (say, age and PSA in men with a PSA>3 ng/ ml) and then the AUC of these markers plus the new marker (e.g. age, PSA and PHS). I should note that the PPV data is quite out of keeping with other studies in the literature, which show relatively minor improvement in discrimination with germline data.

Minor comments

1. Please avoid reporting p values such as 10E-53, which is roughly the probability of correctly identifying a randomly selected atom in the entire Milky Way galaxy. Vanishingly small p values are meaningless and spuriously precise.
2. It is unclear why results for any cancer and aggressive cancer are given equal prominence. We want to detect the latter not the former, indeed, it is somewhat of a problem to detect non-aggressive prostate cancer. I strongly suggest that the paper focuses on aggressive cancer with discussions of overall cancer restricted to supplementary material.
3. Figure 1 is extremely unusual, indeed, it could be misleading to a reader who didn't attend to the fine print. Why not use a more conventional methods and plots to test proportional hazards?

Additional Questions:
Please enter your name: Andrew Vickers

Job Title: Attending

Institution: MSKCC

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here: not sure whether this is really competing at all, but I earn royalties for the sale of a test given to men with high PSA to determine the need for biopsy.

Reviewer: 3

Recommendation:

Comments:
The authors report a novel SNP-based signature to help guide decision-making regarding prostate cancer screening, based on a large, multi-cohort analysis. The sample size is very robust and the use of ProtecT as a validation cohort adds strength. The goal is laudable, though the use of such a score does not seem practical for contemporary primary care practice today. There are some other methodological concerns related to the PRACTICAL cohorts which should be addressed to strengthen the manuscript.

Major concerns:

1. More details are needed re: the PRACTICAL consortium. Clearly the collaboration has allowed a large number of samples to be gathered, but table S1 needs to provide much more detail than numbers of cases and control. Review of previous PRACTICAL publications suggests that some of the cohorts have come and gone, and there may be multiple sources of bias. Some of these cohorts are population-based, some are based on treated cohorts, and the secular rates of screening in the relevant populations are quite divergent both over region and over time. Given how dependent being a prostate “case” is on background screening history, these seem like major potential problems. Trials like the PCPT, in which all men were biopsied after a fixed period of followup, provide clarity, but many other designs will not.
2. Lead time is also a big consideration. How old are the men in the PRACTICAL cohorts, and how long were they followed? It’s one thing for a genetic screen to inform a man at a fixed age to defer testing; it’s quite another to advise that man that he will never need to worry about prostate cancer.
3. This is not the first polygenic risk score to be proposed. How does it compare to previous SNP-based scores such as the PRS (Pashayan et al BJC 2015), the GRS (Oh et al Oncotarget 2017), not to mention the STHML3 model which some of the study authors published? How many SNPs overlap between PHS and these signatures?
4. The PHS signature should also be compared explicitly to early _baseline_ PSA, which is much more accurate and can also guide decision making many years into the future (e.g., Vickers BMJ 2013, Preston JCO 2016).

5. Given variations in PRACTICAL studies (see point #1 above), is Cox proportional hazards analysis legitimate? What tests of proportionality were performed across the cohorts?
6. SEER data from the 1990s and 2000s reflect a heavily, and variably, screened population. This population is quite different from ProtecT and most PRACTICAL cohorts.

Minor concerns:

1. Ref 2 is substantially out of date. The update (Lancet 2014) showed a 21% to 27% mortality reduction with screening.
2. Consideration of family history as a binary variable is a bit inadequate. There's a big difference between a brother diagnosed with low-risk disease at age 80 and three close relatives diagnosed at young ages with lethal disease.
3. This tool will not apply for men with non-European ancestry—who often face substantially higher rates of both prostate cancer and potentially lethal disease.
4. PHS clearly offers some potential advantages over PSA alone as a screening tool. But given the ongoing controversies about screening in primary care (in the US, for example, the USPSTF just switched its recommendation again from a "D" to a "C"), it seems a stretch that primary care providers will be able to implement germline testing before even tackling the prostate cancer screening decision. A simple (and cheap) PSA test can essentially rule out prostate cancer for more the half the population tested. So perhaps the better place for a test like this is to help guide decision-making regarding biopsy among those with a borderline elevated PSA.
5. Citing refs 19-22 seems superfluous—SEER data are downloadable or analyzable via SEER/Stat. Moreover, the refs are ACS publications, and reflect other data sources in addition to SEER.

Additional Questions:
Please enter your name: Matthew Cooperberg

Job Title: Associate Professor, Depts of Urology and Epidemiology & Biostatistics

Institution: UCSF Helen Diller Family Comprehensive Cancer Center

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-check

lists/declaration-competing-interests'target='_new'> (please see BMJ policy)
please declare them here:

Reviewer: 4

Recommendation:

Comments:
Not the easiest paper to read, particularly in the Methods – the authors need to improve the readability of what they are attempting to do in the Methods. Some of the Methods are tucked away in the supplementary material and these should be moved and described more clearly in the main body of the paper. Consulting the GRIPS Statement (Janssens et al, PLoS Med 2011) and how to report genetic risk scores should be considered to help improve the clarity and reporting of the manuscript.

Avoid spin terms such as 'discovery data set' – This is the development dataset. The description of the PRACTICAL consortium is brief (uninformative). Much more information is required. Supplementary Table S1 is brief! If this were a prediction model using non-genetic information from multiple data sources, we would expect much more information to be presented, basic characteristics, dates, information on the controls etc…

I have some queries regarding how the authors have evaluated the predictive performance of the PHS (the approaches the authors use and present are non-standard). Typically this is assessed by calculating measures such as calibration and discrimination. (See the GRIPS or TRIPOD statements for reporting such studies). I don't fully understand what the authors are bombarding the reader with – they are complex and difficult to disentangle what is going on. What we want to know is whether the risk score predicts sufficiently accurately in the validation data, and whether family history alone or added to the model improves prediction, and I'm struggling to pull out this information.

I'm not convinced that regressing the PHS as a single variable in a Cox model and focussing on the resulting p-value will reveal whether the PHS accurately predicts age of PCa onset. The coefficient from fitting this model is the calibration slope.

There is no mention of data quality – completeness of data, missing data.

It's not clear to me why the validation data set is being used to compare family history, why aren't models with and without family history using the development data set that are subsequently tested on the validation data set.

I don't think I fully understand what they are doing when 'examining the impact genetic risk on PSA screening'.

The reporting of the p-values as Xe-XX should be avoided. The very small P-values are reported with too much precision.

The discussion of cost-effectiveness is over-hyping things somewhat - nothing by hyperbole.

There is generally quite a bit of enthusiastic interpretation of the findings – beyond what the results show. The conclusions in the Abstract being one example.

Additional Questions:
Please enter your name: Gary Collins

Job Title: Professor of Medical Statistics

Institution: University of Oxford

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here: