## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to the BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Infrequent and Incomplete Registration of Test Accuracy Studies: Analysis of Recent Study Reports |
| **AUTHORS** | Korevaar, Daniël; Bossuyt, Patrick; Hooft, Lotty |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | Trikalinos, Thomas<br>Tufts University |
| **REVIEW RETURNED** | 17-Oct-2013 |

| | |
|---|---|
| **GENERAL COMMENTS** | We read the manuscript by Korevaar and colleagues with great interest. The authors motivate the paper as follows: It is very likely that publication bias and selective reporting bias are major problems in test accuracy studies. Unlike clinical trials, prospective registration for these studies is not required. The authors assert that prospective registration of test accuracy studies is necessary, and conduct empirical research to estimate the proportion of registered accuracy studies in a 2-month sample of recent reports published in journals with impact factor of 5 or more. This clearly written manuscript addresses an interesting issue. We have one comment, and several observations of much smaller import.<br><br>Comment: The authors' (apparent) position is that all diagnostic accuracy studies should be registered prospectively. We are very sympathetic to a mandate for prospective registration, but not as resolved that it should apply universally. Indulge us in this brief digression:<br>For RCTs (and trials in general), the research community largely agrees that the arguments favor mandatory prospective registration. Benefits of registration include reducing the impact of publication/reporting bias, promoting transparency, and fulfilling an ethical obligation to study participants (primarily) and future patients (secondarily). The main counterargument relates to the inconvenience of the added bureaucracy. For RCTs we have extensive data that publication/reporting biases are very much a problem. We have almost universal buy-in from the industry, academe, governments and other stakeholders that the RCT represents a large enough investment to justify the added costs of registration. The ethical obligations to the participants seem (and probably are) crisp: Those who consented to play dice with their (well-)being must be respected. Overall, demanding universal registration makes sense. Fortunately, prospective registration is also practical. |

| | |
|---|---|
| | Studies of diagnostic accuracy are a mixed bag when it comes to designs and purpose, and this can tip the balance of the pros and cons for prospective registration, or even render prospective registration impractical. It may help to consider three groups of studies:<br>-- In one extreme we have protocol-driven test accuracy studies that collect data de novo and primarily for assessing accuracy (some can even be RCTs). We submit that the aforementioned arguments probably transfer, and we favor prospective registration for these studies.<br>-- In the other extreme we have exploratory ('opportunistic', without a protocol or a hypothesis) analyses of existing datasets (e.g., a predictive model fit in a long-standing cohort), or trivially obtained samples (e.g., lab measurements in an institution with substantial health IT infrastructure), or data collected during routine quality control monitoring that can inform on test accuracy. It is not obvious that the same arguments transfer to exploratory studies. Publication and reporting bias are likely an issue, but the authors admit in their ultimate sentence that we need serious empirical data to understand their extent and impact. Transparency is important, but exactly how it is conceptualized in the context of exploratory studies is not as clear. Does pre-registration of hypotheses somehow guard against false positives? (But how can we practically define the family of hypotheses that were examined? And how do we make the FDR calculations?) And surely, the validity of the hypothesis is not dependent on the timing of the acquisition of the data. Finally, the ethical aspects appear less crisp. The ethical obligation towards patients whose data were recorded as a matter of routine does not seem to be of the same kind as the obligation to those who consented to be the subjects of an experiment. The ethical obligation to future patients is also rather generic as a concept, and it appears to us, of secondary importance. The added bureaucracy may be a negligible hassle compared to the large investment that went into expensive studies, but is not negligible in itself. Based on these considerations, we are at best ambivalent about the practicality or value of mandating that all exploratory studies be pre-registered.<br>-- Between the two extremes we might consider analyses that are protocol-driven and assess test accuracy using data gathered prospectively but for other purposes. In some cases, collection of data may be completed before the test-at-hand has been identified, e.g., assessing novel molecular markers in prospectively collected archival samples, or developing predictive models in large government-funded cohorts. In some cases prospective registration is meaningful only for the analysis and not for data collection. In some cases registration is of dubious practicality (should we really register yet-another-analysis in the Framingham study? To what end – surely not to calculate the number of false research claims!). In other cases, as in a preplanned spin-off study of test accuracy nested in a large RCT we might think differently.<br>Our ambivalence is genuine, and stems from our assessment that we have to understand better the tradeoffs of mandatory study pre-registration. To this end, as the authors mention, we need more empirical data.<br>Suggested actions for the authors: The authors do allude to a discussion in the observational studies literature, but rather briefly and almost dismissively. If they indeed believe that the benefits are clearly in favor of universal mandatory registration, they should at least entertain the aforementioned thoughts in a fair manner in the discussion. It would also be reasonable to stratify the reported |

| | |
|---|---|
| | proportions according to the three groups of studies we proposed, or according to an operationally feasible definition of strata.<br><br>Minor questions for clarification of the study execution:<br>• Does the term diagnostic tests subsume screening tests, and tests that are done for treatment guidance (give estrogen therapy in early breast cancer or not?), or for clinical course monitoring (e.g., viral load in HIV carriers under treatment)?<br>• Did the authors exclude studies without accuracy measures at the abstract or full-text screening stage?<br>• How was in-kind support handled when describing funding sources? Should the reader interpret funding as different that sponsorship?<br><br>A few comments pertaining to typo's and other edits:<br>• In-text citation calls should be after punctuation marks throughout the document.<br>• Abstract-Conclusion: Consider using "Because" instead of "Since" in third sentence.<br>• What is already known box: Consider hyphenating "well-known" in first point.<br>• Introduction: 1st paragraph, 3rd sentence: Should state "This policy improves…"<br>• Methods: Change 2nd paragraph, 2nd sentence to read "We limited our search to papers published in English and had an abstract available."<br>• Methods: 2nd paragraph, last sentence should read "area under operator curve."<br>• Results: 5th paragraph, 2nd sentence should read "…but it was described much more vaguely or slightly differently."<br>• Discussion: 2nd paragraph, last sentence should read "…due to scarce and substandard reporting."<br><br>Overall, we found this study insightful, relevant to the BMJ audience, and congruent with the BMJ editors' past statements about the prospective registration of observational studies. We hope our comments will be helpful to the authors. Thank you kindly for providing us with the opportunity to review the manuscript.<br><br>Cordially,<br>Alexandra G. Ellis & Thomas A. Trikalinos, Center for Evidence-based Medicine, Brown University. |

- The manuscript received two reviews at The BMJ but the other reviewer declined to make the reviews public. Please contact BMJ Open editorial office for any further information.

## VERSION 1 – AUTHOR RESPONSE

*Comments:*
*We read the manuscript by Korevaar and colleagues with great interest. The authors motivate the paper as follows: It is very likely that publication bias and selective reporting bias are major problems in test accuracy studies. Unlike clinical trials, prospective registration for these studies is not required. The authors assert that prospective registration of test accuracy studies is necessary, and conduct empirical research to estimate the proportion of registered accuracy studies in a 2-month sample of recent reports published in journals with impact factor of 5 or more. This clearly written manuscript addresses an interesting issue. We have one comment, and several observations of much smaller import.*

*Comment: The authors' (apparent) position is that all diagnostic accuracy studies should be registered prospectively. We are very sympathetic to a mandate for prospective registration, but not as resolved that it should apply universally. Indulge us in this brief digression:*
*For RCTs (and trials in general), the research community largely agrees that the arguments favor mandatory prospective registration. Benefits of registration include reducing the impact of publication/reporting bias, promoting transparency, and fulfilling an ethical obligation to study participants (primarily) and future patients (secondarily). The main counterargument relates to the inconvenience of the added bureaucracy. For RCTs we have extensive data that publication/reporting biases are very much a problem. We have almost universal buy-in from the industry, academe, governments and other stakeholders that the RCT represents a large enough investment to justify the added costs of registration. The ethical obligations to the participants seem (and probably are) crisp: Those who consented to play dice with their (well-)being must be respected. Overall, demanding universal registration makes sense. Fortunately, prospective registration is also practical.*
*Studies of diagnostic accuracy are a mixed bag when it comes to designs and purpose, and this can tip the balance of the pros and cons for prospective registration, or even render prospective registration impractical. It may help to consider three groups of studies:*
*-- In one extreme we have protocol-driven test accuracy studies that collect data de novo and primarily for assessing accuracy (some can even be RCTs). We submit that the aforementioned arguments probably transfer, and we favor prospective registration for these studies.*
*-- In the other extreme we have exploratory ('opportunistic', without a protocol or a hypothesis) analyses of existing datasets (e.g., a predictive model fit in a long-standing cohort), or trivially obtained samples (e.g., lab measurements in an institution with substantial health IT infrastructure), or data collected during routine quality control monitoring that can inform on test accuracy. It is not obvious that the same arguments transfer to exploratory studies. Publication and reporting bias are likely an issue, but the authors admit in their ultimate sentence that we need serious empirical data to understand their extent and impact. Transparency is important, but exactly how it is conceptualized in the context of exploratory studies is not as clear. Does pre-registration of hypotheses somehow guard against false positives? (But how can we practically define the family of hypotheses that were examined? And how do we make the FDR calculations?) And surely, the validity of the hypothesis is not dependent on the timing of the acquisition of the data. Finally, the ethical aspects appear less crisp. The ethical obligation towards patients whose data were recorded as a matter of routine does not seem to be of the same kind as the obligation to those who consented to be the subjects of an experiment. The ethical obligation to future patients is also rather generic as a concept, and it appears to us, of secondary importance. The added bureaucracy may be a negligible hassle compared to the large investment that went into expensive studies, but is not negligible in itself. Based on these considerations, we are at best ambivalent about the practicality or value of mandating that all exploratory studies be pre-registered.*
*-- Between the two extremes we might consider analyses that are protocol-driven and assess test accuracy using data gathered prospectively but for other purposes. In some cases, collection of data may be completed before the test-at-hand has been identified, e.g., assessing novel molecular markers in prospectively collected archival samples, or developing predictive models in large government-funded cohorts. In some cases prospective registration is meaningful only for the analysis and not for data collection. In some cases registration is of dubious practicality (should we really register yet-another-analysis in the Framingham study? To what end – surely not to calculate the number of false research claims!). In other cases, as in a preplanned spin-off study of test accuracy nested in a large RCT we might think differently.*

*Our ambivalence is genuine, and stems from our assessment that we have to understand better the tradeoffs of mandatory study pre-registration. To this end, as the authors mention, we need more empirical data.*

*Suggested actions for the authors: The authors do allude to a discussion in the observational studies literature, but rather briefly and almost dismissively. If they indeed believe that the benefits are clearly in favor of universal mandatory registration, they should at least entertain the aforementioned thoughts in a fair manner in the discussion. It would also be reasonable to stratify the reported proportions according to the three groups of studies we proposed, or according to an operationally feasible definition of strata.*

Reply: We agree with the reviewer that the benefits of prospective registration do not equally apply to test accuracy studies with different study designs. We sympathize with the arguments the reviewer has given.

We therefore have added the following paragraph to the discussion section:

> "We also strongly recommend that authors of test accuracy studies register their studies before initiation, and that journal editors start to think about expanding required registration to this type of research. An important question that should be addressed before such a requirement can be implemented is whether this should apply to any study on test accuracy, or only to those with specific study designs. The recent announcement of Lancet and the British Medical Journal that they would encourage researchers to register observational studies in a manner similar to what has become requirement for clinical trials caused some disapproving reactions. Criticism especially focused on the fact that observational studies vary widely in their design, and that prospective registration is not as useful for one type of study as it is for the other. These design issues also apply to test accuracy studies. Study data can be collected prospectively or retrospectively, and study aims, hypotheses and protocols can be formulated before or after the analysis of the data. All the reasons for registering clinical trials seem to equally apply to protocol-driven test accuracy studies with a-priori defined aims, irrespective of whether their data collection was prospective or retrospective. Some test accuracy studies, however, are exploratory in nature. Such studies often do not have a pre-defined protocol or aim, and existing data-sets are used to explore potentially interesting findings. The benefits of study registration are not as clear for such studies. For example, although non-publication and selective reporting are likely to be prevalent among exploratory studies, it would be impossible to determine whether the study has been registered before the post-hoc hypothesis was formulated. In addition, the bureaucratic load of registering every post-hoc analysis would be enormous and probably outweigh the benefits. We believe that at least all protocol-driven test accuracy studies with a-priori defined aims should be registered."

*Minor questions for clarification of the study execution:*

- *Does the term diagnostic tests subsume screening tests, and tests that are done for treatment guidance (give estrogen therapy in early breast cancer or not?), or for clinical course monitoring (e.g., viral load in HIV carriers under treatment)?*

Reply: We have added the following sentence to the methods sections:

> "Tests for screening, diagnosis, staging, monitoring, prediction, or prognosis were all eligible."

- *Did the authors exclude studies without accuracy measures at the abstract or full-text screening stage?*

Reply: We have added the following sentence to the methods section:

> "Studies that did not provide an accuracy measure in their abstract, but were deemed likely to publish one in their full-text, were also tagged as potentially eligible."

- *How was in-kind support handled when describing funding sources? Should the reader interpret funding as different that sponsorship?*

Reply: We categorized studies that clearly described a source of support as “industry involvement” or “sources of funding not including an industrial party” (e.g. universities, general hospitals, or grants from other non-profit organizations). Studies that did not report a funder, or that only used terms such as “no external funding” (or alike), were categorized as “no (external) funding reported”.

To further clarify this, we have changed the original sentence referring to our handling of funding sources in the methods section as follows:

> “We extracted the funding sources from the full publication. Studies that clearly described a source of support were categorized into those reporting some form of industry involvement and those reporting sources of funding not including an industrial party. Studies that did not report a source of support, or only indicated that “no external funding” was obtained, were categorized as “no (external) funding reported”.”

*A few comments pertaining to typo’s and other edits:*

- *In-text citation calls should be after punctuation marks throughout the document.*
- *Abstract-Conclusion: Consider using “Because” instead of “Since” in third sentence.*
- *What is already known box: Consider hyphenating “well-known” in first point.*
- *Introduction: 1st paragraph, 3rd sentence: Should state “This policy improves…”*
- *Methods: Change 2nd paragraph, 2nd sentence to read “We limited our search to papers published in English and had an abstract available.”*
- *Methods: 2nd paragraph, last sentence should read “area under operator curve.”*
- *Results: 5th paragraph, 2nd sentence should read “…but it was described much more vaguely or slightly differently.”*
- *Discussion: 2nd paragraph, last sentence should read “…due to scarce and substandard reporting.”*

Reply: We have changed these items according to the reviewer’s suggestions.

*Overall, we found this study insightful, relevant to the BMJ audience, and congruent with the BMJ editors’ past statements about the prospective registration of observational studies. We hope our comments will be helpful to the authors. Thank you kindly for providing us with the opportunity to review the manuscript.*