PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Submissions from the SPRINT Data Analysis Challenge on Clinical Risk Prediction: A Cross-Sectional Evaluation
AUTHORS	Jackevicius, Cynthia; An, JaeJin; Ko, Dennis T.; Ross, Joseph; Angraal, Suveen; Wallach, Joshua; Koh, Maria; Song, Jeeeun; Krumholz, Harlan

VERSION 1 - REVIEW

REVIEWER	Ruilope Institute of Investigation Hospital 12 de Octubre Madrid 28041 Spain
REVIEW RETURNED	09-Sep-2018

GENERAL COMMENTS	Interesting paper reflecting the different interpretations from different authors about a defined series of data. This probably has been facilitated by the enormous repercussion of the SPRINT trial

REVIEWER	Suzanne Oparil, MD Division of Cardiovascular Disease Department of Medicine, School of Medicine The University of Alabama at Birmingham Birmingham, Alabama, United States I served as Director/PI of the UAB Clinical Center Network (CCN) for the NIH/NHLBI-funded Systolic Blood Pressure Intervention Trial (SPRINT)
REVIEW RETURNED	10-Sep-2018

GENERAL COMMENTS	The authors have done a commendable job of assessing the methods, results and clinical performance of submissions to the SPRINT Challenge for clinical prediction tools or clinical risk scores. It is striking that of 143 submissions, only 29 met inclusion criteria and only 9 of those developed and provided risk prediction tools. Interestingly, clinical performance of the evaluable risk prediction tools was poor. The methods employed in the analysis are clearly stated and the analysis is interpretable, even to the non-statistician. The manuscript is useful in illuminating some of the tangible results of the SPRINT Challenge for clinical prediction tools in particular, and sheds light on the SPRINT Challenge in general, an area about which we have heard little in recent months. Finally, the manuscript clearly indicates that there is an unmet need for developing risk prediction tools in SPRINT eligible populations.

	Minor Specific Comment: Reference 3 is not cited in the manuscript.

REVIEWER	Dan Weiner Tufts Medical Center
REVIEW RETURNED	10-Sep-2018

GENERAL COMMENTS	In this manuscript, Jackevicius and colleagues describe the characteristics of 29 submissions to the SPRINT Data Analysis Challenge that described the development of a clinical prediction tool or risk score. This is an interesting exercise given the nature of the SPRINT Challenge, which put a finite number of variables into the public domain for a semi-competitive challenge for best use of the data. Jackevicius and colleagues describe some remarkable heterogeneity in methods and results. In part, this is consistent with the advertising of the challenge, which promoted access to trainees and junior investigators as well as more experienced teams. The heterogeneity of approaches is interesting, but we already know from published literature that tremendous heterogeneity exists with predictive modeling, even when the data sources and 'outcome' are similar. I am hesitant on the conclusion of the abstract, which states: "Our findings may be used to stimulate researchers to further optimize the development of risk prediction tools in SPRINT-eligible populations" as I am not really sure what the authors mean by this. It is critical to note that these submissions to the SPRINT Challenge did not undergo any standardized vetting or peer review process. My major comment is that I am really unsure as to what to do with these results. As a SPRINT investigator and an attendee at the meeting in Boston, this research satisfies some of my curiosity about what was produced. However, I am not sure if that curiosity is generalizable to other researchers or clinicians given the somewhat unique SPRINT Challenge structure and timing. My other comment is that I cannot quite figure out the precise intent of the low risk case that was included. At first I thought that the authors deliberately included a case that would not be SPRINT eligible in order to assess whether, using data derived from SPRINT, this patient would be deemed sufficiently high risk for targeting a more intensive BP goal. But, in the discussion, they state 'to a high and a low risk SPRINT-eligible patient case.' The 60 yo woman, as described is not eligible. This vignette suffers from a lack of detail in the kidney function variable as the eGFR is not reported; rather the serum creatinine is reported as 1.0 mg/dL. Of note, a serum creatinine of 1.01 mg/dL results in an MDRD eGFR of 59 in a white 60 yo woman, while 1.00 mg/dL results in an eGFR of 60 mL/min per 1.73m2. Using the ACC/AHA calculator and assigning an HDL of 50 and total cholesterol of 180, this patient has a 4.5% 10-year risk of MI – so low risk. Depending on the precise serum creatinine measure, this patient would have qualified for SPRINT based on eGFR criteria (if the serum

	creatinine is 1.01 mg/dL) and would have been lumped into gaining benefit if one solely looks at SPRINT data; in contrast, if 1.0 represents 1.00 mg/dL, the eGFR is 60.1 mL/min and she would NOT have qualified for SPRINT. This is a really important point for the exercise that you are engaging in with the current manuscript. On what factors were the experts (HMK and JSR) ranking risk in the hypothetical cases? I think it is critical to note that these abstracts were just that, abstracts, produced fairly quickly in response to a competition. These were not deemed ready for 'prime time' and, unless the group competing had cleaned data from a second similar data source, external validation and other standard methodologies would not be pursued. Minor comment: SPRINT wasn't terminated after 3.3 years; the trial itself approached 5 years when the allocation to the randomization strata was discontinued. There was a median follow-up of ~3.3 years (depending on the outcome), but that is not the same as the trial being terminated after 3.3 years. Would just clarify this as it implies that maximum total follow-up was only 3.3 years. In the discussion, the phrase that begins: "While some investigators…" is not a sentence.

REVIEWER	Mattias Brunström Department of Public Health and Clinical Medicine, Umeå University, Sweden
REVIEW RETURNED	11-Sep-2018

GENERAL COMMENTS	Thank you for the opportunity to review the manuscript by Prof. Jackevicius and colleagues. The manuscript is a summary of different risk prediction models derived from the SPRINT Data Analysis Challenge, with the aim to assess their characteristics and applicability. The methods are clearly described, results are reported in a straightforward way, and the manuscript is overall well written. I´ve got some general thoughts about the manuscript. - Firstly, is this really a systematic review? The paper describes different models derived from one single dataset, published at the same time, on the same platform. Generally, the scope of a systematic review is defined by a scientific question for which all available evidence, from all available data sources, is sought and critically assessed. When the scope of the review is defined by the data source, it is by definition not comprehensive. In my mind this is a cross-sectional meta-study of SPRINT-derived prediction models. - As a result of the above, the generalizability of the findings from this study is not clear. It is not a comprehensive assessment of studies derived from open data initiatives, nor is it a comprehensive assessment of prediction models/decision tools within the field of hypertension/cardiovascular prevention. It might not even be a comprehensive assessment of prediction models

	derived from SPRINT, because a search for such models outside the framework of the Data Analysis Challenge was not conducted. These limitations should be more clearly described in the discussion, but also highlights potentially interesting future extensions of the project. Apart from the above, I think the paper includes several interesting results. Especially, the discrepancy between different prediction models in terms of estimated risks, and the very low frequency of publication in peer-reviewed journals for the submitted abstracts. In a way, such abstracts represent a new form of grey literature that may become increasingly important for future authors of systematic reviews.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Ruilope

Institution and Country: Institute of Investigation, Hospital 12 de Octubre, Madrid, 28041, Spain

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Interesting paper reflecting the different interpretations from different authors about a defined series of data. This probably has been facilitated by the enormous repercussion of the SPRINT trial

Response: Thank you for the positive feedback.

Reviewer: 2

Reviewer Name: Suzanne Oparil, MD

Institution and Country: Division of Cardiovascular Disease, Department of Medicine, School of Medicine, The University of Alabama at Birmingham, Birmingham, Alabama, United States

Please state any competing interests or state ‘None declared’: I served as Director/PI of the UAB Clinical Center Network (CCN) for the NIH/NHLBI-funded Systolic Blood Pressure Intervention Trial (SPRINT)

Please leave your comments for the authors below

General Comments:

The authors have done a commendable job of assessing the methods, results and clinical performance of submissions to the SPRINT Challenge for clinical prediction tools or clinical risk scores. It is striking that of 143 submissions, only 29 met inclusion criteria and only 9 of those developed and provided risk prediction tools. Interestingly, clinical performance of the evaluable risk prediction tools was poor. The methods employed in the analysis are clearly stated and the analysis is

interpretable, even to the non-statistician. The manuscript is useful in illuminating some of the tangible results of the SPRINT Challenge for clinical prediction tools in particular, and sheds light on the SPRINT Challenge in general, an area about which we have heard little in recent months. Finally, the manuscript clearly indicates that there is an unmet need for developing risk prediction tools in SPRINT eligible populations.

Response: Thank you for your interest and for your positive feedback.

Minor Specific Comment:

Reference 3 is not cited in the manuscript.

Response: We have now cited reference 3 in the introduction. Thank you for noticing this.

Reviewer: 3

Reviewer Name: Dan Weiner

Institution and Country: Tufts Medical Center

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

In this manuscript, Jackevicius and colleagues describe the characteristics of 29 submissions to the SPRINT Data Analysis Challenge that described the development of a clinical prediction tool or risk score. This is an interesting exercise given the nature of the SPRINT Challenge, which put a finite number of variables into the public domain for a semi-competitive challenge for best use of the data.

Jackevicius and colleagues describe some remarkable heterogeneity in methods and results. In part, this is consistent with the advertising of the challenge, which promoted access to trainees and junior investigators as well as more experienced teams. The heterogeneity of approaches is interesting, but we already know from published literature that tremendous heterogeneity exists with predictive modeling, even when the data sources and 'outcome' are similar.

Response: Thank you for the considerate feedback.

I am hesitant on the conclusion of the abstract, which states: "Our findings may be used to stimulate researchers to further optimize the development of risk prediction tools in SPRINT-eligible populations" as I am not really sure what the authors mean by this.

Response: By collating the diversity of approaches taken for developing risk prediction tools, we provide researchers interested in this area an overview of various approaches taken by others, as well as highlight the methodological areas, such as, external validation that would benefit from further work. We hope this clarifies the intention of this statement in our abstract conclusion.

We revised the abstract conclusion as follows:

"Our findings may be used to stimulate By collating an overview of the range of approaches taken, researchers to may further optimize the development of risk prediction tools in SPRINT-eligible populations, and our findings may as well as to inform the conduct of future similar open science projects."

It is critical to note that these submissions to the SPRINT Challenge did not undergo any standardized vetting or peer review process.

Response: We agree with the reviewer. In the discussion section, we added the following sentences: “Moreover, these SPRINT Challenge submissions did not undergo a standardized peer review process. Therefore, the quality of the abstracts may be lower than those in peer-reviewed publications, which may have impacted our study findings.”

My major comment is that I am really unsure as to what to do with these results. As a SPRINT investigator and an attendee at the meeting in Boston, this research satisfies some of my curiosity about what was produced. However, I am not sure if that curiosity is generalizable to other researchers or clinicians given the somewhat unique SPRINT Challenge structure and timing.

Response: We agree that the SPRINT Challenge experience may be somewhat unique. However, we anticipate seeing more data sharing opportunities in the near future with the recent interest in the open science movement. Therefore, our findings are likely to be of interest to researchers, clinicians, and organizers of open science competitions, even outside of the SPRINT Challenge. Moreover, as reviewer #4 notes, these abstracts represent a new form of grey literature that may become increasingly important for future authors of systematic reviews, and provides important information to other researchers.

We have revised the discussion as follows:

“Furthermore, we anticipate seeing more data sharing opportunities in the future with the recent interest in the open science movement. Therefore, our findings are likely to be of interest to researchers and clinicians, and that those organizing future open science initiatives may also benefit from our systematic evaluation. We offer the following suggestions to organizers of open science competitions to enhance the experience and potential productivity…”

My other comment is that I cannot quite figure out the precise intent of the low risk case that was included. At first I thought that the authors deliberately included a case that would not be SPRINT eligible in order to assess whether, using data derived from SPRINT, this patient would be deemed sufficiently high risk for targeting a more intensive BP goal. But, in the discussion, they state ‘to a high and a low risk SPRINT-eligible patient case.’ The 60 yo woman, as described is not eligible. This vignette suffers from a lack of detail in the kidney function variable as the eGFR is not reported; rather the serum creatinine is reported as 1.0 mg/dL. Of note, a serum creatinine of 1.01 mg/dL results in an MDRD eGFR of 59 in a white 60 yo woman, while 1.00 mg/dL results in an eGFR of 60 mL/min per 1.73m2. Using the ACC/AHA calculator and assigning an HDL of 50 and total cholesterol of 180, this patient has a 4.5% 10-year risk of MI – so low risk. Depending on the precise serum creatinine measure, this patient would have qualified for SPRINT based on eGFR criteria (if the serum creatinine is 1.01 mg/dL) and would have been lumped into gaining benefit if one solely looks at SPRINT data; in contrast, if 1.0 represents 1.00 mg/dL, the eGFR is 60.1 mL/min and she would NOT have qualified for SPRINT. This is a really important point for the exercise that you are engaging in with the current manuscript.

Response: Our intention was to create a low risk patient case who is eligible for the SPRINT trial, therefore, an eGFR of 59.

On what factors were the experts (HMK and JSR) ranking risk in the hypothetical cases?

Response: The clinical experts were asked to use their clinical judgment as they would do in their typical practice setting in ranking risk in the hypothetical cases.

I think it is critical to note that these abstracts were just that, abstracts, produced fairly quickly in response to a competition. These were not deemed ready for ‘prime time’ and, unless the group competing had cleaned data from a second similar data source, external validation and other standard methodologies would not be pursued.

Response: We agree with the reviewer on this important point, although it is not possible to know how much time they devoted to their abstracts. We have added this point on page 15 in the discussion: “It is possible that other research teams may not have published their work yet in order to complete their validation, or given the short timeline for the competition, may not have had access to a similar external data source with which to conduct external validation.”

We also have included on page 16 in the discussion, the following sentence: “Some investigators may have viewed the competition as preliminary work, or did not enter the competition with the intent to publish.”

Minor comment:

SPRINT wasn’t terminated after 3.3 years; the trial itself approached 5 years when the allocation to the randomization strata was discontinued. There was a median follow-up of ~3.3 years (depending on the outcome), but that is not the same as the trial being terminated after 3.3 years. Would just clarify this as it implies that maximum total follow-up was only 3.3 years.

Response: Thank you for noting this. We have changed this in the introduction to add that 3.3 years is the median duration of follow-up (Page 4).

In the discussion, the phrase that begins: “While some investigators…” is not a sentence.

Response: Thank you for noticing this. We removed “while” and it is now a sentence.

Reviewer: 4

Reviewer Name: Mattias Brunström

Institution and Country: Department of Public Health and Clinical Medicine, Umeå University, Sweden

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Thank you for the opportunity to review the manuscript by Prof. Jackevicius and colleagues. The manuscript is a summary of different risk prediction models derived from the SPRINT Data Analysis Challenge, with the aim to assess their characteristics and applicability. The methods are clearly described, results are reported in a straightforward way, and the manuscript is overall well written.

Response: Thank you for your kind feedback.

I´ve got some general thoughts about the manuscript.

- Firstly, is this really a systematic review? The paper describes different models derived from one single dataset, published at the same time, on the same platform. Generally, the scope of a systematic review is defined by a scientific question for which all available evidence, from all available data sources, is sought and critically assessed. When the scope of the review is defined by the data source, it is by definition not comprehensive. In my mind this is a cross-sectional meta-study of SPRINT-derived prediction models.

Response: Thank you for noting this. As the reviewer and the associate editor suggested, we have changed the title to, “Submissions from the SPRINT Data Analysis Challenge on Clinical Risk Prediction: A Cross-Sectional Evaluation”.

- As a result of the above, the generalizability of the findings from this study is not clear. It is not a comprehensive assessment of studies derived from open data initiatives, nor is it a comprehensive assessment of prediction models/decision tools within the field of hypertension/cardiovascular prevention. It might not even be a comprehensive assessment of prediction models derived from SPRINT, because a search for such models outside the framework of the Data Analysis Challenge was not conducted. These limitations should be more clearly described in the discussion, but also highlights potentially interesting future extensions of the project.

Response: Thank you for noting this. We added the following to the discussion section (page 16); “It is important to note that this study reviewed SPRINT Challenge submissions only, and did not review clinical prediction models or clinical risk scores outside of the SPRINT Challenge. Future research can further evaluate prediction models outside of the SPRINT Challenge.”

Apart from the above, I think the paper includes several interesting results. Especially, the discrepancy between different prediction models in terms of estimated risks, and the very low frequency of publication in peer-reviewed journals for the submitted abstracts. In a way, such abstracts represent a new form of grey literature that may become increasingly important for future authors of systematic reviews.

Response: Thank you for pointing out the important potential for the abstracts to represent grey literature. We have added this point to a sentence in the discussion on page 16: “Just as few meeting abstracts get translated into publications, the SPRINT Challenge submissions may be experiencing the same fate, creating a new form of grey literature.14”

VERSION 2 – REVIEW

REVIEWER	Mattias Brunström Department of Public Health and Clinical Medicine, Umeå University, Sweden
REVIEW RETURNED	15-Dec-2018

GENERAL COMMENTS	The authors have considered previous comments in their revision. I have nothing to add.