Joint forces for making clinical prediction models contribute to science

Junfeng Wang[1], Yue Li[2,3]

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [2]Department of Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [3]Risk Modelling Department, ABN AMRO Bank N.V, Amsterdam, The Netherlands

Correspondence to: Junfeng Wang. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584 CX Utrecht, The Netherlands. Email: j.wang-4@umcutrecht.nl.

Recently, the ATM journal has published a special report "*In-depth mining of clinical data: the construction of clinical prediction model with R*" (1), which is a series of methodology of clinical prediction model (CPM) construction. We read the 96 pages long article with great interest, especially because this is a statistical paper written by a group of clinicians. We are very glad to see clinicians are interested in statistics and predictive modelling, and they have reasonable knowledge in the development and validation of CPMs.

Just as the authors mentioned, nowadays researchers can get access to huge amount of data, and data availability is no longer a limitation for clinical researches. Meanwhile, a lot of data need to be (properly) analyzed, and many studies could not be done or are with delay due to lack of statistical capability. It will be more efficient if clinicians can have the ability to perform some statistical analyses themselves, at least for the sake of quantity of publications.

To achieve this goal, a methodological paper with examples and implementations (i.e., R code), which can be easily understood by clinicians is more than welcome. This is the significant added value of this paper. The authors deliberately wrote the paper in a more practical way, so readers can quickly get hands-on experience in using R to perform almost all the analyses necessary in a prediction model paper. The paper can serve as a manual for starters and will guide them to go through the entire process of developing and evaluating a CPM.

As always, statisticians are very critical in methodology papers. We would suggest improving the current paper at least in the following points: Figure 2 in page 14 shows the research process suggested by the authors. Actually, a more logical thinking is first to determine the outcome and then try to find variables which can predict the outcome, but not the other way around.

In both page 31 and page 63, the explanation of useful number of pairs is questionable. To the best of our knowledge, the C Index for survival outcome needs to take into account the censored observations (2-4). In the meanwhile, there is no need to repeat exactly the same paragraph within one paper. The same applies to where the authors introduced the concept of discrimination ability, in both page 27 and page 30.

Some important elements are missing, e.g., why internal validation is needed in model development, how to perform internal validation with R.

Some terminologies in the paper are not commonly used in the CPM field, which may bring difficulties in communication with other researchers especially with statisticians.

The language is a bit difficult to follow, which is understandable since the paper was written by a group of non-native English speakers. With the help of a professional academic writer, the quality of this report can be greatly improved.

For the readers with more experience in data analysis or with a statistical background, this paper may not be

informative enough for them. For their further development in the CPM field, we recommend the classic text book Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis (5) by Frank Harrell and another book more specific for CPM, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating (6) by Ewout Steyerberg.

From the words in the paper, I got the feeling, most clinicians think statisticians are the ones always say "*Wait, the analysis looks not right*" but never give them a definite answer what is correct. Maybe that is because statisticians are good at rejecting the null hypothesis but can never prove a hypothesis is true. This is just joking. Statisticians were struggling with preventing mis-use of the statistical methods and mis-interpretation of the analysis results. The increasing needs for CPMs is definitely a new opportunity as well as a new challenge for them.

When I was attending the Cochrane Colloquium in Quebec city, Canada, a senior researcher said to me "*You are the rock star in this room*", when he knew I am a statistician. However, from my experience it seems not true. We are more like the drummer in a band: everyone can hear the rhythm (as every reader will read the methods, tables, figures, analyses, results in a paper produced by statisticians) but most of the time we are invisible to the audiences (Did statisticians get enough credits for their work? At which place are the statisticians in the authors list?).

In most conventional clinical studies, statistical models are used as tools to explore or confirm some scientific findings, while in CPM studies, the model itself is the final product. Thanks to CPM studies, statisticians finally have the opportunity to stand in the middle of the stage.

Statisticians may not pursue the vocal role, but they do want to voice for the new challenges brought by CPMs. A lot of efforts were made to improve the quality of CPM studies by groups of statisticians and methodologists, including but not limited to the PROGnosis RESearch Strategy (PROGRESS) Partnership (7-10), Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD) statements (11), and STRATOS (STRengthening Analytical Thinking for Observational Studies) initiative (12). All these contributions should be acknowledged.

To many clinicians, predictive modelling is supposed to be advanced clinical studies, since more complicated statistical methods and heavy data analyses work are needed in such kind of studies. However, from statistical point of view, it may not be true.

Recently, Bradley Efron, who is the real "statistical rock star" (13) gave a lecture at the opening ceremony of the Leiden University Center for Statistical Science with his insights that "*Prediction is easier than estimation and estimation is easier than attribution.*" (14). Even a weak learner CPM can yield an acceptable or even good prediction performance. One of the audiences, Ewout Steyerberg, a big name in the CPM field, commented to this talk "*New prediction algorithms have had impact on the public perception of predictability rather than on science*".

A lot of CPMs are produced with the following formulation:

1 dataset + 1 algorithm = 1 new prediction model

Some better studies use 2 datasets (1 for derivation and 1 for validation) and N algorithms (although I wouldn't say the latter makes the study better). It seems that old CPMs can only have very short shelf life due to the booming of new CPMs, but the fact is probably that the new CPMs might be never on the shelf.

Are we producing rubbishes? What can we do to improve the quality of CPM studies?

The quality requirements of CPMs are sub-optimal compared to the leading practices in other fields. Prediction modelling is widely used in banks, insurance companies and government economic research institutes. In these fields, the quality of the predictive models is strictly overseen by model validators, model risk auditors and regulators. If the modelers there follow the routine of current CPMs to build their models and make predictions, these models cannot even meet company's internal standards and compliance requirements, not to mention publishing them as advanced researches in top journals.

A good quality CPM study is not only about large dataset, proper statistical analysis, and good model performance, researchers should pay more attention to the conceptual soundness of the CPM and the study should have some impacts on the clinical practice.

The model performance (i.e., predictive accuracy) is usually considered as the gold standard to judge a CPM. However, a CPM study should not be a Kaggle competition, it needs to be part of the science. So, we encourage researchers, especially clinicians to put more efforts into the conceptual soundness of a CPM.

Conceptual soundness includes at least the following aspects: the model design should be logic and transparent; the correctness of physiology or pathology mechanism underpinning the model; and the methodology should be

more or less consistent with other models in the same field. Even in external validations, the conceptual soundness of a CPM is seldom evaluated by researchers. We expect this issue can be solved in systematic reviews and meta-analyses of CPMs (15), which is developing fast.

The improvement of quality of CPM studies needs the input from joint forces of clinicians, methodologists, statisticians and other stakeholders. Let's collaborate more closely in CPM!

Acknowledgments

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Zhou ZR, Wang WW, Li Y, et al. In-depth mining of clinical data: the construction of clinical prediction model with R. Ann Transl Med 2019;7:796.
2. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.
3. Uno H, Cai T, Pencina MJ, et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011;30:1105-17.
4. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics 2005;61:92-105.
5. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd edition. Berlin: Springer, 2015.
6. Steyerberg E. Clinical prediction models. Berlin: Springer, 2009.
7. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ 2013;346:e5595.
8. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med 2013;10:e1001380.
9. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10:e1001381.
10. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013;346:e5793.
11. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594.
12. Sauerbrei W, Abrahamowicz M, Altman DG, et al. Strengthening analytical thinking for observational studies: the STRATOS initiative. Stat Med 2014;33:5413-32.
13. Castelvecchi D. Statistical 'rock star' wins coveted international prize. Nature 2018. doi: 10.1038/d41586-018-07395-w.
14. Efron B. Prediction, Estimation, and Attribution. Available online: https://statprize.org/pdfs/materials/2019Efron-presentation.pdf
15. Debray TP, Damen JA, Riley RD, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2019;28:2768-86.