Clinical prediction models: evaluation matters

Hong-Qiu Gu¹², Chelsea Liu³

¹China National Clinical Research Center for Neurological Diseases, ²National Center for Healthcare Quality Management in Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing 100085, China; ³Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA

Correspondence to: Hong-Qiu Gu, PhD. China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing 100085, China. Email: guhongqiu@yeah.net.

Provenance: This is an invited article commissioned by the Editorial Office, Annals of Translational Medicine.


Submitted Nov 18, 2019. Accepted for publication Nov 25, 2019.
doi: 10.21037/atm.2019.11.143

View this article at: http://dx.doi.org/10.21037/atm.2019.11.143

Clinical prediction models, also known as “prognostic models”, “risk scores”, or “prediction rules”, have received increasing attention in recent years (1,2). Clinical prediction models appear to be the next rising star in personalized medicine, just as systematic reviews and meta-analyses were in the era of evidence-based medicine. Essentially, a clinical prediction model is a mathematical equation that estimates the probability of having a disease or condition in the present (diagnostic prediction model) or the probability of developing a particular disease or outcome in the future (prognostic prediction model) (3). The core techniques for developing a clinical prediction model, such as logistic regressions for binary outcomes or Cox proportional hazards regressions for time-to-event outcomes, are commonly used for effective size estimation in traditional clinical research as well. However, clinical prediction models focus on prediction rather than hypothesis testing. Additionally, the process of developing, internally validating, and externally validating a well-performing prediction model requires understanding concepts including discrimination, calibration, concordance statistic, calibration slope, net reclassification index, integrated discrimination improvement, etc. (4-6). Therefore, a detailed handbook will serve as a much-needed introduction for clinical researchers.

In this issue of Annals of Translational Medicine (ATM), Zhou and colleagues present an elaborate report on the construction of clinical prediction models with the R software (7). This report consists of a series of technical notes comprising 16 sections, including a general framework of clinical prediction models, model development, internal validation, and external validation based on logistic regressions or Cox models with or without competing risks. Authors also discuss the procedures for evaluating clinical utility using decision curve analysis, dealing with outliers and missing data, and selecting variables using ridge regression or LASSO regression.

This report by Zhou et al. serves to provide valuable guidance to readers without extensive background in statistics. It was composed by a group of clinicians rather than statisticians or methodologists, and thus minimized the use of technical terms to interpret the concept and process of clinical prediction model construction and evaluation. Authors included easily understandable examples in R, a user-friendly (with the support of RStudio), cross-platform, open-source software that can be quickly installed and easily used by all readers. However, several concerns should be carefully considered. In order to provide a broad overview for readers, some important frameworks on the types of prediction model studies and validation studies should be included in the first section of the report (Table 1). The report emphasizes data-driven methods for variable selection and model development, but a sophisticated analysis cannot salvage a poorly-designed study or poor data collection procedures. Other considerations such as existing evidence in the literature on the value of a particular prediction model, the cost of the collecting data required by the model, and potential applications in a clinical setting
warrant further discussion. Lastly, the structure of this report may be more easily understandable if arranged in order of general concept, followed by a simple example with the core processes of model development and validation, and concluding with a detailed step-by-step description of the process.

Although clinical prediction models are promising tools for decision-making, risk stratification and prognosis management, whether their use will eventually lead to improvements in healthcare delivery or patient outcomes remains unclear, owing to methodological shortcomings, incomplete presentation, and especially to the lack of external validation and model impact studies (8). As such, prospective validation to examine model stability, reproducibility and external validity in independent samples is needed. Model impact studies, usually in the form of cluster randomized controlled trials, are also necessary to assess the validity of a prediction model before it is recommended for use in clinical practice (9). In addition, when introducing a well-validated prediction model into clinical practice, it should be nested or integrated into clinical workflows. The AF-ALERT study may serve as a template for implementation (10).

In summary, this special report by Zhou et al. has provided critical technical notes on the R programming language with regards to clinical prediction models. Challenges remain, such as choosing the most urgent clinical needs for clinical prediction model studies, developing a study protocol and statistical analysis plan, and integrating the models into clinical practice. In particular, efforts should be made to systematically evaluate clinical prediction models before routine clinical use through external validation studies, systematic reviews and model impact studies. Nevertheless, joint efforts from clinicians, methodologies, statisticians, and engineers can drive increasing interest and knowledge in clinical prediction models, an encouraging development for an emerging field with strong potential to improve the quality of patient care.

**Acknowledgments**

None.

**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**


---

**Table 1** Five broad categories of prediction model studies and 6 types of development and validation studies

<table>
<thead>
<tr>
<th>Five broad categories</th>
<th>Six types of development and validation studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1: prognostic or diagnostic predictor finding studies</td>
<td>–</td>
</tr>
<tr>
<td>Category 2: prediction model development studies without external validation</td>
<td>Type 1: development only</td>
</tr>
<tr>
<td></td>
<td>Type 2: development and validation using resampling</td>
</tr>
<tr>
<td></td>
<td>Type 3: random split-sample development and validation</td>
</tr>
<tr>
<td>Category 3: prediction model development studies with external validation</td>
<td>Type 4*: nonrandom split-sample development and validation</td>
</tr>
<tr>
<td></td>
<td>Type 5: development and validation using separate data</td>
</tr>
<tr>
<td>Category 4: prediction model validation studies</td>
<td>Type 6: validation only</td>
</tr>
<tr>
<td>Category 5: model impact studies</td>
<td>–</td>
</tr>
</tbody>
</table>

*, although Type 4 may be considered an intermediary between internal and external validation, it is commonly referred to as “external validation studies”.

© Annals of Translational Medicine. All rights reserved. Ann Transl Med 2019 | http://dx.doi.org/10.21037/atm.2019.11.143
2015;162:W1-73.

Cite this article as: Gu HQ, Liu C. Clinical prediction models: evaluation matters. Ann Transl Med 2019. doi: 10.21037/atm.2019.11.143