Dissipating the Fog at the Crossroad: Predicting Survival after the Initiation of Kidney Replacement Therapy

Jean-Maxime Côté [1,2] and William Beaubien-Souligny [1,2]

The ability to predict the course of a critical illness complicated by AKI may be particularly crucial at the time when RRT is contemplated. Identifying overwhelming odds of death may be of great relevance for shared decision making because RRT initiation usually represents an escalation in the invasiveness of medical care and an opportunity to reevaluate the goals of care. Recently, the widespread use of electronic health records and easier access to administrative databases coupled with the emergence of innovative data science approaches, including machine learning, have led to the creation of prediction models for this specific purpose (1–6). Such models use clinical information available at RRT initiation to predict short-term outcomes. However, they have had mitigated success so far, and none have been meaningfully integrated into clinical practice.

The recent study published by Ganguli *et al.* (7) aimed to achieve this noble objective. The authors intended to develop a novel predictive model for in-hospital mortality in critically ill patients after RRT initiation. Using retrospective data from a single center between 2010 and 2015, they derived a model on the basis of data from 416 patients who were initially treated with either continuous (63%) or intermittent (37%) modality, from which 48% survived at hospital discharge. Using a combination of 14 demographic, clinical, and laboratory variables, they derived a promising model with very good discrimination after internal validation (c-statistic of 0.93; 95% confidence interval, 0.92 to 0.95). If this parameter alone was to be considered in isolation, we could argue that this model outperforms other efforts in this field, including sophisticated attempts using machine learning technology (5). However, as is almost always the case with prediction models, the devil is in the detail.

Although the authors should be congratulated for their efforts, there is still a long way to go before this model—or any other contemporary model—can reliably be used in clinical care for prognostication in the setting of acute RRT initiation (8). Indeed, it is important to consider carefully how prediction models should be developed and validated. From initial conceptualization to clinical adoption, five steps need to be planned and implemented to create a useful prediction model (Figure 1).

Step 1: Determine the Objective

Before performing any statistical analysis, clear modeling objectives should be planned: what is the intended purpose? This process may orient decisions related to subsequent steps. As previously mentioned, the most obvious clinical use of such models would be to orient discussions regarding goals of care at a critical point of care escalation. In this case, a simple model with a high ability to predict clinical futility (short-term mortality) would be sought. In contrast, the model could serve the purpose of selecting patients who might benefit from potential interventions in a trial, or for analytical use in epidemiologic research or quality improvement processes (*i.e.*, comparing mortality in different AKI cohorts). Understanding the purpose of the model is key in guiding decisions related to determining candidate variables and choosing an appropriate modeling strategy.

Step 2: Ensure the Quality and Representativeness of Source Data

Representativeness of the cohort from which data are extracted is an essential consideration. A model derived from a homogeneous and highly selected cohort of patients, for example a dataset from a large randomized clinical trial, may lead to poor generalizability, despite promising initial performance. Datasets used should be large but without compromise in regard to the quality and completeness of the data, most importantly regarding the studied outcome. Quality control and a deep understanding of the strength and weakness of the available data are invaluable. Although easy access to a large dataset may be seductive for researchers, overlooking these critical steps is setting up for failure before any analysis has taken place.

[1]Division of Nephrology, Centre Hospitalier de l'Université de Montréal, Montréal, Canada
[2]Research Center, Centre Hospitalier de l'Université de Montréal, Montréal, Canada

Correspondence: Dr. William Beaubien-Souligny, 1000 rue Saint Denis, Montréal, Canada, H2X0C1.
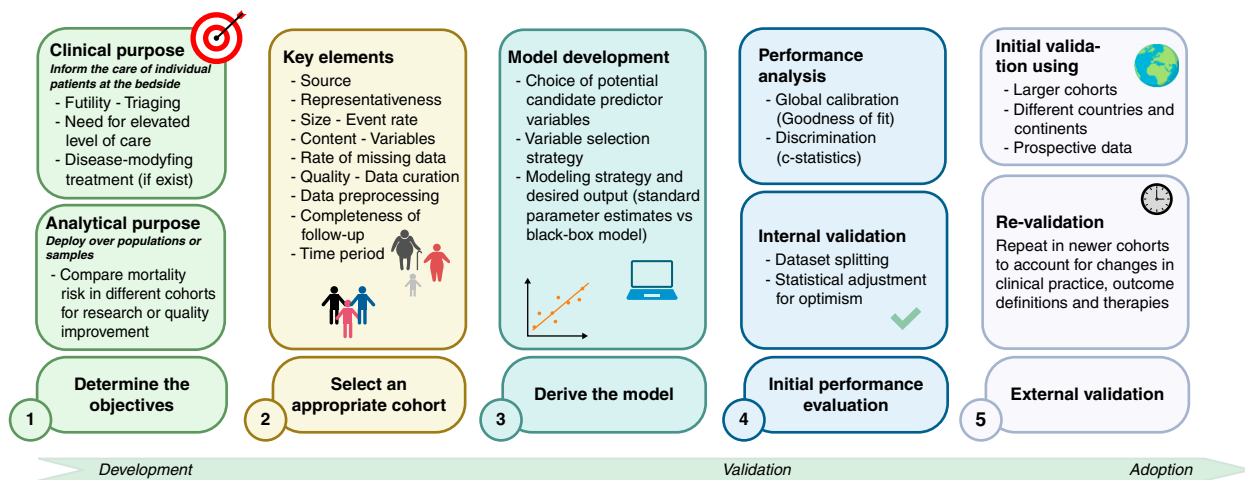Email: William.beaubien@gmail.com

Figure 1. | **Proposed scheme for prediction model development.** Created with BioRender.com.

Step 3: Model Development

The choice of variables included in a prediction model and how that choice was made are still frequently neglected. Candidate variables should respect three criteria: (1) present at the time when prediction would take place, (2) must have a clear definition, and (3) must be reliably collected, particularly for external validation in mind. Hence, urine output 1 week after RRT initiation does not respect the first criteria and should not be selected as a potential variable in a predictive model to be applied at the time of RRT initiation. Similarly, the choice to include subjective variables such as a semiquantitative fluid overload grading or the retrospective adjudication of the indication to initiate RRT, as chosen by *Ganguli et al.* (7), may raise concerns regarding external validity because they have a subjective quality.

Some variable selection methods have been known to produce poor generalizability by resulting in overly optimistic initial performance. In stepwise selection, the strategy used by Ganguli *et al.* (7), selection of variables is based on the effect of including (forward) or excluding (backward) each candidate variable on how closely the resulting model fits the data to subsequently determine whether each variable will be included in the final model (8). This process of optimizing the model's fit will naturally result in a high risk of overfitting. In other words, the derived model will be optimized for the training dataset but lack generalizability and underperform in other cohorts (4,9). An overly optimistic model may therefore mislead and prompt external validation efforts that are doomed to fail. To avoid the pitfalls of these methods, the TRIPOD guidelines and other experts recommend alternative strategies for variable selection, which include penalization methods (*LASSO, Ridge,* or *Elastic-Net* regularization) or even relying on *Full prespecification* (*a priori*) of predictor variables as detailed elsewhere (8,10).

Beyond these technical details, it is always important to keep the intended purpose in mind. This often involves striking a balance between simplicity (*e.g.,* a simple score with few predictors) and complexity (*e.g.,* machine learning algorithm) for which clinical implementation may be problematic. There are no black or white solutions because a simple score-based model for AKI-associated mortality could be easy to implement at the bedside but may be limited in its accuracy. Conversely, the widespread access to a smartphone (and applications) has made "complex" multivariate equations easier to use and therefore potentially applicable.

Step 4: Performance Analysis

Multiple metrics exist to quantify the performance of predictive models (11) according to the nature of data and outcomes reported, and that performance should be interpreted according to the context of its application, where the distinction between clinical and statistical significance is particularly important. A hypothetic model that discriminates mortality after RRT initiation with 70% accuracy might be statistically significant but is mostly useless if the aim is to guide important clinical decisions. On the other hand, a well-validated model with moderate discrimination (c-statistic of 80%) might be acceptable to select participants for an interventional trial who are most likely to benefit. Furthermore, discrimination is not the only consideration related to performance. Calibration is important to consider to understand the reliability of risk estimates remain mindful of the possibility of over- or underestimating the risk. The model proposed by *Ganguli et al.* (7) suggests both high discrimination and good calibration.

Step 5: Model Validation

Confirming the performance of the predictive model on data distinct from the derivation cohort is mandatory for clinical use, but most will fail at this essential and final step (4). An external validation strategy should be planned, usually when the project of creating a prediction tool is initially designed. Although several strategies have been proposed to alleviate the need for a cumbersome and costly multicenter prospective validation process, dataset splitting (to derive the model and then to validate it) or attempting validation on contemporary data from the same site(s) may

Table 1.   Proposed avenues to improve the success of prediction models

Steps	Avenues
Determine the objectives	• Identify the clinical need and potential future use by engaging relevant stakeholders, including patients, clinicians, researchers, and administrators • Carefully plan the subsequent steps and collaboration network to ensure feasibility and availability of appropriate expertise
Selection of source data for model derivation	• Develop high-quality curated datasets for the specific purpose of creating generalizable prediction models • Consider the exploration of new predictors such as biologic biomarkers and functional status
Model development	• Avoid the use of stepwise variable selection methods that produce overfitted and nongeneralizable models • Only use variables with a clear definition that can be collected prospectively in a reproducible manner • Align model development with the intended purpose and future integration strategy to clinical and/or research contexts
Performance analysis	• Appropriate internal validation and calibration to identify subsets in which the model may underperform
External validation	• Plan for external validation strategy in advance during the design phase • External validation may be repeated over time to account for a change in clinical practice • Plan for interventional studies to investigate the effect of implementation in a real-life setting

raise issues regarding generalizability. In nephrology, the Kidney Failure Risk Equation, derived from five "classic" variables and aiming to predict RRT at 2 and 5 years in patients with CKD, has succeeded the external validation process on more than 700,000 patients and is now easily available online and on smartphones (12). It is clear that achieving this goal is the result of a well-designed process and extensive collaboration in a given field.

Despite commendable efforts by Ganguli *et al.* (7) and other groups, the quest to derive a clinically useful and externally validated model to predict mortality after RRT initiation for AKI remains daunting. We believe that wide collaboration, careful planning, and using innovative approaches, some of which are presented in Table 1, are key elements in achieving meaningful progress. Hopefully, despite the many challenges, we will eventually dissipate the fog of uncertainty at the crossroad of RRT initiation.

Author Contributions

J.-M. Côté wrote the original draft of the manuscript. J.-M. Côté and W. Beaubien-Souligny were responsible for the visualization and reviewed and edited the manuscript. W. Beaubien-Souligny was responsible for conceptualization and supervision.

References

1. Demirjian S, Chertow GM, Zhang JH, O'Connor TZ, Vitale J, Paganini EP, Palevsky PM; VA/NIH Acute Renal Failure Trial Network: Model to predict mortality in critically ill adults with acute kidney injury. *Clin J Am Soc Nephrol* 6: 2114–2120, 2011 https://doi.org/10.2215/CJN.02900311
2. da Hora Passos R, Ramos JG, Mendonça EJ, Miranda EA, Dutra FR, Coelho MF, Pedroza AC, Correia LC, Batista PB, Macedo E, Dutra MM: A clinical score to predict mortality in septic acute kidney injury patients requiring continuous renal replacement therapy: the HELENICC score. *BMC Anesthesiol* 17: 21, 2017 https://doi.org/10.1186/s12871-017-0312-8
3. Kim Y, Park N, Kim J, Kim DK, Chin HJ, Na KY, Joo KW, Kim YS, Kim S, Han SS: Development of a new mortality scoring system for acute kidney injury with continuous renal replacement therapy. *Nephrology (Carlton)* 24: 1233–1240, 2019 https://doi.org/10.1111/nep.13661
4. Li DH, Wald R, Blum D, McArthur E, James MT, Burns KEA, Friedrich JO, Adhikari NKJ, Nash DM, Lebovic G, Harvey AK, Dixon SN, Silver SA, Bagshaw SM, Beaubien-Souligny W: Predicting mortality among critically ill patients with acute kidney injury treated with renal replacement therapy: Development and validation of new prediction models. *J Crit Care* 56: 113–119, 2020 https://doi.org/10.1016/j.jcrc.2019.12.015
5. Kang MW, Kim J, Kim DK, Oh KH, Joo KW, Kim YS, Han SS: Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. *Crit Care* 24: 42, 2020 https://doi.org/10.1186/s13054-020-2752-7
6. Lee BJ, Hsu CY, Parikh R, McCulloch CE, Tan TC, Liu KD, Hsu RK, Pravoverov L, Zheng S, Go AS: Predicting renal recovery after dialysis-requiring acute kidney injury. *Kidney Int Rep* 4: 571–581, 2019 https://doi.org/10.1016/j.ekir.2019.01.015
7. Ganguli A, Farooq S, Desai N, Adhikari S, Shah V, Sherman MJ, Veis JH, Moore J: A novel predictive model for hospital survival in critically ill dialysis-dependent acute kidney injury (AKI-D) patients: A retrospective single-center exploratory study [published online ahead of print January 25, 2022]. *Kidney360* https://doi.org/10.34067/KID.0007272021
8. Leisman DE, Harhay MO, Lederer DJ, Abramson M, Adjei AA, Bakker J, Ballas ZK, Barreiro E, Bell SC, Bellomo R, Bernstein JA, Branson RD, Brusasco V, Chalmers JD, Chokroverty S, Citerio G, Collop NA, Cooke CR, Crapo JD, Donaldson G,

Fitzgerald DA, Grainger E, Hale L, Herth FJ, Kochanek PM, Marks G, Moorman JR, Ost DE, Schatz M, Sheikh A, Smyth AR, Stewart I, Stewart PW, Swenson ER, Szymusiak R, Teboul JL, Vincent JL, Wedzicha JA, Maslove DM: Development and reporting of prediction models: Guidance for authors from editors of respiratory, sleep, and critical care journals. *Crit Care Med* 48: 623–633, 2020 https://doi.org/10.1097/CCM.0000000000004246

9. Mundry R, Nunn CL: Stepwise model fitting and statistical inference: Turning noise into signal pollution. *Am Nat* 173: 119–123, 2009 https://doi.org/10.1086/593303

10. Collins GS, Reitsma JB, Altman DG, Moons KG: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD statement. *Ann Intern Med* 162: 55–63, 2015 https://doi.org/10.7326/M14-0697

11. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW: Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* 21: 128–138, 2010 https://doi.org/10.1097/EDE.0b013e3181c30fb2

12. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, Levin A, Levey AS: A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 305: 1553–1559, 2011 https://doi.org/10.1001/jama.2011.451