EDITORIAL

AKI biomarkers are poor discriminants for subsequent need for renal replacement therapy, but do not disqualify them yet

Harm-Jan de Grooth¹, Jean-Jacques Parienti^{2,3} and Miet Schetz^{4*}

© 2018 Springer-Verlag GmbH Germany, part of Springer Nature and ESICM

An important controversy and uncertainty still exists around the optimal timing to start renal replacement therapy (RRT) in ICU patients with acute kidney injury (AKI) that do not present with absolute indications for RRT. Recent RCT's comparing “early” and “late” initiation of RRT provided contradictory messages, potentially related to the difference in study population, design (single- versus multicenter), illness severity, modality of RRT and definitions of early and late [1–3]. Reliable quantitative tools that predict whether AKI patients will recover would allow avoiding unnecessary RRT. Moreover, by identifying a more severe subset of patients with AKI, an evaluation of the advantages and disadvantages of early starting of RRT could be performed in an enriched population that would probably need RRT sooner or later during their ICU stay. Biomarkers of kidney damage have been extensively studied for their accuracy in early diagnosis of AKI. Their role in distinguishing patients with high versus low likelihood of renal recovery is less clear.

In a recent article in this journal, Klein et al. report the results of a systematic review which meta-analyzed the ability of biomarkers to identify patients with and without subsequent need for RRT [4]. The largest body of evidence exists for NGAL that showed only fair discrimination. The markers of cell cycle arrest, the product of [TIMP-2]*[IGFBP-7], showed the best discrimination with a pooled area under the ROC curve (AUROC) of 0.857 (0.789–0.925). This result (based on four studies with 280

patients and 50 RRT received) was rather homogeneous ($I^2 = 24.9\%$), increasing the confidence of this estimate. On the other hand, the homogeneity of this result probably relates to the more homogeneous patient populations with clearly defined insult (mainly surgery) in the included studies whereas studies on NGAL included more heterogeneous populations where the timing of the renal insult is less clear. A summary of the most important findings can be found in the Supplementary Table (Table S).

This enormous amount of work, for which the authors are to be congratulated, cannot overcome the limitations of the included original studies. The most important limitation is the absence of a gold standard for the endpoint (RRT): substantial practice variation exists in the initiation of RRT and many of the included studies did not use predefined indications. In 85% of the studies, the prediction of RRT was not the main purpose. The study populations are heterogeneous and not all patients had AKI at the time of biomarker sampling, which could explain the large variability of RRT need (Table S). The delay between sampling of the biomarker and the start of RRT, when reported, was variable, increasing the observed between-study heterogeneity and biomarker performance (Fig. 1). Biomarkers were not compared with readily available clinical parameters nor was their additive value assessed, for example with the net reclassification index.

Finally, we share the authors' interpretation that the use of biomarkers to discriminate patients requiring RRT has currently no implication in routine clinical decision-making at the bedside because the benefit of early RRT initiation remains to be demonstrated, and the predictive accuracy of biomarkers is at best fair. Such discrimination is particularly useful when the disease state is already determined but not directly observable (for example the

*Correspondence: marie.schetz@uzleuven.be

⁴ Clinical Department and Laboratory of Intensive Care Medicine, Division of Cellular and Molecular Medicine, KU Leuven University, Herestraat 49, 3000 Leuven, Belgium

Full author information is available at the end of the article

use of d-dimer for a pulmonary embolism diagnostic). However, an important purpose of AKI biomarkers is to predict future states that are not yet determined, such as the future risk of AKI or the need for RRT. AUCs are of limited value for this type of risk stratification under stochastic uncertainty. Differences in population risk distribution may contribute to an increase between study heterogeneity regarding discrimination (Fig. 1).

Discrimination vs. risk stratification

When trying to guess a future event, two statistical predictions can be applied: “black or white” (discrimination)

versus “several shades of gray” (risk stratification). This dichotomy can be illustrated with a simple example. Suppose that among 1000 patients, 100 have a 10% chance of needing RRT, 800 have a 50% chance and 100 have a 90% chance. Also, suppose that a biomarker perfectly predicts RRT need: It reads ‘10’, ‘50’ and ‘90’ for the patients with 10, 50 and 90% risk, respectively. This biomarker has obvious clinical potential, as it can be used to stratify low-, intermediate- and high-risk patients with great fidelity. Yet the calculated AUC of this biomarker in this population is only 0.60, which would be commonly interpreted as ‘poor’ (0.60–0.70) discrimination. The

counterintuitive disparity between excellent risk prediction and poor discrimination is a well-proven phenomenon [5–7]. In a population with normally distributed risk characteristics, any biomarker that is perfectly calibrated for risk stratification will be a poor discriminant. Conversely, a biomarker can achieve perfect discrimination only when the predicted risks are 0 and 100%—i.e., when the outcomes are already determined but not yet known (Fig. 1).

In spite of these limitations, most AKI biomarker studies report discriminant statistics. Consequently, Klein et al. had to rely on AUCs as the common denominator to compare the performance of biomarkers for prediction of the need for RRT [4]. This has some validity, as the future need for RRT is contingent on renal failure that may already be present at the time of biomarker sampling.

How biomarkers may help identify patients that can benefit from treatment

AKI biomarkers are only useful if they can be proven to influence treatment decisions that result in patient-oriented benefit [8, 9]. In the PrevAKI study [10] cardiac surgery patients at increased risk of AKI were identified by a product of [TIMP-2]·[IGFBP-7] above 0.3. Although a renal protective treatment bundle was shown to reduce the incidence of AKI, the protective effect was not observed in the subgroup of patients with the highest [TIMP-2]·[IGFBP7] (above 2.0). This underscores the importance of risk stratification over discrimination: outcomes in patients with very high or very low risk may not be amenable to therapy. It is often intermediate risk patients in the “grey zone” that benefit from protective interventions, as is now widely recognized in sepsis research [11–13]. Generating and validating biomarker-based risk predictions using logistic regression models and decision-tree models are more suitable than AUCs for this purpose [5, 11, 14].

Readers should bear in mind that poor discriminatory performance does not necessarily disqualify these biomarkers as risk predictors. In the future, AKI biomarker studies should complement AUCs with risk prediction parameters, such as odds ratio, relative risks or hazard ratios. Although sensitivity and specificity are familiar to most clinicians and appear readily interpretable, their clinical applications may be limited. Instead, early identification of patients that can benefit from protective therapies requires more focus on risk stratification.

Electronic supplementary material

The online version of this article (<https://doi.org/10.1007/s00134-018-5151-7>) contains supplementary material, which is available to authorized users.

Author details

¹ Department of Intensive Care, VU University Medical Center, De Boelelaan 1118, 1081 HZ Amsterdam, The Netherlands. ² Department of Biostatistics

and Clinical Research, Centre Hospitalo-Universitaire de Caen, Caen, France. ³ EA2656 Groupe de Recherche sur l'Adaptation Microbienne, Université Caen Normandie, Caen, France. ⁴ Clinical Department and Laboratory of Intensive Care Medicine, Division of Cellular and Molecular Medicine, KU Leuven University, Herestraat 49, 3000 Leuven, Belgium.

Compliance with ethical standards

Conflicts of interest

The authors have no conflicts of interest to disclose.

Received: 19 March 2018 Accepted: 22 March 2018

Published online: 12 April 2018

References

- Gaudry S, Hajage D, Schortgen F, Martin-Lefevre L, Pons B, Boulet E, Boyer A, Chevrel G, Lerolle N, Carpentier D, de Prost N, Lautrette A, Bretagnol A, Mayaux J, Nseir S, Megarbane B, Thirion M, Forel JM, Maizel J, Yonis H, Markowicz P, Thiery G, Tubach F, Ricard JD, Dreyfuss D, AKIKI Study Group (2016) Initiation strategies for renal-replacement therapy in the intensive care unit. *N Engl J Med* 375:122–133
- Zarbock A, Kellum JA, Schmidt C, Van Aken H, Wempe C, Pavenstädt H, Boanta A, Gerß J, Meersch M (2016) Effect of early vs delayed initiation of renal replacement therapy on mortality in critically ill patients with acute kidney injury: The ELAIN randomized clinical trial. *JAMA* 315:2190–2199
- Bagshaw SM, Lamontagne F, Joannidis M, Wald R (2016) When to start renal replacement therapy in critically ill patients with acute kidney injury: comment on AKIKI and ELAIN. *Crit Care* 20:245. <https://doi.org/10.1186/s13054-016-1424-0>
- Klein SJ, Brandtner AK, Ulmer H et al (2018) Biomarkers for prediction of renal replacement therapy in acute kidney injury: a systematic review and meta-analysis. *Intensive Care Med* 44(3):323–336. <https://doi.org/10.1007/s00134-018-5126-8>
- Cook NR (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115:928–935. <https://doi.org/10.1161/CIRCULATIONAHA.106.672402>
- Gail MH, Pfeiffer RM (2005) On criteria for evaluating models of absolute risk. *Biostatistics* 6:227–239. <https://doi.org/10.1093/biostatistics/kxi005>
- Diamond GA (1992) What price perfection? Calibration and discrimination of clinical prediction models. *J Clin Epidemiol* 45:85–89. [https://doi.org/10.1016/0895-4356\(92\)90192-P](https://doi.org/10.1016/0895-4356(92)90192-P)
- Prowle JR (2015) Measurement of AKI biomarkers in the ICU: still striving for appropriate clinical indications. *Intensive Care Med* 41:541–543. <https://doi.org/10.1007/s00134-015-3662-z>
- Prowle JR, Schetz M (2017) Validating benefit of biomarker-directed therapy for acute kidney injury: can you have your cake and eat it? *Intensive Care Med* 43:578–580. <https://doi.org/10.1007/s00134-017-4710-7>
- Meersch M, Schmidt C, Hoffmeier A et al (2017) Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. *Intensive Care Med* 43:1551–1561. <https://doi.org/10.1007/s00134-016-4670-3>
- Wong HR, Lindsell CJ, Pettilä V et al (2014) A multibiomarker-based outcome risk stratification model for adult septic shock. *Crit Care Med* 42:781–789. <https://doi.org/10.1097/CCM.0000000000000106>
- Cohen J, Guyatt G, Bernard GR et al (2001) New strategies for clinical trials in patients with sepsis and septic shock. *Crit Care Med* 29:880–886
- Marshall JC (2008) Sepsis: rethinking the approach to clinical research. *J Leukoc Biol* 83:471–482. <https://doi.org/10.1189/jlb.0607380>
- Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S (1997) A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 16:965–980