

**Article:**

Lama Ghazi, Kassem Farhat, Melanie P Hoenig, Thomas J S Durant, Joe M El-Khoury.

Biomarkers vs Machines: The Race to Predict Acute Kidney Injury

Clin Chem 2024; 70(6): 805–819. <https://doi.org/10.1093/clinchem/hvad217>

Guests: Dr. Joe El-Khoury from the Yale School of Medicine and Yale-New Haven Health and Dr. Lama Ghazi from the University of Alabama at Birmingham.

Bob Barrett:

This is a podcast from *Clinical Chemistry*, a production of the Association for Diagnostics & Laboratory Medicine. I'm Bob Barrett.

Acute kidney injury, or AKI, is a substantial health concern affecting an estimated 50% of patients in the intensive care unit. AKI is associated with high morbidity and mortality but fortunately, poor outcomes are preventable in some cases if recognized at an early stage. Current diagnostic criteria rely on changes in serum creatinine or urine output, but both metrics have fundamental flaws that hinder early diagnosis and effective intervention. Recognizing these limitations, substantial efforts are underway to find a better alternative. Proteins released by damaged kidneys have been studied for years, but are any primed to take a central role in AKI diagnosis? Others have focused on developing machine learning algorithms to predict AKI using information in the patient's medical record. How accurate are these models, and are they likely to positively impact patient care? A review article appearing in the June 2024 issue of *Clinical Chemistry* provides an update on novel methods to predict AKI and asks which will get to the finish line first, the biomarkers or the machines?

In this podcast, we welcome the article's lead and senior authors. Dr. Joe El-Khoury is an Associate Professor of Laboratory Medicine at Yale School of Medicine and Director of the Clinical Chemistry Laboratory and Fellowship Program at Yale New Haven Health. Dr. Lama Ghazi is an Assistant Professor of Epidemiology at the University of Alabama at Birmingham, and her research interests lie at the intersection of chronic disease prevention, health disparities, and digital health. And I've got a first question here I'd like you both to tackle. What inspired the shift from serum creatinine and urine output to exploring alternative biomarkers and AI in AKI diagnosis?

Joe El-Khoury:

So, serum creatinine and urine output have been used for decades in this space and as mentioned, they have their limitations. So, urine output is a problem because it's impractical and requires the insertion of a Foley catheter to

be done. So, it's quite invasive and it's done mainly for patients who are already in the hospital as inpatients, but not for all procedures, so, it's not as widely available to monitor the development of acute kidney injury. While serum creatinine, the complaints have been that it's too slow to rise after an AKI event. Usually, studies have reported taking 24 to 48 hours after an AKI before serum creatinine goes up and is able to help with the diagnosis.

So, those have been really the big complaints with these two. And if we look at myocardial infarction, for example, as a corollary, creatinine has been around since the 1950s, while for MI, myocardial infarctions, we went from using, believe it or not, AST at the time in the 1950s, to LDH, to CKMB, to troponin, to high-sensitivity troponin now, while the whole time we've been using creatinine in the AKI space. So, we're really on a hunt for a kidney troponin but so far, it's been elusive. So that really covers it for biomarkers. And I'll leave it to Dr. Ghazi to discuss what's happening in the AI space and why that's also evolving.

Lama Ghazi:

Thank you, Dr. El-Khoury. So, yeah, I think everyone has heard about big data, AI, machine learning. Those are, like, pretty popular buzzwords. They do have a space in AKI research and pretty much with the digitalization of medical records and the ability to sort of get data from different places and not just labs. That is the beauty of it is that you can use this data to predict in real time even acute kidney injury. So being able to do that, researchers now are using all the data available to optimize the way you either detect AKI or prevent AKI. And that's ideally what you need to do, is sort of identify those patients who might develop AKI and try to prevent it early on.

For example, if you know a patient is likely to have AKI during their hospital stay at day two, and you're like, okay, there's a flag that's like, oh, the patient might have AKI in the next two days, then you're less likely, or the care team is less likely, to prescribe nephrotoxins, pay attention more to the volume overload or volume load, and things like that. So, this is why there has been a shift toward moving, to using sort of machine learning and AI.

Bob Barrett:

Dr. El-Khoury, how do the proposed expansions to the KDIGO definition and the introduction of structural biomarkers contribute to early AKI detection and patient outcomes?

Joe El-Khoury:

Thanks, Bob. So, to kind of answer that question, we need to first recognize what are the limitations of the current KDIGO definition, which originally came out in 2012. At the time, it was important and great because it standardized three other definitions that were very variable and were causing confusion in the space.

Unfortunately, since that time, you've identified, and there are studies that have shown repeatedly, that it has a high false positive rate in patients with high creatinine. Basically, if you're using the 0.3 mg per deciliter change, you have over 30% chance of false positive results in patients with creatinine over 1.5 mg per deciliter. So patients with high creatinine are having very high rates of false positive by that definition. And on the flip side, patients with low creatinine were also having issues of false negatives because a 0.3 change is very large for somebody whose creatinine is 0.3 or 0.4 to begin with, which is particularly affecting children and older adults, especially older women, who have lower creatinine values. So those are the limitations of that definition.

And since that time, studies and groups have come out trying to figure out better definitions using creatinine first. And one of them is pROCK, which specifically targeted the pediatric population, and the other is the AACC AKI, which looked at both adults and pediatric, and both used the similar approach, which uses the concept of reference change value. It's a complicated kind of nomenclature, but really what it means is "what is a significant change in creatinine?" If I know what are my analytical variation on the instrument and what are the biological variations that happen in healthy people. And both groups tracked those and even linked them to outcomes and came up with newer, better definitions for creatinine to use for AKI.

The other changes that we were seeing is people were talking about using kind of like we do for troponin, a rise and fall in creatine, not just the rise, which is what we're currently doing, so that's another expansion. And the third and more significant expansion is adding damage criteria and monitoring those. And that's important in the context of speaking about NGAL or TIMP-2.IGFBP7, which are these structural biomarkers. How does that information is included in what we know about AKI today? So KDIGO right now, since you have stage 1, stage 2, stage 3, based on creatinine, with the new damage criteria, you basically now have stage 1A and stage 1B, stage 2A, stage 2B, where the A and B represent structural damage, present or absent. So this gives additional information, and with the goal of helping identify if this damage is due to intrinsic AKI, so are we seeing a problem in the kidney itself that's causing the acute kidney injury, or is it a prerenal issue, where you may not have structural kidney damage, but you have functional loss because of prerenal issues? And so that's the type of information that would also be helpful in managing and identifying the cause that would be different than what we're doing today.

There are problems, of course, with the markers. They're not perfect. And it's important for me to say they're not troponin to the kidney. So very important for labs looking to adopt them to recognize that there's a lot of false positives associated with those as well. And I'll just use an example for TIMP-2.IGFBP7, where even in the FDA validation studies, they said 50% of healthy people will show up positive beyond the first cutoff that they use which is 0.3. That's not great. So that's why you have to be careful where you're implementing it and the population you're using and making sure they're being evaluated there. Overall, the areas under the curve for their performance for both NGAL and TIMP-2.IGFBP7 are really good. They stand around 0.8 on average. But it's really important to recognize that they haven't really shown great improvements in patient outcomes when you look at the bigger picture. So more studies are needed there.

This is why there are many organizations on the laboratory side and like the National Institute of Health in the UK, and basically, they don't currently recommend using those tests. So that's the current issue. And I would really encourage laboratories who are looking at these to be proactive and review the literature to your specific population as you're trying to apply these tests.

Bob Barrett: Okay, thanks for that. Now, Dr. Ghazi, what has machine learning contributed so far?

Lama Ghazi: Unlike, well, more biomarkers, their machine learning is still in its infancy sort of for AKI. There are quite a few predictive models that have been published specifically in ICU populations. So, they're very narrow to which population they serve. There are quite a few models that can predict acute kidney injury development. They can predict outcomes as well, or even development of AKI, but those have been limited. So even though it is exciting that machine learning is on the horizon, and it does quite a good job in some models of identifying who has AKI, again, those are in specific populations such as the ICU. It's still not there yet. You do see in the literature, some AUCs like Dr. El-Khoury mentioned for some biomarkers of predicting AKI that can go up to 0.9. So, what that means, it's the likelihood that a patient with AKI will be correctly identified over someone who doesn't have it. So, we do need to use other measures. But, long story short is that, that there's a lot going on in the machine learning world about identifying AKI.

And I think what's exciting actually there is, there have been studies where people have taken those models and integrated them into the clinical setting and care to see if they make changes in the provider's behavior. So some models, for example, went from just developing a model, but not just developing the model, but you deploy the model in a clinical

setting and see if patient outcomes improve. Unfortunately, those trials so far have been negative. They didn't really make changes in the provider sort of behavior if they knew what the risk of a patient's AKI is. So maybe more actionable models are better. So, if you actually tell the provider, like, do this or do that, those are more likely to work. I won't dismiss the model or the machine learning models. I would say those are like stepping stones and they provide the groundwork. You do need to know your limitations to know what the next step is. So, I would say in the future, those models should be more dynamic. You should actually test them in the real world setting to see what you can use and maybe move forward beyond just ICU or critical care settings toward more general populations. And we do touch base on that as well, later on.

Bob Barrett: Okay, now, in this review, you also offer some recommended future directions. Could you both provide a summary for our audience?

Joe El-Khoury: Absolutely, Bob, and I'm happy to leave the recommendations on the biomarker space, and I'll leave it to Dr. Ghazi to discuss the AI recommendations. But basically, the main issue with studies we've reviewed today is heterogeneity and lack of clear reporting of what they exactly did. So the recommendations for biomarker studies are: you have to list detailed assay information. For example, NGAL from one vendor is not the same as an NGAL from another vendor. So you have to specify in your reports which NGAL assay you've used and sourced from which specific company, for example.

Also, AKI definitions used. As we've mentioned, there are now improved and alternate definitions out there. It's no longer just KDIGO. pROCK, AACC-AKI, linked to even patient outcomes. So as studies come out, it is worth exploring which definition do you want to use, and how has that impacted your analysis? Another thing we talked about is the choice of population. Dr. Ghazi mentioned the importance of going out of the ICU, and if you do that, the test will perform differently because there's different incidences of AKI in these different settings, and that can impact performance. So AUCs alone are not the, be all, you have to look at how does that AUC reflect it and what was the incidence of API in that population? And then other things to focus on are choice of the sample collection time relative to the AKI event, choice of the cutoffs used for those assays, choice of the prediction window. Are you trying to predict AKI happening in 12, 24, or 48 hours, and we've seen that heterogeneity again in studies. And finally, if risk stratification is used, because some for like TIMP-2.IGFBP7, which is again commonly called Nephrocheck, some have recommended only testing these in patients who are at high risk for AKI. And that is theoretically

improve the performance of the test. And that's true because I mentioned it has a very high false positive rates in healthy people. One way you do that is you risk stratify and that improves the performance. But you have to be clear about what are you doing step by step to classify these patients and that's important to also outline in the report, if done. And I'll turn it over to Dr. Ghazi for the AI side.

Lama Ghazi:

Well, a lot of it, I'll just piggyback on what you said. A lot of it is common, whether it's on the AI side or the biomarker side. We do need more structured definitions of what AKI is, with more clarity, as well as more transparency in reporting, whether it's your exposures, your outcome, what models you used. You need to stay away from that black box of "I used machine learning model and it spit out those results." So we do strongly recommend that researchers be as transparent as possible when they are writing their articles or when they're reporting their results. So you need to document which strategy has been used for various things such as missing data, whether imputation was used.

We recommend that more fairness in representation of the population should be there, because unfortunately, this is the case in all medical literature, well most of it, is that we underrepresent certain minority populations. So this should be taken into consideration when you're developing models, sort of its generalizability as well. We also recommend that maybe even using unstructured data, or there is some untapped potential in the amount and the type of data that is available in the medical records. That can be anything from physician's note, imaging, and that might improve sort of model performance. We highly recommend dynamic models be used. I think Dr. El-Khoury also alluded to that, in which when you are planning to predict AKI is important. So those dynamic models not just predict AKI on an up-to-date basis, but they also change as the patient's status changes.

We also emphasize the importance of rigorous validation of these models. It is great if you have a model that works in your own setting, but what if you take it away from sort of your institution to a different setting?

So, again, testing those models and applicability of those models outside a controlled setting is very important. So I do echo Dr. El-Khoury as well with saying that AUC is not the best way, or the recommended way, to just assess the performance of a model, but we should be thinking about other ways to assess model performance beyond that. So, yeah, it's an exciting field, and I think there's a lot of future research opportunities to make sure that those machine learning models and AI models improve on the accuracy, applicability, and the impact they have on clinical care.

Bob Barrett: Well, Dr. El-Khoury, finally, who's winning this race to predict acute kidney injury? Is it the biomarkers, or is it the machines?

Joe El-Khoury: I mean, that's the million-dollar question, right, Bob? So, given the limitations that we've discussed today, I think it's fair to say that the race is closely balanced between both, but both are really stuck in the mud right now. We really need more prospective studies that focus on the impact of these diagnostic tools on patient outcomes and we need them to more clearly report exactly what they did. Ultimately, this race is not about biomarkers versus machines. I see it as a relay race, and both are on the same side, racing against AKI. But to get out of the mud and beat it, we really need to better design our studies and more clearly report our findings. This was really the take home message for us when our team sat down and reviewed the literature and prepared this review. I would also therefore, like to take this moment to thank our co-authors, Kassem Farhat, Drs. Melanie Hoenig and Tom Durant, who also helped in a big way with this review. So, thank you, Bob, for taking the time, and obviously -- always a pleasure to be with Dr. Ghazi discussing this topic.

Bob Barrett: That was Dr. Joe El-Khoury from the Yale School of Medicine and Dr. Lama Ghazi from the University of Alabama at Birmingham. They co-authored a review article on the use of biomarkers and machine learning algorithms to predict acute kidney injury in the June 2024 issue of *Clinical Chemistry*, and they have been our guests in this podcast on that topic. I'm Bob Barrett. Thanks for listening.