Leveraging Big Data and Electronic Health Records to Enhance Novel Approaches to Acute Kidney Injury Research and Care

Scott M. Sutherland a, Stuart L. Goldstein b, Sean M. Bagshaw c

a Department of Pediatrics, Division of Nephrology, Stanford University, Stanford, CA; b Center for Acute Care Nephrology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, USA; c Department of Critical Care Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

Keywords
Acute kidney injury · Electronic health record · Electronic medical record · Big data

Abstract
While acute kidney injury (AKI) has been poorly defined historically, a decade of effort has culminated in a standardized, consensus definition. In parallel, electronic health records (EHRs) have been adopted with greater regularity, clinical informatics approaches have been refined, and the field of EHR-enabled care improvement and research has burgeoned. Although both fields have matured in isolation, uniting the 2 has the capacity to redefine AKI-related care and research. This article describes how the application of a consistent AKI definition to the EHR dataset can accurately and rapidly diagnose and identify AKI events. Furthermore, this electronic, automated diagnostic strategy creates the opportunity to develop predictive approaches, optimize AKI alerts, and trace AKI events across institutions, care platforms, and administrative datasets.

Introduction
Acute kidney injury (AKI), which is defined as an abrupt decline in kidney function, is a common complication encountered among hospitalized adults and children [1, 2]. Over the past decade, the critical care and nephrology communities have made great strides in refining and standardizing AKI definitions and care; these efforts culminated in formulating the Kidney Disease: Improving Global Outcomes (KDIGO) AKI clinical practice guidelines [3]. These guidelines contain the first fully harmonized AKI definition based upon relative changes in serum creatinine and urine output (UOP). With this development, AKI can now be identified in a uniform, consistent manner across all care environments and information technology platforms.

In parallel with the development of these standard AKI guidelines, we have seen a marked growth in electronic health record (EHR) adoption and impressive advances in clinical informatics techniques [4–6]. While these occurrences are far from being causally related, the now ubiquitous presence of EHRs, when combined with novel “big data” informatics approaches, create unique opportunities for AKI research and care improvement.

The data contained within the EHR is “big” from the standpoint of volume (amount of data present), velocity (speed at which new data is generated), and variety (number of different types of data) [7, 8]. To put this in the context of AKI, the EHR contains every creatinine and documented UOP value for each patient across the entire institution, accumulates new creatinine and UOP data as they become available, and contains a myriad of AKI-relevant data elements generated through the routine provision of clinical care. Within this EHR environment, it is...
possible to apply the KDIGO definition to accurately and efficiently identify AKI events, thereby anchoring them in time. AKI can then be explored prospectively, investigated retrospectively, and studied for quality assurance/improvement or research purposes; AKI can be evaluated by an individual practitioner, the institution itself, or the medical community at large. The goal of this manuscript is to highlight the unique opportunities that EHR data and informatics techniques create within the AKI domain.

AKI Diagnosis: The Fulcrum of EHR-Enabled Approaches

The key to EHR-enabled AKI research, prediction, tracking, and care improvement is accurate and timely diagnosis. The ability to precisely identify AKI is the core aspect of all such approaches (Fig. 1). This underscores the importance of utilizing and adhering to the KDIGO AKI criteria when diagnosing and identifying AKI within EHR platforms [3, 9]. The KDIGO AKI definition contains both serum creatinine and UOP criteria (Table 1) and, ideally, all definitional criteria should be applied [9].

Within the EHR, applying the creatinine criteria is straightforward. As each creatinine becomes available, it can be compared to the patient’s prior creatinines and AKI can be diagnosed when the relative (≥1.5 × increase) or absolute (≥0.3 mg/dL [27 μmol/L]) change threshold is met. Creatinine values are discrete elements and accrue with an associated date/time, allowing full application of the temporal components (7 days for relative changes and 48 h for absolute changes). One definitional aspect that is challenging, however, is baseline creatinine determination; accurate assessment of baseline is of paramount importance as it is the basis of all relative changes used to diagnose and stage AKI. Many have addressed this issue and while there is yet no consensus in this matter, several approaches have been utilized [9–14]. If creatinine values are available, one may use a pre-hospitalization creatinine value as the baseline. Most studies utilizing this approach have chosen the lowest creatinine obtained during the 3–6 months prior to hospitalization; other than in neonates and small infants, creatinine is unlikely to change physiologically in that timeframe. However, often no creatinine values are available within the EHR prior to admission. Patients may receive their ambulatory care in a different health system or may have never had

Fig. 1. All EHR-enabled approaches to AKI care improvement and research hinge on the ability to accurately diagnose AKI. Once this is accomplished, it becomes possible to predict and identify events, alert practitioners when their patients develop AKI, standardize and improve the quality of care provided in the setting of AKI, and track patients/events across institutions, populations, datasets, and care platforms.
a creatinine obtained prior to hospitalization. In these cases, one option is to define the admission creatinine as the baseline. This is simple and effective; however, it can miss cases of AKI, which are present upon admission, underestimating the burden by as much as one-third [12]. The second option in this scenario is to estimate a baseline creatinine by back-calculating from an assumed creatinine clearance (CrCl); adult and pediatric studies have tended to presume the CrCl to be 75 and 100–120 mL/min/1.73 m², respectively [10, 12, 13]. This approach will capture AKI events present on admission; however, it fails to account for patients with chronic kidney disease (CKD), misclassifying them as AKI. In adult patients, this approach can overestimate the incidence of AKI by 50% [12]. Additionally, this method requires computation, adding complexity to any attempt to automate an AKI diagnostic tool. Estimating equations also require data elements that are not reliably captured in the EHR such as ethnicity and height; the absence of this data could render the tool useless.

The UOP criteria pose a second definitional challenge within the EHR environment. The most substantial problem is that UOP is not obtained with the same rigor as creatinine. Few hospitalized patients outside of intensive care units (ICUs) have indwelling urinary catheters capable of providing hourly data; patients, especially those outside of ICUs, may not have UOP recorded for hours at a time. Furthermore, in children, UOP data is often documented only as a void count, without volumetric data. Given the relatively short diagnostic timeframe set by the KDIGO UOP criteria, many patients with normal renal function would be diagnosed with AKI. Second, the EHR itself can pose technical challenges. Most EHRs sum intake and output data at static intervals; usually these intervals coincide with 8 or 12-h hospital shifts. The UOP criteria, however, require a dynamic calculation with a rolling 6–24-h window. Although not impossible, running such a rolling sum for all patients in the hospital may pose logistic, resource, and computational challenges.

With these issues in mind, given the fact that data regarding automated AKI identification are scant and that no truly comprehensive method has been developed, many have advocated an incremental approach [9]. It is not possible to anticipate all issues, and initial experience with these tools will enhance refinement. Thus, with the goal of creating an effective and efficient automated AKI diagnostic tool, 2 compromises might be made. The first would be to focus initially on integrating the creatinine criteria with the expectation of adding the UOP criteria in future iterations; this will allow earlier and simpler adoption and will begin to build a dataset that we can use to improve such tools. It is important to note, however, a recent international multicenter pediatric study demonstrated that 2/3 of patients who met AKI by UOP criteria would have been missed by the creatinine criteria alone; additionally, these UOP-diagnosed AKI events carried an increased risk for mortality [15]. This underscores the need to work toward the integration of the full KDIGO criteria. With the same concept in mind, the second com-

### Table 1. KDIGO AKI criteria

<table>
<thead>
<tr>
<th>Definition of AKI</th>
<th>Staging/grading of AKI severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Increase in SCr by ≥0.3 mg/dL (≥26.5 μmol/L) within 48 h; or</td>
<td>Stage 1: Increase in SCr by 1.5–1.9 times baseline; or</td>
</tr>
<tr>
<td>– Increase in SCr to ≥1.5 times baseline, which is known or presumed to have occurred within the prior 7 days; or</td>
<td>Increase in SCr by ≥0.3 mg/dL (≥26.5 μmol/L); or</td>
</tr>
<tr>
<td>– Urine volume &lt;0.5 mL/kg/h for 6 h</td>
<td>Urine output &lt;0.5 mL/kg/h for 6–12 h</td>
</tr>
<tr>
<td></td>
<td>Stage 2: Increase in SCr by 2.0–2.9 times baseline; or</td>
</tr>
<tr>
<td></td>
<td>Urine output &lt;0.5 mL/kg/h for ≥12 h</td>
</tr>
<tr>
<td></td>
<td>Stage 3: Increase in SCr by 3.0 times baseline; or</td>
</tr>
<tr>
<td></td>
<td>Increase in SCr to 4.0 mg/dL (353.6 μmol/L); or</td>
</tr>
<tr>
<td></td>
<td>Initiation of renal replacement therapy; or</td>
</tr>
<tr>
<td></td>
<td>In patients &lt;18 years, decrease in eGFR to 35 mL/min/1.73 m²; or</td>
</tr>
<tr>
<td></td>
<td>Urine output &lt;0.3 mL/kg/h for ≥24 h; or</td>
</tr>
<tr>
<td></td>
<td>Anuria for ≥12 h</td>
</tr>
</tbody>
</table>
promise might be to initially use the lowest creatinine within the 3 months prior to admission, inclusive of the day of admission, as the baseline. Although this would not be a flawless approach and would likely miss some patients with AKI whose severity peaks on admission, it would reduce the complexity of the initial identification tool by not requiring a computational creatinine estimation and would effectively handle patients with CKD. In situations where it becomes apparent that the admission creatinine did not represent the patient’s baseline (i.e., the creatinine rapidly declines in the first several hospital days), it would be possible to retroactively enter a more accurate baseline; although this would not help in real-time, it would help more accurately define disease burden and would be beneficial when machine learning techniques were applied predictively. Eventually, however, through an iterative process, an optimized tool could be developed, which utilized creatinine changes and UOP changes when available, and when baseline data was not available, used the admission creatinine as a baseline for patients with CKD and a back-calculated creatinine for non-CKD admissions.

**AKI Prediction: The Pre-Disease State**

Once AKI can be reliably and accurately diagnosed, all EHR-enabled approaches naturally unfold around that event. Prior to discussing techniques that are subsequent to the development of AKI, it is important to examine the antecedent time frame. The temporal anchoring of AKI within the EHR clearly delineates a pre-disease state that contains all the EHR data which accumulate prior to the development of AKI. High-content, high-throughput predictive analytic techniques can be applied to this colossal dataset in order to identify a pre-AKI signal (Fig. 2). This signal can be used to discriminate between patients who are likely or unlikely to develop AKI, allowing patients to be categorized by AKI risk. The ability to predict AKI risk has profound implications since currently no treatments exist for AKI once it develops [16–18]. When patients at high AKI risk are identified, it becomes possible to modify care and implement preventative and harm avoidance strategies among these patients [19–24].

A comprehensive appraisal of EHR-enabled AKI prediction is beyond the scope of this review; however, several groups have evaluated such approaches and the technique was the subject of a recent Acute Dialysis Quality Initiative (ADQI) consensus conference [8, 25–27]. The goal of this review, therefore, is to highlight several aspects related to AKI prediction and risk stratification. The first concept worthy of discussion is the prediction target. While it may initially seem obvious that the models should predict “AKI,” upon reflection, a more nuanced approach might be needed. Across many patient populations, AKI has been associated with poorer outcomes, regardless of severity. However, the strength of those associations is greater for stages 2 and 3 AKI and in some populations, milder AKI has not always portended a worse prognosis [1, 2, 15, 28–36]. Given these parameters, an incremental approach might again be optimal. One option would be to begin by building a model to predict moderate/severe (stage 2/3) AKI; this would likely generate a more robust initial model as associations between predictors and AKI would be greater and it would eliminate some of the “noise” associated with stage 1 AKI [8]. However, given that stage 1 AKI is indeed associated with worse outcomes in some populations, subsequent iterations would aim to predict AKI of all severities. An initial model aimed at stage 2/3 AKI would also have the advantage of allowing prediction of AKI relevant outcomes in addition to AKI events. Given the greater associational strength between moderate and severe AKI and outcomes, the prediction model could identify not only patients at high risk for AKI development but additionally those who are likely to experience poorer AKI-related outcomes. This would add additional relevance to the model and provide practitioners with more useful clinical information.

The second aspect of AKI prediction that merits discussion is the technique utilized. To date, a majority of AKI prediction models have used a supervised approach; even among massive datasets, the potential predictors...
have been chosen a priori based upon their association with AKI in previous studies [37–41]. While this is an excellent initial approach, it does not take full advantage of big data methodologies. Unsupervised or machine-learning techniques allow predictors to be identified without prior prejudice; this allows the identification of predictors that may be both novel and strong. Using a dynamic technique such as random forest modeling, cluster analysis, or principal component analysis, although it is a departure from classic model building approaches, would optimize the use of EHR data [8].

AKI Alerting: The Disease State

The ability to accurately and efficiently diagnose AKI in real time also paves the way for automated notification systems and alerting. Essentially, an AKI alert notifies care providers immediately when their patient meets the diagnostic criteria for AKI. In theory, the knowledge that their patient has AKI would allow practitioners to modify or optimize care with the goal of eliminating ongoing injury, preventing further disease progression and mitigating AKI sequelae. These alerts would seem at first glance to be straightforward effective techniques to improve AKI-related care and outcomes; however, in practice, the concepts that surround alerting are complex and subtle choices in implementation can have profound implications.

This is perhaps best exemplified by 2 recently published AKI alert trials. The first trial was a pre/post intervention study of an alert within an ICU at a single center [42, 43]. The alert was based on changes in UOP and creatinine and the attending physicians were notified by Digital Enhanced Cordless Technology phones whenever a patient transitioned from one AKI stage to the next. This study found that the alert led to an increase in the frequency of therapeutic interventions among patients with AKI (28.7 vs. 7.9%, p < 0.001) [42]. During the alert phase of the study, patients with AKI were more likely to receive fluid therapy, diuretics, and vasopressor support; additionally, the time to intervention was significantly reduced by the alert [42]. Finally, patients who developed “risk”-staged AKI were more likely to experience a return to baseline renal function within 8 h [42]. The second study was a single-center, hospital-wide, randomized controlled trial (RCT) of an AKI alert [3, 44]. The covering intern, resident, or advanced practice provider (NP or PA) was notified when their patient met criteria for AKI based upon changes in serum creatinine values. Notably, in addition to identifying the patient as having AKI and requesting that the practitioner “take appropriate diagnostic and therapeutic measures,” the text of the alert informed the practitioner that the alert “does not fire for all patients with AKI [44].” This trial was essentially negative and did not find differences in maximum creatinine change, receipt of dialysis, death, or nephrotoxin administration [44].

While it may be tempting to take the negative results from this RCT as evidence that AKI alerts are ineffective, we believe this is the wrong conclusion to draw. There is plentiful evidence that AKI alerts can be effective; the ADQI working group given the task of reviewing the literature surrounding such alerts identified 11 studies demonstrating some degree of improvement with alert implantation [45]. What these 2 studies do underscore, however, is the complexity behind the alerting mechanism [46].

Some aspects of the alert are straightforward. Alerts and studies that examine them should employ the consensus definition for AKI. The ADQI consensus group felt that the full KDIGO definition should be employed, meaning that when available, the UOP criteria should be integrated in addition to the serum creatinine criteria [45]. Additionally, if the alert is meant to have a significant clinical impact, the alert must fire in real or near real time; given the rapidity with which AKI can progress, a delay in even 6–12 h could substantially reduce the utility of the alert. Also, any alert should, at a minimum, grade the severity of the AKI in the alerting process. Additional contextual details are likely to improve the performance of the alert and will help inform provider decision making; ideally alerts will strike a balance between adequate contextual information and minimal complexity. Finally, although the alert is capable of increasing AKI awareness, it must be combined with some sort of evidence-informed clinical decision support [42, 45]. This may be a care bundle, a set of institutional guidelines, or electronic recommendations and physician order support.

Other aspects of alerting are more variable and are highly dependent upon the goals and targets of the alert. The first such feature is the location of the alert. Prior to integrating an alert, it is important to determine where it will fire. Will it be active in the ICUs where AKI is common but more overt? Or would it be more effective on non-critical care wards where AKI is less ubiquitous but more commonly missed? Should it be active only in the inpatient setting or might there be a role among ambulatory patients? The second characteristic is the target of the alert – the person to whom the alert is delivered. In the aforementioned examples, the alert delivered directly to
the attending physician was more effective than the alert delivered to the resident/intern/AHP. Although it is obvious that the alert should be delivered to the primary service, it is challenging to know which member of that service is likely to derive the most value from the information. It is likely that this may differ between institutions, but it is an aspect that must be discussed prior to implementation. Another attribute is the alerting method or medium. There are a variety of ways the alert can be delivered. The studies above utilized text pages and phones. Other options include EHR alerts or reports delivered to providers at certain time intervals. Data on alerting medium are lacking, but it is clear that the choice of the alerting method will need to be tailored to the individual goals. The last aspect that is necessary to consider is the disruptiveness of the alert. Some alerts do not need to be acknowledged, whereas others require intervention and may disrupt bedside workflow. Clearly, the disruptiveness will impact physician acceptance and alert effectiveness; compliance has been found to be greater with increasing disruptiveness [46]. As stated, all of these variable aspects must be carefully identified and defined in the context of the alerting goal and target outcome. If the aim is to identify AKI cases more accurately, a successful alert might be standalone, non-disruptive, and delivered to ancillary staff using a near-real time report. However, if the goal is to reduce AKI progression or improve mortality, the AKI alert would ideally be in real time, highly disruptive, delivered to the provider most intimately involved in the patients’ care, and connected to specific, highly actionable clinical decision support. What is clear, however, is that a comprehensive implementation strategy and a robust mechanism for auditing, feedback, and improvement must be developed prior to putting any alerting tool into practice [46].

Longitudinal AKI Care and AKI Tracing: The Post-Disease State

Although AKI was once thought to be a transient, self-limited event, over the past decade, the mid- and long-term consequences of kidney injury have been well established. In a number of patient populations, AKI has been associated with increased risk for CKD, proteinuria, hypertension, cardiovascular disease, and stroke [47–51]. Despite the ongoing morbidity risk that exists well after the event itself, patients who experience AKI rarely receive the follow-up care they warrant [52]. For the most part, the lack of follow-up can be traced to a lack of provider and patient awareness regarding the occurrence of AKI and the potential long-term hazard associated with the diagnosis [53]. This lack of recognition also contributes to an inability to trace patients with AKI on institutional, administrative, and population levels [53, 54]. Although barriers exist, it is possible that EHR-enabled AKI diagnosis as described above can help address this issue (Fig. 3).

Tracking patients with AKI hinges on the ability to “tag” these patients with an AKI identifier [54]; once the diagnosis of AKI is made electronically within the EHR, any such tag must be applied reliably. A number of different types of identifiers exist; however, perhaps the most straightforward and established method would be to tag patients with the specific International Classification of Diseases Ninth/Tenth Revision (ICD-9/10) code for AKI. Traditionally, such ICD-9/10 coding has demonstrated acceptable specificity but poor sensitivity [55, 56]; this is primarily related to the aforementioned lack of AKI awareness and lack of a clear operational definition. However, as described above, application of the KDIGO AKI criteria to the EHR dataset in real time allows accurate, automatic diagnosis of AKI events; this, in turn, could routinely apply the ICD-9/10 code to the patient’s electronic chart; should a better tag be developed, the same strategy could be used to attach it electronically as well. This tagging could be performed autonomously (without any provider input) or could be subject to provider adjudication. Regardless, once this tag was applied, it suddenly becomes possible to track AKI at the patient, institution, and population level with relative ease.

At the patient level, an AKI tag could help direct patients into the appropriate follow-up clinics; there is observational evidence that seeing a nephrologist after an AKI event improves outcomes [57]. One approach would be to use the AKI tag at discharge to automatically notify primary care physicians and generate a nephrology follow-up appointment [58]. The discharge order itself could trigger the follow-up encounter and populate the patient’s discharge instructions. It would even be possible to generate ambulatory orders for serum creatinine and albumin/creatinine ratios, which would be consistent with the KDIGO AKI guidelines [3]. While it may not be feasible for every single patient with AKI to be seen by a nephrologist, a graded strategy where patients with stage 2/3 AKI were directed into follow-up clinics might be a good initial approach. Additionally, the AKI tag could increase awareness among other providers; physicians might be less likely to prescribe nephrotoxic medications or more likely to monitor kidney function in patients.
whom they know have experienced a severe episode of AKI previously. Clearly, this type of patient-level tagging is most effective in a closed healthcare organization. The possibility that subsequent care will be provided outside of the tagging institution and similar fragmentation of care is a definite barrier to this approach [54].

At the population level, accurate diagnosis and tagging of AKI events allow patients to be traced longitudinally. If patients were assigned a unique patient identifier, they could be followed across institutions and throughout administrative databases. This would have a myriad of benefits. First, accurate diagnosis and long-term tracking of patients with AKI would allow for better assessment of the healthcare burden of AKI. Currently, administrative databases rely upon physician awareness and ICD9/10 coding to flag AKI events. This leads to a lack of sensitivity and a bias toward more severe AKI events or receipt of dialysis [53]; the cost and morbidity data acquired from such databases are an inaccurate reflection of the entirety of the AKI spectrum. Second, accurate identification of patients with AKI could facilitate recruitment into clinical trials or registries, allowing for scientific advancement of AKI management strategies. Third, institutions would have the opportunity for benchmarking and quality-improvement initiatives if approaches were standardized across the disease spectrum.

**Conclusions**

The combination of a standard AKI definition, the pervasive growth of EHR adoption, and the development of novel informatics tools has created a unique set of circumstances capable of transforming AKI-related care and research. Application of the KDIGO criteria to the EHR dataset allows accurate and timely diagnosis of AKI events. This, in turn, allows us to identify patients at high AKI risk, inform practitioners about the development of AKI in real time, provide care givers with effective care strategies, and link AKI events across institutional and care system boundaries. Eventually, these patient-, institution-, and population-level techniques will lead to bet-
ter short- and long-term outcomes for patients, hospitals, and broad care systems (Fig. 3). Although potential barriers exist and many nuances must be considered, these EHR-enabled approaches have the potential to markedly improve AKI-related knowledge and care.

Acknowledgments

Dr. Sean M. Bagshaw is supported by a Canada Research Chair in Critical Care Nephrology.

References


Disclosure Statement

Dr. Sean M. Bagshaw has received grant support and fees from Baxter Healthcare Corp for consulting/speaking.

Dr. Stuart L. Goldstein has received grants and personal fees from Baxter Healthcare, grants and personal fees from Astute Medical, grants and personal fees from Bellco, grants and personal fees from AM Pharma, personal fees from Biopoint, personal fees from Akebia, personal fees from Otsuka, personal fees from Kaneka, grants from VitaFlo, grants from Mallinckrodt, grants from Alexion, grants and personal fees from La Jolla Pharmaceuticals, outside the submitted work.


