Acute kidney injury (previously known as acute renal failure) is a common complication in hospitalized patients, and its incidence has risen significantly in the past 15 yr. As a conservative estimate, roughly 17,000,000 admissions annually in the United States are complicated by AKI, resulting in more than $10 billion in costs to the health care system (4). Between 1988 and 2002, the incidence of AKI in the community increased from 61 to 500 per 100,000 population (3,5). Despite significant technical advances in therapeutics, the mortality and morbidity rates associated with AKI remain dismally high and have not appreciably improved during the past four decades. Although the serum creatinine concentration performs fairly well for estimating kidney function in patients with stable chronic kidney disease, it performs poorly in the setting of acute disease. An ideal biomarker for acute kidney injury would help clinicians and scientists diagnose the most common form of acute kidney injury in hospitalized patients, acute tubular necrosis, early and accurately and may aid to risk-stratify patients with acute kidney injury by predicting the need for renal replacement therapy, the duration of acute kidney injury, the length of stay, and mortality. Herein is reviewed the diagnostic and prognostic performance of several types of urinary biomarkers for the diagnosis and risk stratification of acute kidney injury. The major types of urinary biomarkers fall into three classes: (1) Inflammatory, (2) renal tubular proteins that are excreted into the urine after injury, and (3) surrogate markers of tubular injury. Also discussed are statistical issues in evaluating the accuracy of biomarkers as diagnostic tests. It is likely that a panel of biomarkers, rather than a single biomarker, will be needed to perform extremely well in these three situations.

Abstract

Acute kidney injury (AKI; previously known as acute renal failure) is a common complication in hospitalized patients, and its incidence has risen significantly in the past 15 yr (1–3). As a conservative estimate, roughly 17,000,000 admissions annually in the United States are complicated by AKI, resulting in more than $10 billion in costs to the health care system (4). Between 1988 and 2002, the incidence of AKI in the community increased from 61 to 500 per 100,000 population (3,5). Despite significant technical advances in therapeutics, the mortality and morbidity rates associated with AKI remain dismally high and have not appreciably improved during the past four decades (6–12).

Why Do Clinical Trials in AKI Fail?

Several agents that work remarkably in animal models (13–15) of AKI have failed in clinical trials (16–19). This disparity in the efficacy of therapies in animal and human studies has been attributed, in part, to lack of suitable laboratory methods that can be used as physiologic surrogate end points for the clinical trials. For example, the standardized use of serum cardiac enzyme concentrations and electrocardiographic criteria has facilitated rapid progress in the management of coronary insufficiency, markedly decreasing the morbidity and mortality of acute myocardial infarction. In contrast, AKI prevention and therapy studies using variables such as change in serum creatinine, urine output, and urine chemistries have not yielded interventions that are proved to decrease the morbidity (requirement for dialysis) or mortality.

Limitations of Serum Creatinine for Timely and Accurate Diagnosis of AKI

The traditional diagnosis of AKI involves a measurement of a surrogate marker of reduction in the glomerular filtration rate (GFR), which is a rise in serum creatinine. GFR may be influenced by several types of renal and nonrenal insults. Prerenal factors such as true volume depletion, decreased effective circulating volume, and alterations in the caliber of afferent arterioles that lead to the glomerulus, all cause reductions in GFR and elevations in serum creatinine without significant structural damage to the kidney or activation of inflammatory pathways. Postrenal factors that cause urinary obstruction similarly result in elevations in serum creatinine without damage to the renal tubules. Finally, a multitude of “intrinsic renal diseases” including acute tubular necrosis, rapidly progressive glomerulonephritis, thrombotic microangiopathy, allergic interstitial nephritis, and other glomerular and tubulointerstitial diseases cause ischemic or inflammatory damage to the tubular cells and result in an abrupt rise in serum creatinine. Thus, an increase in serum creatinine clinically does not distinguish among prerenal, postrenal, and intrinsic renal causes. Although various other maneuvers are used to distinguish these various forms of clinical syndromes associated with AKI, the need for a biologically meaningful and accurate biomarker remains.

Urinary Biomarkers for Acute Kidney Injury: Perspectives on Translation

Steven G. Coca and Chirag R. Parikh

Clinical Epidemiology Research Center, Veterans Affairs Medical Center, West Haven, and Department of Medicine, Yale University School of Medicine, New Haven, Connecticut

Acute kidney injury (previously known as acute renal failure) is a common complication in hospitalized patients, and its incidence has risen significantly in the past 15 yr. Despite significant technical advances in therapeutics, the mortality and morbidity rates associated with acute kidney injury remain dismally high and have not appreciably improved during the past four decades. Although the serum creatinine concentration performs fairly well for estimating kidney function in patients with stable chronic kidney disease, it performs poorly in the setting of acute disease. An ideal biomarker for acute kidney injury would help clinicians and scientists diagnose the most common form of acute kidney injury in hospitalized patients, acute tubular necrosis, early and accurately and may aid to risk-stratify patients with acute kidney injury by predicting the need for renal replacement therapy, the duration of acute kidney injury, the length of stay, and mortality. Herein is reviewed the diagnostic and prognostic performance of several types of urinary biomarkers for the diagnosis and risk stratification of acute kidney injury. The major types of urinary biomarkers fall into three classes: (1) Inflammatory, (2) renal tubular proteins that are excreted into the urine after injury, and (3) surrogate markers of tubular injury. Also discussed are statistical issues in evaluating the accuracy of biomarkers as diagnostic tests. It is likely that a panel of biomarkers, rather than a single biomarker, will be needed to perform extremely well in these three situations.
across renal dysfunction, such as microscopic urine examination for various casts, determination of fractional excretion of sodium, and renal ultrasound, these tests are often imprecise and have not advanced the field of nephrology significantly (20–23).

Although the serum creatinine concentration performs fairly well for estimating kidney function in patients with stable chronic kidney disease (CKD), it performs poorly in the setting of acute disease. In AKI, serum creatinine is a poor reflection of kidney function for three reasons. First, a large amount of renal mass can be lost without appreciable changes in serum creatinine because of a concept known as “renal reserve” (24,25). A prototypical example of this is kidney donation. After extraction of one of the two kidneys, there is virtually no change in the serum creatinine concentration in the donor despite the loss of 50% of functioning renal mass (26,27). Second, the kinetics of the rate of rise and peak steady state of serum creatinine concentration in each patient with AKI depends upon the following: 1) relatively constant rate of non-enzymatic conversion of creatine and phosphocreatine in skeletal muscle to creatinine; 2) release of creatinine into the bloodstream; 3) circulation, filtration and excretion of creatinine into the urine (under normal circumstances) (28). Even when glomerular filtration is reduced abruptly to 0% in experimental animals after renal artery clamping, the serum creatinine rise is delayed because it takes time for the creatinine that is released from muscles to accumulate in serum (29). Finally, serum creatinine is also influenced by several nonrenal factors, including body weight, race, age, gender, total body volume, drugs, muscle metabolism, and protein intake (30).

These limitations of serum creatinine increase the risk for failure in drug development. Of the interventions that have been successful in smaller, phase II–level efficacy studies of AKI, most prominently exemplified by the experiences with atrial natriuretic peptide (31) and IGF (32), none was successful at improving clinical outcomes end points such as dialysis requirement or mortality in larger phase III trials (17,18,33). The interventions may have been successful had they been initiated at the onset of AKI rather than after several days, while the creatinine rise was awaited.

### Phases of Biomarker Development

A biomarker is a biologic characteristic that is measured and evaluated objectively as an indicator of normal biologic processes, pathogenic processes, or pharmacologic response to therapeutic intervention (34,35). The field of oncology has been developing biomarkers for diagnostic purposes for decades. Five phases of biomarker development have been proposed by a prominent biostatistician, Dr. Margaret Pepe (Table 1). In the following text, we summarize her paradigm as it pertains to the development of biomarkers for the diagnosis of AKI.

Phase 1 involves preclinical exploratory studies that identify genes or proteins that are over- or underexpressed in AKI compared with without AKI in mice and rat models. Although tissues from organs of experimental animals are vital for phase 1, a noninvasive method (e.g., serum, urine) to assay concentrations of proteins expressed by the identified genes would be preferable. Immunohistochemistry, Western blots, gene arrays, and protein expression profiles have been extensively used for these purposes. The objective of this phase is to identify genes or clusters of genes or proteins that may be present in serum or urine. An example of phase 1 development of biomarkers for AKI is a study that demonstrated via a genome-wide interrogation strategy that the gene for neutrophil gelatinase-associated lipocalin (NGAL) was upregulated more than 10-fold in animals with AKI (36).

Phases 2 and 3 are the translational phases of biomarker development. Phase 2 involves development of clinical assays for clinical disease on the basis of specimens that can be obtained noninvasively. This phase usually requires development of an ELISA or an immunoassay for the protein under investigation. The clinical assay must be shown to distinguish patients with AKI, preferably established AKI, from those without AKI to be considered promising for screening. The specimen collection in this phase occurs concurrently with disease; thus, phase 2 studies typically do not allow determination of whether the disease can be detected early with the proposed biomarkers. In AKI studies, this phase can also involve comparing the results from a variety of control subjects without AKI such as healthy population, chronic kidney disease, urinary tract infection, and other kidney diseases (37,38). The primary aims of phase 2 are

<table>
<thead>
<tr>
<th>Goals</th>
<th>Study Design</th>
<th>Phase</th>
<th>Selected Examples of Specific Aims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery phase</td>
<td>Preclinical exploratory</td>
<td>Phase 1</td>
<td>Promising directions identified</td>
</tr>
<tr>
<td>Translational phase</td>
<td>Clinical assay and validation</td>
<td>Phase 2</td>
<td>Clinical assay detects established disease</td>
</tr>
<tr>
<td></td>
<td>Retrospective longitudinal</td>
<td>Phase 3</td>
<td>Biomarker detects disease early before it becomes clinically obvious</td>
</tr>
<tr>
<td>Validation phase</td>
<td>Prospective screening</td>
<td>Phase 4</td>
<td>Determine sensitivity/specificity</td>
</tr>
<tr>
<td></td>
<td>Disease control</td>
<td>Phase 5</td>
<td>Use biomarker to screen population</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“False referral” rates are identified</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Impact of screening on reducing the burden of disease</td>
</tr>
</tbody>
</table>

*Adapted from reference (73).
to estimate the true-positive rate and the false-positive rate or receiver operating curve (ROC) for the clinical biomarker assay and to assess the ability of the assay to distinguish individuals with AKI from individuals without AKI. Within this stage, other observations, such as behavior of the marker with patient characteristics such as age, race, and gender, and variations in biomarker concentration with spectrum of disease and other disease characteristics should also be noted. If the phase 2 results are robust, then the biomarker can be advanced to the phase 3 development stage. Phase 3 is similar to phase 2, except that the biomarkers are now evaluated for their capacity to detect preclinical disease at earlier time points than the clinical diagnosis (39,40). The other goal of phase 3 could also be to define the cutoffs for a positive screening test that can be validated in phase 4 studies. Availability of biorepository of stored specimens with various clinical phenotypes of diseased and control cases can considerably accelerate the phase 2 and phase 3 testing and development of the biomarkers.

Finally, phases 4 and 5 studies involve large-scale validation of the biomarkers. Compared with phases 1, 2 and 3, these phases involve larger scale prospective cohort studies. Because the disease prevalence of AKI is low, the sample sizes here will be usually large. Thus, adequate piloting, planning, and ethical considerations of the clinical study are very important in phase 4. The primary questions addressed in the phase 4 studies include the understanding of the performance of the biomarker-based “screening” test in a relevant population by determining the true positive rate and false positive rate. Phase 4 studies also help to determine the characteristics of disease that are determined by the test. For example, in the setting of AKI, the ability of the biomarker to detect and differentiate between nonsustained (recovery in 24 h) versus sustained AKI or oliguric versus nonoliguric AKI may be relevant questions. Studies in phase 4 can also help with understanding of the feasibility of applying the newer test and make preliminary assessment of costs and mortality. Phase 5 studies are usually postmarketing studies in which the assessment of reduction of disease burden is made as a result of the availability of the new screening test. For AKI, it would be essential that availability of biomarkers reduce the mortality associated with the disease.

Types of Urinary Biomarkers for AKI
An ideal biomarker for AKI would help clinicians and scientists diagnose the most common form of AKI in hospitalized patients, acute tubular necrosis (ATN), early and accurately. Potential biomarkers thus can reflect tubular inflammation (because inflammation frequently accompanies and magnifies tubular injury in ATN), can reflect change in excretion of a normally occurring renal tubular protein that is “shed” into the urine after injury, or can reflect frank tubular dysfunction by the appearance of proteins or enzymes in the urine that are normally filtered at the glomerulus but reabsorbed or metabolized by the tubules.

Biomarkers may not only aid with the differential diagnosis of AKI by differentiating true tubular injury (e.g., ATN) from other forms of renal dysfunction but may also potentially identify the location of tubular injury (proximal versus distal), the cause (ischemia versus toxin), and the temporal course (acute versus chronic) of renal dysfunction. These same biomarkers or perhaps other sets of biomarkers will hopefully also detect AKI at an earlier time point than the typically delayed rise in serum creatinine so that a rapid diagnosis can be made and interventions can be applied in a timely manner for patients who are at high risk for AKI, such as patients who undergo cardiac surgery, receive deceased-donor kidneys, and receive intra-arterial radiocontrast loads and for critically ill patients. Finally, these or other biomarkers may aid to risk-stratify patients with AKI by predicting the need for renal replacement therapy (RRT), the duration of AKI, the length of stay, and ultimately mortality (Figure 1).

Urinary Biomarkers for Improving the Differential Diagnosis of Established AKI
Identifying which patients have ATN and which have reversible prerenal dysfunction or CKD at the time of evaluation for an acutely elevated creatinine remains a clinical dilemma. Five phase 2 studies (37,38,41–43) have examined urine biomarkers for improving the differential diagnosis of established AKI (Table 2).

Inflammatory Biomarkers
NGAL. NGAL is a 25-kD protein that is covalently bound to gelatinase from neutrophils. NGAL is normally expressed at very low levels in several human tissues, such as the kidney, lungs, stomach, and colon. Several phase 1 studies in experimental animals demonstrated that NGAL is one of the earliest and most robustly induced genes and proteins in the kidney after ischemic or nephrotoxic injury (36,43–45). In a cross-sectional study of 57 adults, urinary NGAL was elevated 25-fold in patients with established ATN compared with patients with chronic renal failure or control patients without AKI (43). Patients with chronic renal failure had less prominent eleva-
Table 2. Major objectives for future studies of biomarkers

- Validate all biomarkers that performed moderate to strong in phase 2/3 studies in larger multicenter cohorts
- Validate biomarkers in different clinical settings (e.g., cardiac surgery, sepsis, DGF)
- Validate AKI biomarkers in patients with pre-existing CKD
- Examine trajectory of biomarker elevation during the course of AKI
- Identify biomarkers that identify imminent AKI recovery
- Assess the predictive value of biomarkers for risk stratification and additional improvement to existing models of severity
- Develop a panel of biomarkers to maximize performance for differential diagnosis, early diagnosis, and risk stratification in AKI
- Identify biomarkers that will distinguish between ischemic and nephrotoxic AKI
- Use a validated set of biomarkers for early diagnosis and/or risk stratification that can be used in clinical trials of interventions for AKI
- Determine the cost-effectiveness of biomarkers

*AKI, acute kidney injury; CKD, chronic kidney disease; DGF, delayed graft function.

...tions in urinary NGAL. Studies to date have not examined whether urine NGAL is elevated in prerenal or obstructive forms of AKI. Studies that involve early diagnosis of AKI have not differentiated between causes of AKI; therefore, it is plausible that urinary NGAL may be elevated in forms of AKI other than pure ATN.

IL-18. IL-18 is a proinflammatory cytokine that is induced and cleaved in the proximal tubule after AKI. The intracellular cysteine protease, caspase-1, converts the proform of the cytokines IL-1β and IL-18 to their active forms (46,47). The active form of IL-18 exits the cell and may enter the urine after being activated in proximal tubules. In mice, urinary IL-18 concentration was increased in ischemic AKI compared with sham-operated mice (46). In a cross-sectional study of 72 patients with various renal diseases, urine IL-18 was significantly greater and had a high sensitivity and specificity for the diagnosis of ATN versus prerenal azotemia, urinary tract infection, chronic kidney disease, and normal renal function in healthy control subjects (38).

Tubular Proteins

Kidney Injury Molecule-1. Kidney injury molecule-1 (KIM-1) is a transmembrane protein that is overexpressed in proximal tubular cells in ischemic (48) or nephrotoxic (49) AKI in animals. A phase 2 study confirmed KIM-1 expression via immunohistochemistry on kidney biopsy sections taken from six patients with confirmed ATN (37). In the second phase of the study, 40 patients had urinary KIM-1 concentrations measured (37). KIM-1 was elevated to a much higher degree in patients with ischemic ATN than in patients with contrast nephropathy or other forms of acute renal failure, in patients with CKD, or in normal control subjects. The results from this study suggest that KIM-1 is useful in identifying ischemic ATN.

Na⁺/H⁺ Exchanger Isoform 3. The Na⁺/H⁺ exchanger isoform 3 (NHE3) is the most abundant sodium transporter in the renal tubules and is located in the apical membrane of proximal tubular cells and thick ascending limb cells. After ischemia and/or necrosis in experimental animals, NHE3 abundance decreases (50), leading to the hypothesis that the tubular damage of AKI will result in liberation of NHE3 from the apical membrane, into the tubular lumen, and thus excretion into the urine. One study (42) demonstrated that urinary NHE-3 was elevated by six-fold in patients with AKI compared with control subjects with prerenal azotemia.

Surrogate Markers of Tubular Injury

Several low molecular weight proteins and enzymes are filtered through the glomerulus and subsequently reabsorbed in the proximal tubule. Tubular injury may result in a decreased ability for these proteins and enzymes to be absorbed along the nephron and cause elevated concentrations in the urine.

**Urinary Biomarkers for the Early Diagnosis of AKI**

**Inflammatory Biomarkers**

NGAL. In a children who underwent cardiac surgery (39) and in adults who received deceased-donor kidneys (51), urinary NGAL concentrations were highly predictive of subsequent clinical diagnosis of AKI (children: sensitivity 1.0, specificity 0.98, area under the curve [AUC] 0.99 at 2 h after surgery; adults: sensitivity 0.9, specificity 0.83, AUC 0.9). In contrast to these remarkable results, subsequent studies have not yielded results as robust, with sensitivities and specificities for detecting AKI <80% in a cohort of adults who underwent cardiac surgery (52) and in a cohort of critically ill children (53).

IL-18. Four phase 3 studies (40,51,54,55) of urinary IL-18 as an early predictive biomarker of AKI have been published. The AUC for IL-18 ranged from 0.54 to 0.9 for the early diagnosis of AKI. In three of the studies that demonstrated lower AUC, IL-18 demonstrated low sensitivity but high specificity (0.85 to 0.94) (40,54,55). As noted for NGAL, the performance of IL-18 also depends on the timing of collection in relation to the exposure (e.g., the stressor such as cardiac surgery) and the outcome (development of AKI) (40,54). Of note, the IL-18 studies adjusted for several clinical variables in multivariate analyses for prediction of AKI and found that IL-18 still independently predicted AKI (40,51,54,55). In summary, IL-18 performed moderately well for detecting AKI early in four different populations (adults who underwent cardiac surgery, children who were admitted with critical illness, adults with acute lung injury, and adults who received kidney transplantation).
**Surrogate Markers of Tubular Injury**

A small study (56) of 26 critically ill patients (only four with AKI) reported AUC of 0.95, 0.93, 0.89, 0.86, and 0.85 for the tubular enzymes γ-glutamyl transpeptidase (γ-GT), GST-π, GST-α, alkaline phosphatase, and NAG, respectively, for predicting the development of AKI. However, other studies of enzymuria have demonstrated that the rise in urinary enzymes is extremely sensitive for renal tubular injury and increases in patients who do not develop clinical AKI after surgery (41,57). Thus, the predictive value of enzymuria is poor.

Albumin in the urine is detectable in significant amounts within a few hours after aortic surgery (58,59). However, albuminuria does not discriminate between patients with and without AKI and also drops to near-normal levels within 24 h of injury even in patients with clinical AKI (58,59).

**Urinary Biomarkers for Risk Stratification of AKI (Severity of AKI)**

In current clinical practice, at the time of AKI diagnosis or nephrology consultation, we lack accurate tools to predict who will have severe AKI (necessitating RRT) or prolonged AKI or who will die with AKI. Several investigators have developed both general clinical and kidney-specific models in an attempt to predict severe AKI. Many of these models were evaluated in the Program to Improve Care in Acute Renal Disease (PICARD), a cohort of 618 adults with AKI (60). The best performing predictors were the APACHE III score (AUC 0.66 and 0.70 to predict on the day of AKI diagnosis and day of renal consultation, respectively) and the Sequential Organ Failure Assessment (SOFA) score (AUC 0.64 and 0.70, respectively). These scores are cumbersome to calculate and time-consuming. Thus, a biomarker that would be able to predict either RRT- or AKI-associated death with an equal or higher accuracy will simplify risk stratification remarkably.

**Inflammatory Biomarkers**

A study (61) of 34 children with diarrhea-associated hemolytic uremic syndrome found that an elevated level of urinary NGAL (>200 ng/ml) within the first 5 d of hospitalization had a sensitivity of 0.90, a specificity of 0.54, a positive predictive value of 0.45, and a negative predictive value of 0.93 for the subsequent need for dialysis. In another study of adult cardiac surgery patients (54), the log of NGAL ng/ml at 4 h and log of IL-18 pg/ml at 4 h after cardiac surgery correlated with number of days with AKI. In another cohort (children admitted to critical care units), urine NGAL and IL-18 performed moderately for predicting persistent AKI (duration >48 h). NGAL had a sensitivity of approximately 0.67 to 0.78 and a specificity of approximately 0.67 to 0.69 (53), and IL-18 had a sensitivity of approximately 0.21 to 0.47 and a specificity of 0.71 to 0.93 (55) for the diagnosis of persistent AKI.

**Tubular Proteins**

A recent study (62) examined the relationship between KIM-1 and the composite end point (dialysis or death) in their cohort of 201 hospitalized patients. Urinary KIM-1 level had an AUC of 0.61 for the prediction of the composite end point. The APACHE II score had an AUC of 0.78, which was increased slightly to 0.80 when KIM-1 was added to the model. The association between the KIM-1 quartiles and composite outcome revealed odds ratios (OR) of 1.4, 1.4, and 3.2 for patients with increasing quartiles as compared with the lowest quartile. These associations were no longer significant after adjustment for covariates; therefore, the independent value of KIM-1 for predicting severe AKI is unclear.

**Surrogate Markers of Tubular Injury**

One study (63) examined the prognostic value of several tubular proteins to predict the need for RRT in 73 patients with nonoliguric AKI. Of the eight urinary biomarkers tested, four performed well for predicting RRT, including three proteins—cystatin C, α-1 microglobulin, and retinol-binding protein—and one enzyme—NAG. The respective AUC were 0.92, 0.86, 0.80, and 0.81. The other four urinary biomarkers (β2-microglobulin, α-GST, γ-glutamyltransferase, and lactate dehydrogenase) did not perform as well (AUC 0.51, 0.64, 0.64, and 0.59, respectively). β2-Microglobulin has been studied often as a predictor of severe AKI but generally performs poorly (57,64) and is also elevated in chronic renal disease (65).

In the study that also examined KIM-1 (62), NAG demonstrated a dosage-dependent association with the prediction of the composite end point of dialysis or death. The association between the NAG quartile and the composite outcome revealed OR of 3.0, 3.7, and 9.1 for patients in the second, third, and fourth quartiles versus the first quartile, respectively, and these results remained robust after adjustment for APACHE II (adjusted OR 3.3, 3.5, and 5.4, respectively). Urinary NAG activity had an AUC of 0.71 to predict the composite outcome. The addition of both NAG activity and KIM-1 level to the APACHE II score increased the AUC for prediction of the composite outcome from 0.78 to 0.83. Thus, it seems that measurement of NAG and KIM-1 increases the ability to predict severe AKI to a modest degree. In summary, several markers of tubular injury (both proteins and enzymes) seem promising for the risk stratification for AKI.
the composite outcome of dialysis and death were examined and added prognostic value (62).

**Statistical Issues in Evaluating the Accuracy of Biomarkers as Diagnostic Tests**

*Summarizing the Performance of a Classifier*

Reporting of OR or relative risk is typical in epidemiologic studies to demonstrate strength of association between risk factor and the disease; however, use of these measures is not adequate to determine the performance of a biomarker for classifying disease or predicting risks in people. Strong statistical associations (OR or relative risk) between a biomarker and an outcome do not necessarily suggest that the biomarker can discriminate between people who are likely to have the outcome and those who are not (66). For example, a biomarker with an OR of 3, which may be considered a strong risk factor, is actually a very poor “classifier” in terms of discriminating patients into disease versus no disease (Figure 2). Pepe et al. (66) argued that the accuracy or validity of a marker for classifying individuals is better summarized by reporting its true-positive fraction (TPF; or sensitivity) and its false-positive fraction (FPF; also known as 1 – specificity). A perfect marker will have TPF = 1 and FPF = 0, but the criteria by which the marker is judged useful will depend entirely on the context in which it is to be used. For example, if one is using the biomarker to determine who will need an invasive procedure or a risky treatment, then an extremely low FPF (<2%) is needed to avoid placing patients at risk for these procedures or treatments. Thus, characterization of false-positive and false-negative errors is not equivalent and must be reported separately. The ROC curve is the prime statistic for evaluation of the accuracy of continuous biomarkers for outcomes because it describes a whole set of potential (FPF, TPF) combinations possible for different cutoff values, does not depend on how the marker is coded, and provides a natural common scale for comparing different markers even when they are measured in completely different units. In Figure 2, the curves closest to the top left corner represent strong ROC curves, and the curves closest to the midline represent weaker ROC curves.

*Errors in Reference Standard*

A major threat to the validity of studies of biomarkers for the diagnosis of AKI is errors in the reference test or gold standard. For most cases of AKI (ATN), true disease status is usually impossible to determine accurately. Reasons for the lack of gold standard are that our best available tests for identifying ATN are imprecise. The limitations of the use of change in serum creatinine have been discussed. Furthermore, some of the biomarkers (e.g., NGAL) are increased in the presence of urinary tract infection. Moreover, nearly 30% of patients with AKI have preexisting CKD (67). Many of the biomarkers have not had their performance rigorously evaluated in patients with preexisting CKD. Finally, the acquisition of tissue via kidney biopsy is not a reasonable alternative for a gold standard for AKI because of its invasiveness, its risk in critically ill patients, and its relative insensitivity for detecting morphologic changes even in cases of obvious ATN. These limitations are difficult to overcome with statistical methods; therefore, the best option may be to choose hard outcomes that do not allow for misclassification (e.g., need for dialysis, death).

*Predictive Values, Risk Models, and Predictiveness*

There are two popular statistical approaches for biomarker evaluation. One models the risk for disease (or disease outcome) using, for example, logistic regression. A marker is useful if it has a strong effect on risk. The second evaluates classification performance using measures such as sensitivity, specificity, predictive values, and ROC curves.

The focus of clinical studies in disease screening is on the capacity of a biomarker to classify accurately individuals as diseased or not. Classification performance parameters such as sensitivity and specificity are of key interest, because ultimately it is the proportion of individuals who have disease detected (sensitivity) and the proportion of individuals who do not have disease and unnecessarily undergo procedures or treatment (1 – specificity) that enter into decisions about screening policy. The ROC curve gives a complete idea of the range of sensitivities and specificities for a continuous marker as a classifier.

The evaluation of markers for disease risk prediction in individual patients, however, requires a different approach. In this context, the goal is to quantify how well a marker identifies people at high or low risk for disease. The predictiveness curve provides a complete and conceptually simple description of the capacity of a marker to predict risk (Figure 3) (68). A disease risk prediction marker might be used to select individuals for a prevention intervention or for screening but does not classify individuals directly.
Nonstatistical Issues for Biomarker Studies

Translation of Data from Animals to Humans

Studies of AKI in experimental animals typically represent very “clean” conditions in that the timing of the insult is known (e.g., ischemia-reperfusion injury) and the animals typically lack any preexisting renal disease. In humans, often the timing of the insult is unknown, the injury is not purely ischemic, and a large proportion of patients may have preexisting renal disease or other systemic comorbidities, which may influence the biomarker concentration. Furthermore, the homology between AKI in experimental animals and humans may be low. For example, whereas ischemia-reperfusion often causes an inflammatory response in the kidney in experimental animals, septic models of AKI do not manifest with a high degree of leukocyte recruitment into the renal tissue (69). The histologic findings in clinical AKI are even more ambiguous. Biopsy studies of patients with septic AKI have demonstrated that only 8 to 50% of patients with clinical “ATN” had pathologic evidence of ATN on histologic specimens (70,71). Thus, biomarkers that are deemed successful in experimental animals and humans may be low. For example, whereas ischemia-reperfusion often causes an inflammatory response in the kidney in experimental animals, septic models of AKI do not manifest with a high degree of leukocyte recruitment into the renal tissue (69). The histologic findings in clinical AKI are even more ambiguous. Biopsy studies of patients with septic AKI have demonstrated that only 8 to 50% of patients with clinical “ATN” had pathologic evidence of ATN on histologic specimens (70,71). Thus, biomarkers that are deemed successful in diagnosing AKI in experimental animals will need to be validated detecting the phenotype of classic AKI (ATN) in humans. Despite these potential pitfalls, biomarkers such as NGAL, KIM-1, and IL-18 have been solid biomarkers in both animal models and humans.

Variability of the Gold Standard Definition of AKI

Because the true phenotype of ATN is unclear in humans, nephrologists have struggled to derive a consensus definition for AKI. There are more than 30 definitions for AKI in the literature. More definitions have recently surfaced, including the recent AKIN (Acute Kidney Injury Network) definition of AKI (72); however, the AKIN definition of AKI does not incorporate cause of AKI into the definition. Furthermore, the studies of biomarkers to date have not used one consistent definition of AKI. Adoption of the AKIN definition of AKI for future biomarker studies may help to provide the ability to compare accuracy results across studies.

Timing of Biomarker Analysis

An ideal biomarker for AKI would be elevated within minutes to hours of injury and would remain elevated for a period of hours to days (similar to serum troponin for the diagnosis of acute coronary syndromes and acute myocardial infarction). Unfortunately, patients do not have “symptoms” of AKI as most have chest pain in the setting of acute coronary syndromes and acute myocardial infarction, which prompts clinical evaluation. Thus, the timing of measurement of urinary biomarkers would vary depending on the clinical situation. In a patient who is at high risk for AKI and has a known stimulus for kidney injury (e.g., cardiac surgery), the biomarkers can be obtained soon after the completion of surgery to facilitate early diagnosis of AKI; however, in situations in which the timing of the insult is not as well defined (e.g., gradual onset of sepsis), early diagnosis of AKI may be less achievable. A sustained elevation in biomarkers, however, would serve to inform about prognosis. Subsequent studies of biomarkers for AKI need to determine comprehensively the course of biomarker concentrations over time in cohorts of patients from different clinical settings.

Conclusions

A systematic phased approach and multidisciplinary involvement are essential for translating urinary biomarkers into diagnostic tests for AKI. On the basis of the carefully designed phase 2 and phase 3 studies, it is essential to make decisions to identify the potential of each marker for further development. This would allow efficient development of individual biomarkers for each indication without wasted resources and energy. On the basis of the phase 2 and phase 3 studies in the literature, IL-18 and KIM-1 have the highest
potential for the differential diagnosis of established AKI; NGAL, IL-18, GST-π, and γ-GST have good potential for early diagnosis of AKI, and NAG, KIM-1, and IL-18 have the highest potential for mortality risk prediction after AKI. However, it is likely that a panel of biomarkers, rather than a single biomarker, will be needed to perform extremely well in these three situations.

Acknowledgments
Studies that were cited in this review and performed by the authors were supported by grants from Donaghue Foundation Grant and Satellite Foundation Grant, Clinical Scientist Award from American Heart Association, and grants from the National Institutes of Health to C.R.P. (K23DK064689 and RO1HL85757) and to S.G.C. (F32 DK076318-01A1).

Disclosures
None.

References

Urinary Biomarkers in AKI


