Nephrol Dial Transplant (2009) 24: 3263–3265
doi: 10.1093/ndt/gfp428
Advance Access publication 7 September 2009

Editorial Comments

NDT
Nephrology Dialysis Transplantation

Creatinine as the gold standard for kidney injury biomarker studies?

Sushrut S. Waikar[1], Rebecca A. Betensky[2] and Joseph V. Bonventre[1,3]

[1]Renal Division, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, [2]Department of Biostatistics, Harvard School of Public Health and [3]Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA

Correspondence and offprint requests to: Sushrut S. Waikar; E-mail: swaikar@partners.org

Keywords: acute renal failure; biomarkers; creatinine

Before 2005, when the Acute Dialysis Quality Initiative proposed a consensus definition [1], acute kidney injury (formerly known as 'acute renal failure') was identified by most clinicians in the way that Justice Potter Stewart identified obscenity: they knew it when they saw it. Epidemiologists and clinical researchers, who needed an objective criterion, seemed to devise a different definition for every new study; indeed, over 35 definitions have been used to define AKI in the nephrology literature [2]. For Homer Smith, who introduced the term 'acute renal failure' in his textbook, *The Kidney: Structure and Function in Disease and Health* [3], a specific definition did not seem to matter: nowhere in his textbook does he propose a way to define AKI.

We know that definitions do matter in modern clinical medicine. The consensus definition of acute myocardial infarction, which has evolved over the years and now requires biochemical evidence of myocardial tissue injury, has facilitated the design and execution of studies that have led to revolutionary changes in the treatment and outcome of patients with acute myocardial infarction around the world. The consensus definitions of sepsis and acute lung injury similarly permitted the critical care community to study and advance the clinical science and management of critically ill patients, with some clear success stories [4,5]. Within nephrology, the new consensus definition of chronic kidney disease, while controversial, has focused attention on early recognition and staging of the clinical syndrome, influenced clinical trial design as well as referral patterns to nephrologists [6], and hopefully will provide a tangible benefit for our patients.

Defining AKI is clearly of paramount importance as nephrology struggles to translate a host of steady advances in the understanding of the pathobiology of AKI into clinical benefit for patients. Unfortunately, however, we face a challenge not faced by cardiologists and pulmonologists: the uncertain relationship between our 'gold standard' biomarkers [serum creatinine (SCr) and urine output] and our disease process. In cardiology, the relationship between the defining biomarker (troponin) and the disease process (myocardial infarction) is direct: injured myocardial cells release troponin into the circulation, where it can be measured and used to diagnose a clinical condition. In pulmonology, one of the biomarkers used is the partial pressure of oxygen: acute lung injury and acute respiratory distress syndrome are defined by the difference between the alveolar and arterial concentration of oxygen. This biomarker reflects the critical life sustaining function of the organ: delivery of oxygen to the circulation to support aerobic metabolism.

What is curious about our universally used biomarker (SCr) is that we do not trust it; it is not an injury marker but rather a functional marker (and a poor one at that, especially in the acute setting). A rise in SCr may not define kidney injury at all: for example, the syndrome 'pre-renal azotaemia' may look biochemically just like 'acute tubular necrosis' by changes in SCr, but differs markedly in underlying pathophysiology, treatment implications and prognosis. All nephrologists know that SCr can be deceptively normal: consider lupus nephritis, in which severe parenchymal injury can coexist with preserved glomerular filtration rate and maintenance of a normal SCr. The ability of SCr to reliably identify AKI is particularly impaired when we try to define relatively milder forms of the disease.

Using mathematical models of creatinine kinetics, we have argued [7] that the prevailing consensus definitions of AKI, which largely use percentage changes in SCr, are misinformed because they will lead to a delay in the diagnosis in patients with elevated baseline SCr levels—precisely the population in which AKI is most common. Any definition of AKI, however, which is based on SCr—whether absolute increases over a defined time period, as we have proposed, or percentage rises over baseline—will be bound to misclassify patients. (The inclusion of urine output criteria present in the RIFLE [1] and AKIN [8] definitions may add even more confusion: oliguria can be masked by diuretics, and can denote simple mechanical obstruction of Foley catheters.)

It is interesting in this context to observe the evolution of the kidney injury biomarker field, where several proteins and urinary enzymes have been studied as potential

biomarkers of AKI. Many study reports published to date (ours included) begin by reciting creatinine's imperfections as a biomarker—non-specificity due to pre-renal azotaemia, non-sensitivity due to renal reserve—and then go on to judge the performance of the biomarker being studied against the same 'gold standard' whose imperfections engendered the need to discover a novel and superior biomarker. Seen in this light, a perfect biomarker with 100% sensitivity and 100% specificity when compared to SCr is just that: SCr redux. The same problems that plague SCr and justified the need for a replacement biomarker are not addressed by such a new biomarker. The diagnosis may be made sooner, but it is no more accurate than SCr. This would be informative if we truly trusted changes in SCr to reliably reflect AKI. Consider, however, the extent of creatinine elevation used to define AKI in most studies. In the first paper by Mishra *et al.* on NGAL's performance (neutrophil gelatinase-associated lipocalin) in paediatric cardiac surgery, AKI was diagnosed as a 50% increase in SCr [9]. Since baseline SCr levels in infants can be as low as 0.3 mg/dL [10], AKI could be diagnosed by a rise of just of 0.15 mg/dL. It is difficult to argue with certainty that a transient rise in SCr of this magnitude can reliably reflect clinically meaningful AKI.

In this regard, the study by Haase–Felitz in this issue of *Nephrol Dial Transplant* has addressed an important question: how does the performance characteristic of NGAL as a biomarker change when demanding a higher percentage change in SCr to define AKI? They show that in adult cardiac surgery, the predictive value of plasma NGAL varied according to the AKI definition, with the best performance seen when using the most strict definition.

Their study raises important methodologic issues relevant to kidney injury biomarker studies. The more general problem is that of the imperfect gold standard [11]. The imperfections of SCr as a gold standard influence directly the apparent performance characteristics of a novel biomarker: if SCr does not rise in a certain proportion of cases of actual AKI (say, for example, as diagnosed by a kidney biopsy) but a novel biomarker does, then the biomarker will appear to be non-specific. If SCr rises in some cases of pre-renal azotaemia but the novel biomarker does not, then the biomarker will appear to lack sensitivity. If the ability of SCr to diagnose AKI differs from one population to the next—for example, if SCr is a better biomarker in children than in adults—then the diagnostic performance characteristics of a novel biomarker will appear to differ in the two populations, not because of its inherent ability to diagnose AKI but because of SCr's differences in reliability.

A second important issue in AKI biomarker studies exemplified by this and many other studies is that of converting a continuous variable into a dichotomous outcome. Though we say 'AKI' and 'no-AKI,' we are in fact dealing with a range of values for SCr and its change. A cutoff is arbitrarily applied (50% or 25%), and those with values above the cutoff are designated as 'AKI' and those below as 'no-AKI'. Where that cutoff is chosen can influence the apparent performance characteristics of a biomarker. In particular, if intermediate values are excluded—no AKI is defined as <0.3 mg/dL increase; and AKI is defined as a 200% increase, while excluding all of the individuals in between—then the diagnostic performance characteristics of a biomarker may appear to improve, simply by excluding intermediate values [12]. This is precisely what was done in the paper by Haase–Felitz; the improvement in NGAL's performance when defining AKI more strictly may be an epiphenomenon of this exclusion rather than an inherent property of the biomarker.

We believe several areas in the biostatistical approach to biomarker studies need to be more fully investigated. First, how can an existing gold standard known to be imperfect be replaced by a new biomarker? If NGAL is indeed the perfect biomarker and correlates perfectly with the true diagnosis of AKI, it is very likely that future studies will show that it is non-specific or non-sensitive (indeed, such studies have been published); but this failure may not reflect NGAL's imperfections, but rather SCr's imperfections. We have to remain open to including endpoints other than SCr to add clarity to the diagnosis of AKI, such as urinary sediment examination, although with current evaluative techniques this marker is not ideal either [13]. Perhaps multiple injury biomarkers that are consistent in their predictive capabilities will be more reliable than SCr to reflect pathology: will we be brave enough to revise our definitions based on these biomarkers? Methods for resolution via a secondary biomarker have been proposed [14] and could be extended to the setting of multiple binary and continuous biomarkers. Second, what is the best analytic approach in a biomarker study using a continuous variable (like change in SCr) as the gold standard? There have been attempts to develop ROC-type measures using a continuous variable outcome [12], and these approaches should be developed more fully for use in AKI biomarker studies.

Biomarker development in nephrology is crucial if we hope to develop therapeutic strategies for AKI prevention and treatment, but may be doomed to stall if we rely solely on SCr as a gold standard for diagnosis. We are on the verge of a possible paradigm shift in nephrology, where the diagnosis of AKI moves away from a functional biomarker like SCr and towards novel tubular injury biomarkers that have been identified in animal models as biologically plausible. While a gold standard measure of disease status may remain illusive, the field of AKI diagnostics does have the benefit of several novel biomarkers, even if potentially imperfect. This advantage should be exploited to validate optimal biomarker(s). Whether nephrology evolves away from SCr to tissue-specific injury biomakers—in the way that cardiology evolved away from lactate dehydrogenase and creatine phosphokinase towards the troponins for the diagnosis of myocardial infarction—will depend on the continued performance of well-conducted clinical studies, with explicit consideration of the limitations of the gold standard being used to define AKI and clear statements of what constitutes true injury. Most importantly, it is time for the biostatistical and epidemiological sophistication of kidney injury biomarker studies to match that of the underlying basic science.

Conflict of interest statement. Dr. Bonventre is an inventor on patents on KIM-1 owned by Partners Health Care and licensed to Johnson and Johnson, Biogen-Idec and Genzyme Corp.

(See related article by A. Haase-Fielitz *et al.* The predictive performance of plasma neutrophil gelatinase-associated lipocalin (NGAL) increases with grade of acute kidney injury. *Nephrol Dial Transplant* 2009; 24: 3349–3354.)

References

1. Kellum JA, Bellomo R, Ronco C *et al*. The 3rd International consensus conference of the acute dialysis quality initiative (ADQI). *Int J Artif Organs* 2005; 28: 441–444
2. Mehta RL, Chertow GM. Acute renal failure definitions and classification: time for change? *J Am Soc Nephrol* 2003; 14: 2178–2187
3. Smith HW. *The Kidney: Structure and Function in Health and Disease*. New York: Oxford University Press, 1951
4. The Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000; 342: 1301–1308
5. Rivers E, Nguyen B, Havstad S *et al*. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001; 345: 1368–1377
6. Jain AK, McLeod I, Huo C *et al*. When laboratories report estimated glomerular filtration rates in addition to serum creatinines, nephrology consults increase. *Kidney Int* 2009 May 13 [Epub ahead of print]
7. Waikar SS, Bonventre JV. Creatinine kinetics and the definition of acute kidney injury. *J Am Soc Nephrol* 2009; 20: 672–679
8. Molitoris BA, Levin A, Warnock DG *et al*. Improving outcomes of acute kidney injury: report of an initiative. *Nat Clin Pract Nephrol* 2007; 3: 439–442
9. Mishra J, Dent C, Tarabishi R *et al*. Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet* 2005; 365: 1231–1238
10. Rudd PT, Hughes EA, Placzek MM *et al*. Reference ranges for plasma creatinine during the first month of life. *Arch Dis Child* 1983; 58: 212–215
11. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. *Stat Methods Med Res* 1998; 7: 354–370
12. Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. *Stat Med* 2006; 25: 481–493
13. Wald R, Bell CM, Nisenbaum R *et al*. Interobserver reliability of urine sediment interpretation. *Clin J Am Soc Nephrol* 2009; 4: 567–571
14. Alonzo TA, Pepe MS. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Stat Med* 1999; 18: 2987–3003

Received for publication: 28.7.09; Accepted in revised form: 30.7.09

Nephrol Dial Transplant (2009) 24: 3265–3268
doi: 10.1093/ndt/gfp010
Advance Access publication 23 March 2009

Kidney injury molecule-1 (KIM-1): a urinary biomarker and much more

Joseph V. Bonventre

Renal Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
Correspondence and offprint requests to: Joseph V. Bonventre; E-mail: joseph_bonventre@hms.harvard.edu

Keywords: acute kidney injury; TIM-1; acute renal failure; phagocytosis; apoptosis

KIM-1 kidney expression and function

The kidney injury molecule-1 (designated as Kim-1 in rodents, KIM-1 in humans) mRNA was identified using techniques of representational difference analysis, a PCR-based technique [1], which we carried out to find genes whose expression was markedly upregulated 24–48 h after ischaemia in the rat [2]. *Kim-1* was the gene found to be most highly upregulated in this screen. A large pharmaceutical company consortium, using an unbiased genomic approach to evaluate genes upregulated with the nephrotoxin cisplatin, determined that *Kim-1* was upregulated more than any other of the 30 000 genes tested [3]. There are a large number of studies in animals showing robust Kim-1 protein production in the affected segments of the proximal tubule whenever a toxin or pathophysiological state results in dedifferentiation of the epithelium (e.g. [4–6]). Dedifferentiation is a very early manifestation of the epithelial cell response to injury [7]. KIM-1 is also expressed, at much lower levels, in lymphocytes and has also been referred to as T-cell immunoglobulin mucin (TIM)-1 and HAVCR-1, hepatitis A virus cellular receptor-1. The protein has also been reported to be expressed in the cochlea in response to cisplatin-induced injury [8]. The KIM/TIM family consists of eight members in mice, six in rats and three in humans [9,10].

Using standard northern or western blot analyses and immunocytochemistry, *KIM-1* gene or protein expression is undetectable in the normal kidney. With injury KIM-1 mRNA is rapidly made and protein is generated and localized at very high levels on the apical membrane of proximal tubule in that region where the tubule is most affected. In the case of experimental ischaemia in rodents, Kim-1 expression is predominantly in the S3 segment of the proximal tubule. In human ischaemic and toxic acute kidney injury (AKI) it is found in the three segments of the proximal tubule.