

Clinical adjudication in acute kidney injury studies: findings from the pivotal TIMP-2•IGFBP7 biomarker study

Kathleen D. Liu¹, Anitha Vijayan², Mitchell H. Rosner³, Jing Shi⁴, Lakhmir S. Chawla⁵ and John A. Kellum⁶

¹Division of Nephrology, Department of Medicine and Division of Critical Care Medicine, Department of Anesthesia, University of California, San Francisco, CA, USA, ²Division of Nephrology, Department of Medicine, Washington University, St Louis, MO, USA, ³Department of Medicine, University of Virginia, Charlottesville, VA, USA, ⁴Statistics, Walker Bioscience, Carlsbad, CA, USA, ⁵Department of Medicine, Divisions of Intensive Care Medicine and Nephrology, Veterans Affairs Medical Center, Washington, DC, USA and ⁶Center for Critical Care Nephrology, Department of Critical Care Medicine, University of Pittsburgh, USA

Correspondence and offprint requests to: John A. Kellum; E-mail: Kellumja@upmc.edu

ABSTRACT

Background. The NEPROCHECK test (Astute Medical, San Diego, CA, USA) combines urinary tissue inhibitor of metalloproteinases-2 (TIMP-2) and insulin-like growth factor binding protein 7 (IGFBP7) to identify patients at high risk for acute kidney injury (AKI). In a US Food and Drug Administration registration trial (NCT01573962), AKI was determined by a three-member clinical adjudication committee. The objectives were to examine agreement among adjudicators as well as between adjudicators and consensus criteria for AKI and to determine the relationship of biomarker concentrations and adjudicator agreement.

Methods. Subjects were classified as AKI 3/3, 2/3, 1/3 or 0/3 according to the proportion of adjudicators classifying the case as AKI. Subjects were classified as Kidney Disease: Improving Global Outcomes (KDIGO) AKI(+) when stage 2 or 3 AKI criteria were met.

Results. Concordance between adjudicators and between adjudicators and KDIGO criteria were lower for AKI than non-AKI subjects [78.9 versus 97.3% ($P < 0.001$) and 91.5 versus 97.9% ($P = 0.01$)]. Subjects who were AKI 3/3 or 2/3 but KDIGO AKI(–) had higher median [TIMP-2]•[IGFBP7] compared with those who were AKI-1/3 or 0/3 but KDIGO AKI(+) {2.78 [interquartile range (IQR) 2.33–3.56] versus 0.52 [IQR 0.26–1.64]; $P = 0.008$ }. [TIMP-2]•[IGFBP7] levels were highest in patients with AKI 3/3 and lowest in AKI 0/3, whereas AKI 2/3 and 1/3 exhibited intermediate values.

Conclusions. In this analysis, urine [TIMP-2]•[IGFBP7] levels correlated to clinically adjudicated AKI better than to KDIGO criteria. Furthermore, in difficult cases where adjudicators overruled KDIGO criteria, the biomarker test discriminated well.

This study highlights the importance of clinical adjudication of AKI for biomarker studies and lends further support for the value of urine [TIMP-2]•[IGFBP7].

Keywords: acute kidney injury, biomarkers, diagnosis, insulin-like growth factor binding protein 7, tissue inhibitor of metalloproteinases-2

INTRODUCTION

Acute kidney injury (AKI) is common among hospitalized patients and is associated with significant morbidity and mortality [1, 2]. At present, there are no therapies specific for AKI, and care of patients with AKI is mainly supportive [3]. Consequently, early identification of AKI is thought to be critical because it allows for prompt implementation of best practices, including reevaluation of drug dosing, nephrotoxin avoidance, close attention to fluid balance and other potential therapeutic maneuvers. In some cases it has been suggested that interventions (e.g. transfer to the intensive care unit or placement of a line for dialysis) should be considered based on the presence or absence of moderate–severe AKI [4]. Ultimately, if patients at high risk of AKI can be reliably identified, this may allow for clinical trials of novel therapies designed to reduce the severity or duration of AKI [5]. Along the same lines, if earlier detection of renal injury or identification of those at risk was feasible, then best practice interventions could be implemented sooner, and potentially improve outcomes.

The NEPROCHECK test (Astute Medical, San Diego, CA, USA) was recently approved by the US Food and Drug Administration (FDA) as the first test to identify patients at high risk of developing AKI [6, 7]. This test measures urinary tissue inhibitor of metalloproteinases-2 (TIMP-2) and insulin-like

growth factor binding protein 7 (IGFBP7) [8]. Elevated urinary [TIMP-2]•[IGFBP7] predicts an increased risk of moderate–severe AKI over the next 12 h; patients with values >0.3 (ng/mL)²/1000 have a 7-fold increased risk of AKI [95% confidence interval (CI) 4–22] compared to those with lower levels [6].

Standardized, validated criteria exist for AKI. The Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guideline on AKI endorsed the use of modified Risk, Injury, and Failure; and Loss; and End-stage renal disease (RIFLE) criteria proposed by the AKI Network (AKIN) with minor modifications to include pediatric RIFLE for children [4, 9, 10]. However, despite well-established consensus criteria, AKI remains a clinical diagnosis [4]. Consequently, as part of the FDA registration process for the Nephrocheck test, a panel of nephrologists with expertise in AKI were convened to adjudicate AKI, rather than relying only on consensus criteria. To the best of our knowledge, this is the first time that AKI has been formally adjudicated in an FDA registration trial. The goal of the current study is to describe the adjudication process, to examine the relationship of [TIMP-2]•[IGFBP7] values to AKI as defined by clinical adjudication both when the determination was unanimous and when it was not and to examine biomarker results when clinical adjudication departed from consensus criteria.

MATERIALS AND METHODS

Patients and procedures

The methods and results of the Topaz trial have been reported previously [6]. Briefly, this was a prospective observational study of 420 critically ill adult patients enrolled within 24 h of admission to an intensive care unit (ICU) at 23 participating sites in the USA from May through December 2012. Paired serum and urine samples for analysis of serum creatinine and urinary [TIMP-2]•[IGFBP7] were obtained immediately upon enrollment and then stored frozen until analysis. The primary endpoint was the diagnosis of AKI within 12 h of enrollment adjudicated by a clinical adjudication committee. Urine samples were analyzed for TIMP-2 and IGFBP7 by technicians blinded to clinical data using a clinical immunoassay (NEPHROCHECK Test and ASTUTE140 Meter, Astute Medical) at three independent hospital laboratories. The test instrument automatically multiplied the concentrations of the two biomarkers together and divided this product by 1000 to report a single numerical test result with units of (ng/mL)²/1000 (the units for all [TIMP-2]•[IGFBP7] tests and cutoff values in this report). The median of the [TIMP-2]•[IGFBP7] test results from the three laboratories was used for analysis.

Adjudication process

The procedures for adjudication were defined in advance of the study. Specifically, the three adjudicators were all board-certified nephrologists with clinical expertise in AKI who determined as a group which variables would be provided to them for adjudication. Importantly, all adjudicators were unaware of the [TIMP-2]•[IGFBP7] values throughout the adjudication process. After a round of practice adjudication, the adjudicators met by teleconference to review the approach to the clinical data review and revisions were made. Variables included all

serum creatinine values for up to 6 months prior to and 72 h after enrollment, all hourly urine output data available for up to 24 h prior to and 72 h after enrollment, daily fluid balance use of diuretics and baseline demographic data including weight, age and clinical setting. However, results of urine microscopy or ultrasonography were not available to adjudicators. AKI was defined according to the KDIGO consensus criteria [4] as a doubling of serum creatinine from baseline or urine output <0.5 cm³/kg/h \times 12 h (Stage 2 or 3 AKI). The adjudicators were asked to determine whether AKI was present or absent (defined as ‘AKI’ or ‘no AKI’, respectively) within the 12 h after enrollment and sample collection based on these criteria.

Adjudicators were encouraged to use their expert judgment for each case. For example, an adjudicator could diagnose AKI in a patient who died prior to reaching KDIGO criteria or who received large volume resuscitation such that creatinine was judged to be less reliable. Conversely, an adjudicator might judge that AKI was not present even though creatinine or urine output criteria were met if extenuating circumstances were present. For example, if urine output criteria were only transiently met (for an hour) and then followed by a period of brisk urine output after diuretics, an adjudicator might judge that AKI was not present. Adjudicators indicated their diagnosis (‘AKI’ or ‘no AKI’) independently without consultation with each other and were blinded to biomarker results. For the current analysis we classified subjects as AKI 3/3, 2/3, 1/3 or 0/3 according to the proportion of adjudicators diagnosing each case as AKI. In other words, if all three adjudicators diagnosed a subject as having AKI, the subject was classified as AKI 3/3. If only one adjudicator diagnosed the subject as having AKI while the other two indicated no AKI, we classified this as AKI 1/3. Adjudicators were required to classify all cases as AKI or no AKI. For the purposes of the Topaz trial, a two-thirds majority was predefined for use as the final diagnosis, so subjects who were either AKI 2/3 or 3/3 were considered to have AKI. We also determined whether subjects were KDIGO AKI(+) or (–) when either Stage 2 or 3 criteria were met or not, respectively, by serum creatinine or urine output using a computer algorithm without any adjudicator input.

Statistical analysis

The association between rater concordance and AKI status was tested using the Fisher exact test. Fleiss’ kappa was used to assess the reliability of agreement between the three clinical adjudicators. Levels of [TIMP-2]•[IGFBP7] were compared in those subjects where there was discordance between the clinical adjudication committee and the KDIGO criteria using the Wilcoxon rank sum test. Subjects were then grouped based on the presence or absence of AKI based on the clinical adjudication committee review and on whether or not there was concordance between all three reviewers, and levels of [TIMP-2]•[IGFBP7] were compared using the Kruskal–Wallis test and Jonckheere–Terpstra trend test. To compare the correlations between [TIMP-2]•[IGFBP7] and clinically adjudicated or KDIGO AKI, the difference in two paired Somer’s D rank correlations was calculated using levels of [TIMP-2]•[IGFBP7] as the response variable and clinically adjudicated and KDIGO AKI as the two predictors. Finally, all P-values were two-sided and a P-value <0.05 was

considered statistically significant. Stata/SE 12.1 (StataCorp, College Station, TX, USA) or R 3.1.0 (R Project for Statistical Computing, Vienna, Austria) were used for all analyses.

RESULTS

Of 408 subjects enrolled with clinical data available, 71 had AKI (56 AKI 3/3 and 15 AKI 2/3) and 337 did not (9 AKI 1/3 and 328 AKI 0/3) based on the clinical adjudication committee review. There was complete agreement between the three clinical adjudicators (AKI 3/3 or 0/3) in 94.1% of cases. However, concordance of all three reviewers was lower for AKI than for no AKI (78.9 versus 97.3%; $P < 0.001$; Table 1). Interreviewer reliability ranged from 0.84 to 0.95 (Table 2), indicating excellent agreement. Overall, Fleiss' kappa was 0.86 (95% CI 0.80–0.92; $P < 0.001$) for all three adjudicators.

When KDIGO AKI criteria were applied, 72 patients were (+) and 336 patients were (–). Agreement between the clinical adjudication committee and the KDIGO criteria was 96.8% overall. Similar to the concordance between clinical reviewers, agreement between the clinical adjudication committee and the KDIGO criteria was lower for AKI than for no AKI (91.5 versus 97.9%; $P = 0.01$; Table 3); seven patients were KDIGO AKI(+) but did not have AKI by clinical adjudication and six were KDIGO AKI(–) but had AKI based on the majority opinion of the clinical adjudicators (AKI 3/3 or 2/3). Furthermore, for these 13 patients where there was not agreement between the clinical adjudication committee and the KDIGO criteria, concordance of all three

Table 1. Concordance between the clinical adjudication committee members, stratified by the presence or absence of AKI

	Concordant	Discordant	Total	Agreement
No AKI	328	9	337	97.3%
AKI	56	15	71	78.9%
Total	384	24	408	94.1%

To have concordance, all three members of the clinical adjudication committee had to agree on the diagnosis.

Table 2. Interreviewer reliability between the three clinical adjudication committee members and the majority opinion

	Member 1	Member 2	Member 3	Majority
Member 1	1			
Member 2	0.84 (0.77–0.91)	1		
Member 3	0.84 (0.76–0.91)	0.90 (0.85–0.96)	1	
Majority	0.89 (0.83–0.95)	0.95 (0.91–0.99)	0.95 (0.91–0.99)	1

Results are presented as reliability (95% CI).

Table 3. Proportion of subjects with AKI based on clinical adjudication (CAC) versus the KDIGO AKI consensus criteria

	KDIGO		Total	Agreement
	(–)	(+)		
CAC				
No AKI (1/3 or 0/3)	330 (80.9%)	7 (1.7%)	337 (82.6%)	97.9%
AKI (2/3 or 3/3)	6 (1.5%)	65 (15.9%)	71 (17.4%)	91.5%
Total	336 (82.4%)	72 (17.6%)	408 (100%)	96.8%

For the purposes of this study, those with KDIGO Stage 1 AKI were not considered to have AKI. Agreement was present if the clinical adjudication committee adjudicated that AKI was not present and the maximum KDIGO stage was 0 or 1 or if the clinical adjudication committee adjudicated that AKI was present and the maximum KDIGO stage was 2 or 3.

clinical adjudicators was much lower than for those where there was agreement (38.5 versus 95.9%; $P < 0.001$).

We next examined urine [TIMP-2]•[IGFBP7] in the 13 cases where clinical adjudication differed from KDIGO. As shown in Figure 1, the median [TIMP-2]•[IGFBP7] levels were lower in KDIGO AKI(+) patients without AKI by clinical adjudication [0.52 (IQR 0.26–1.64)] compared with KDIGO AKI(–) patients who had AKI according to the clinical adjudication committee [2.78 (IQR 2.33–3.56), $P = 0.008$]. Clinically adjudicated AKI correlated more closely with [TIMP-2]•[IGFBP7] than did KDIGO AKI ($P = 0.007$), as determined by Somer's D rank correlation.

Finally, [TIMP-2]•[IGFBP7] levels were compared between the four groups: AKI 3/3, 2/3, 1/3 and 0/3. As shown in Figure 2, the lowest [TIMP-2]•[IGFBP7] levels were observed in AKI 0/3 patients, while AKI 3/3 displayed the highest values [0.32 (IQR 0.16–0.83) versus 1.68 (0.80–2.87), $P < 0.001$]. Discordance among adjudicators (AKI 1/3 and 2/3) was observed with intermediate [TIMP-2]•[IGFBP7] values [0.52 (IQR 0.31–0.99) and 1.13 (0.47–2.27)]. Overall we observed increasing median [TIMP-2]•[IGFBP7] values from AKI 0/3 to 3/3 ($P < 0.001$ for overall trend). Similarly, when we examined the proportion of individuals in each of the four groups who had levels above the two cutpoints in clinical use (0.3 and 2.0), we observed an overall trend toward a greater number of people above the cutpoint as AKI status went from 0/3 to 3/3 ($P < 0.001$ for overall trend; Figure 3).

DISCUSSION

To our knowledge this is the first study to examine the relationships between a biomarker and agreement between members of a clinical adjudication committee and between the majority adjudicator opinion and KDIGO criteria for AKI. Our results are notable first because they demonstrate the importance of clinical adjudication in clinical studies of AKI. While our clinical adjudicators generally agreed with KDIGO, nearly 10% of the subjects judged to have AKI by the adjudicators were KDIGO AKI(–) and in these KDIGO AKI(–)/adjudicator AKI(+)

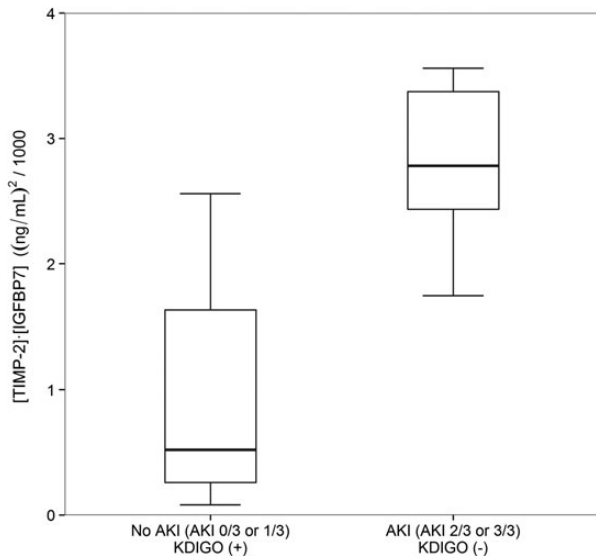


FIGURE 1: [TIMP-2]•[IGFBP7] levels in groups where there was discordance between the clinical adjudication committee (CAC) and the KDIGO criteria. Box and whiskers show IQRs and total observed ranges (censored by 1.5 times the box range), respectively. Seven cases were adjudicated as not AKI (AKI 0/3 or 1/3) but were KDIGO AKI(+), while six cases were adjudicated as AKI (AKI 2/3 or 3/3) but were KDIGO AKI(−). In cases of discordance between CAC adjudication and KDIGO criteria, patients adjudicated as AKI by the CAC had significantly higher levels of [TIMP-2]•[IGFBP7] than patients adjudicated as no AKI ($P = 0.008$).

cases, urinary [TIMP-2]•[IGFBP7] was always elevated, establishing a clear ‘biologic signal’. While our study only used data from an observational study, the results suggest that AKI interventional trials should be adjudicated as well since a substantial number of clinically relevant AKI cases may be missed with solely KDIGO criteria. Second, these results provide further evidence for the utility of urinary [TIMP-2]•[IGFBP7]. Even in cases where experts disagree with one another or with KDIGO criteria, the biomarker test provides further information. This is important because these are precisely the kinds of cases where biomarkers are needed most—both for clinical practice and for clinical studies [7]. For example, two clinicians may differ on whether AKI is present or not, and in these cases, biomarkers can provide incremental information to help solidify the diagnosis.

Of note, [TIMP-2]•[IGFBP7] is not a diagnostic test for AKI, but rather a ‘risk assessment’ test that identifies individuals at increased risk for AKI over the next 12 h. Higher levels portend greater risk for AKI [6] and for adverse long-term outcomes [11]. Consequently, we also sought to better understand the relationship of [TIMP-2]•[IGFBP7] and AKI, as identified by a committee of experts. Our hypothesis was that when experts agree, biomarker discrimination will be maximized. Our results bear this out.

We showed that agreement between the three members of the clinical adjudication committee was excellent, but not 100%. Not surprisingly, agreement was greater for non-AKI than for AKI. A review of study charts where there was discordance between reviewers (AKI 1/3 and 2/3) suggested

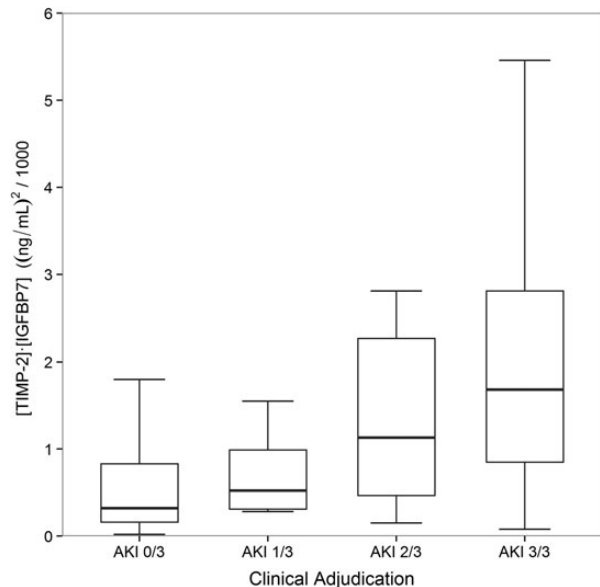


FIGURE 2: [TIMP-2]•[IGFBP7] levels by the proportion of adjudicators classifying each case as AKI. Box and whiskers show IQRs and total observed ranges (censored by 1.5 times the box range), respectively. [TIMP-2]•[IGFBP7] levels increased from AKI 0/3 to AKI 3/3 ($P < 0.001$ for test of trend).

that in those cases, subjects often met AKI criteria for a very short period of time early within the 12 h risk assessment window, just missed meeting urine output criteria for AKI because of transient brisk urine output or were obese (>150 kg), so urine output normalized to weight may have been variably interpreted by adjudicators. It should be noted that the adjudicator training required to achieve excellent consensus between three expert nephrologists was relatively straightforward and should be replicable in future clinical studies. Adjudicators underwent training that was composed of two teleconferences, with a round of practice cases that were reviewed and discussed on the second call. During the clinical trial, adjudicators had the clinical trial protocol and the training slides available as references.

Our review of the small number of KDIGO AKI(+), clinically adjudicated non-AKI (AKI 0/3 or 1/3) subjects revealed that KDIGO AKI(+) non-AKI was identified when subjects met AKI criteria for a very short period of time early within the risk assessment window or when the serum creatinine was falling but more than double the pre-morbid baseline. In such cases, the clinical adjudication committee determined that those patients did not have clinical AKI within the risk assessment window, yet the patients technically met KDIGO Stage 2–3 criteria.

Major strengths of our study include blinding of clinical adjudicators to biomarker measurements as well as detailed clinical data collected as part of the trial that allowed for hour-by-hour assessment of AKI. Limitations include the fact that we have not yet demonstrated that this approach can be extended beyond the three members of the clinical adjudication committee. However, as previously stated, the approach was predefined and training was standardized, so our approach should be applicable to future studies as well.

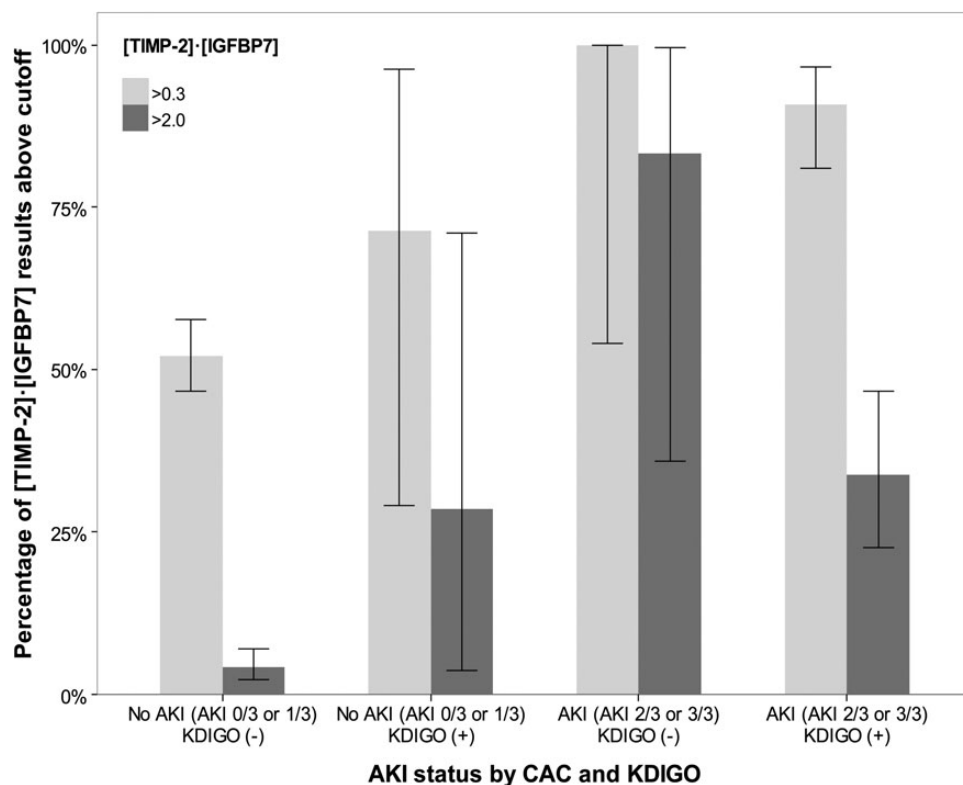


FIGURE 3: Proportion of [TIMP-2]•[IGFBP7] above the cutpoints of 0.3 and 2.0. Box and whiskers show proportion and 95% CI, respectively. For both cutpoints there was a trend toward a higher proportion of individuals having [TIMP-2]•[IGFBP7] levels above the cutpoint from AKI 0/3 to AKI 3/3 ($P < 0.001$ for test of trend).

In sum, we examined agreement between three nephrologists for diagnosis of AKI in the context of a trial of novel biomarkers focused on risk assessment for AKI. We demonstrated that with limited training that can be easily accomplished, agreement was excellent, but not universal, and clinical adjudication did not always agree with strict KDIGO criteria. Interestingly, biomarker levels were higher in those with clinically adjudicated AKI rather than those with AKI by consensus definition alone—in general [TIMP-2]•[IGFBP7] results supported the majority opinion of the experts. Our results highlight the importance of clinical adjudication committees in studies of AKI.

AUTHOR CONTRIBUTIONS

Study concept and design: K.D.L., A.V., M.H.R., L.S.C., J.A.K. Acquisition of data: K.D.L., A.V., M.H.R., L.S.C., J.A.K. Analysis and interpretation of data: K.D.L., J.S., J.A.K. Drafting of the manuscript: K.D.L., J.A.K. Critical revision of the manuscript for important intellectual content: A.V., M.H.R., L.S.C., J.A.K. Statistical analysis: J.S. Obtained funding: L.S.C., J.A.K. Study supervision: J.A.K.

ACKNOWLEDGEMENTS

The authors thank the investigators, staff and coordinators who worked on the clinical trial, as well as the study participants. This

work was presented in abstract form at the American Society of Nephrology Kidney Week 2013.

CONFLICT OF INTEREST STATEMENT

K.D.L., A.V., M.H.R. were members of the clinical adjudication committee and served as paid consultants to Astute Medical. J.S. has served as a paid consultant to Astute Medical. J.A.K. has received consulting fees and grant support from Astute Medical as well as royalties from licensing fees for unrelated technologies paid by Astute Medical to the University of Pittsburgh. L.S.C. has received consulting fees and grant support from Astute Medical as well as royalties from licensing fees for unrelated technologies paid by Astute Medical to George Washington University. The Topaz study was funded by Astute Medical.

REFERENCES

- Chertow GM, Burdick E, Honour M *et al.* Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. *J Am Soc Nephrol* 2005; 16: 3365–3370
- Hsu RK, McCulloch CE, Dudley RA *et al.* Temporal changes in incidence of dialysis-requiring AKI. *J Am Soc Nephrol* 2013; 24: 37–42
- Faubel S, Chawla LS, Chertow GM *et al.* Ongoing clinical trials in AKI. *Clin J Am Soc Nephrol* 2012; 7: 861–873
- KDIGO Acute Kidney Injury Work Group. KDIGO clinical practice guideline for acute kidney injury. *Kidney Int Suppl* 2012; 2: 1–138

5. Okusa MD, Molitoris BA, Palevsky PM *et al.* Design of clinical trials in acute kidney injury: a report from an NIDDK workshop—prevention trials. *Clin J Am Soc Nephrol* 2012; 7: 851–855
6. Bihorac A, Chawla LS, Shaw AD *et al.* Validation of cell-cycle arrest biomarkers for acute kidney injury using clinical adjudication. *Am J Respir Crit Care Med* 2014; 189: 932–939
7. Kellum JA, Chawla LS. Cell-cycle arrest and acute kidney injury: the light and the dark sides. *Nephrol Dial Transplant* 2016; 1: 16–22
8. Kashani K, Al-Khafaji A, Ardiles T *et al.* Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. *Crit Care* 2013; 17: R25
9. Mehta RL, Kellum JA, Shah SV *et al.* Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Crit Care* 2007; 11: R31
10. Bellomo R, Ronco C, Kellum J *et al.* Acute renal failure—definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative Group. *Crit Care* 2004; 8: R204–R212
11. Koyner JL, Shaw AD, Chawla LS *et al.* Tissue inhibitor metalloproteinase-2 (TIMP-2)-IGF-binding protein-7 (IGFBP7) levels are associated with adverse long-term outcomes in patients with AKI. *J Am Soc Nephrol* 2015; 26: 1747–1754

Received for publication: 8.1.2016; Accepted in revised form: 24.5.2016