

Characterization of renal biomarkers for use in clinical trials: biomarker evaluation in healthy volunteers

David A Brott¹
Scott H Adler¹
Ramin Arani²
Susan C Lovick³
Mark Pinches⁴
Stephen T Furlong¹

¹Translational Patient Safety and Enabling Sciences, AstraZeneca Pharmaceuticals, AstraZeneca Pharmaceuticals, Wilmington, DE, USA; ²AstraZeneca Pharmaceuticals, Wilmington, DE, USA; ³AstraZeneca Pharmaceuticals, AstraZeneca Pharmaceuticals, Macclesfield, Cheshire, UK

Background: Several preclinical urinary biomarkers have been qualified and accepted by the health authorities (US Food and Drug Administration, European Medicines Agency, and Pharmaceuticals and Medical Devices Agency) for detecting drug-induced kidney injury during preclinical toxicologic testing. Validated human assays for many of these biomarkers have become commercially available, and this study was designed to characterize some of the novel clinical renal biomarkers. The objective of this study was to evaluate clinical renal biomarkers in a typical Phase I healthy volunteer population to determine confidence intervals (pilot reference intervals), intersubject and intrasubject variability, effects of food intake, effect of sex, and vendor assay comparisons.

Methods: Spot urine samples from 20 male and 19 female healthy volunteers collected on multiple days were analyzed using single analyte and multiplex assays. The following analytes were measured: α -1-microglobulin, β -2-microglobulin, calbindin, clusterin, connective tissue growth factor, creatinine, cystatin C, glutathione S-transferase- α , kidney injury marker-1, microalbumin, N-acetyl- β -(D) glucosaminidase, neutrophil gelatinase-associated lipocalin, osteopontin, Tamm-Horsfall urinary glycoprotein, tissue inhibitor of metalloproteinase 1, trefoil factor 3, and vascular endothelial growth factor.

Results: Confidence intervals were determined from the single analyte and multiplex assays. Intersubject and intrasubject variability ranged from 38% to 299% and from 29% to 82% for biomarker concentration, and from 24% to 331% and from 10% to 67% for biomarker concentration normalized to creatinine, respectively. There was no major effect of food intake or sex. Single analyte and multiplex assays correlated with $r^2 \geq 0.700$ for five of six biomarkers when evaluating biomarker concentration, but for only two biomarkers when evaluating concentration normalized to creatinine.

Conclusion: Confidence intervals as well as intersubject and intrasubject variability were determined for novel clinical renal biomarkers/assays, which should be considered for evaluation in the next steps of the qualification process.

Keywords: clinical, drug development, biomarkers, kidney, healthy volunteers, qualification

Introduction

Drug-induced kidney injury (DIKI) is recognized to occur throughout the drug research and development process, with histology considered to be the gold standard for pre-clinical screening. Routine laboratory tests for creatinine and blood urea nitrogen are used as biomarkers for DIKI, but are considered insensitive since more than 50% of functional kidney loss occurs prior to any significant biomarker change.^{1,2} This has led to the development of additional renal biomarkers, some of which have been qualified for detecting DIKI in the preclinical setting. Qualification of these biomarkers was

Correspondence: Stephen T Furlong
AstraZeneca Pharmaceuticals,
1800 Concord Pike, D3C-421,
Wilmington, DE 19850, USA
Tel +1 302 886 8588
Email stephen.furlong@astrazeneca.com

accomplished in a stepwise process involving biomarker candidate identification, assay validation, determining levels in urine of naïve rats, and then qualification in rats after nephrotoxicant treatment.^{3–5} Each of these biomarkers was anchored to DIKI using histopathology by verifying the presence of the marker within the kidney as well as determining if biomarker expression was increased and/or staining loss was associated with DIKI.

In the clinical setting, DIKI is more difficult to detect, and reliance on changes in renal function through measurement of creatinine and blood urea nitrogen has been in routine standard use. However, measurement of serum creatinine and blood urea nitrogen is considered to be insensitive for clinical monitoring of DIKI.^{6,7} Unlike the preclinical setting, it is not typically possible to anchor a biomarker to histopathology, but clinical DIKI biomarker qualification still requires a stepwise process entailing candidate biomarker identification and assay validation, followed by a process of establishing confidence intervals (CIs, pilot reference intervals) and subject variability in healthy volunteers, patients with normal renal function treated with known nephrotoxicants, such as oncology patients treated with cisplatin, patients with specific underlying diseases for which a drug is being developed, and finally in patients with underlying renal disease.⁸

The initial step in identifying candidate renal biomarkers was based on biomarkers under investigation in the hospital setting for critically ill patients and translation of the qualified preclinical biomarkers.^{3,4,9–16} Human assays for many of these candidate biomarkers have become commercially available and in this study we set out to characterize such biomarkers in healthy volunteers. Table 1 lists the human biomarkers under characterization within this study, along with some preclinical and/or clinical rationale for inclusion.

The objective of this study was to characterize the renal biomarkers α -1-microglobulin (A1M), β -2-microglobulin (B2M), calbindin, clusterin, connective tissue growth factor (CTGF), cystatin C, glutathione S-transferase alpha (GST α), kidney injury marker-1 (KIM-1), N-acetyl- β -(D) glucosaminidase (NAG), neutrophil gelatinase-associated lipocalin (NGAL), osteopontin, Tamm-Horsfall urinary glycoprotein (THP), tissue inhibitor of metalloproteinase 1 (TIMP-1), trefoil factor 3 (TFF3), and vascular endothelial growth factor (VEGF) in a healthy volunteer population typical of that included in Phase I clinical trials in order to determine CIs (pilot reference intervals), intersubject and intrasubject variability, effect of food intake, effect of sex, and vendor assay comparisons.

Material and methods

Clinical study

The study was performed in accordance with the ethical principles set out in the Declaration of Helsinki and Good Clinical Practice, approved by the appropriate internal review boards, and consisted of 20 males and 19 females aged 18–70 years. The inclusion criteria were similar to those for typical Phase I clinical studies: signed informed consent; body mass index 18–30 kg/m² (inclusive); healthy as judged by an acceptable medical record; acceptable laboratory values and vital signs (blood pressure and pulse rate); negative serum hepatitis B surface antigen, hepatitis C antibody and human immunodeficiency virus status; and a negative pregnancy test for female subjects of child-bearing potential.

Exclusion criteria consisted of: any clinically significant illness within 2 weeks; use of any prescription medication within 2 weeks or nonprescription medication (other than paracetamol) within one week; significant history of alcohol abuse or consumption greater than 28 units/week in males or 21 units in females, with one unit equaling a half-pint of beer, one glass of wine, or one measure of spirits; history of drug abuse or a positive drug abuse test; or involvement in another investigational medicinal project within 4 months.

Restrictions during the study included: no drinking alcohol for 48 hours prior to study initiation until conclusion of study; no exercise for 72 hours prior to study initiation until conclusion of study; no foods containing poppy seeds; must eat within 2 hours prior to urine collection on days 1 and 2; and no food or drink (except water) from midnight before morning samples on day 3 until after urine collection.

Prestudy and demographic measurements included date of birth, sex, ethnicity, body mass index, height, weight, medical status questionnaire, blood pressure, and pulse rate. At approximately 8 am on three consecutive days, spot urine was collected (up to 120 mL) and placed on ice as soon as possible after voiding. Up to 50 mL of urine was centrifuged at 2,000 \times g at 4°C for 10 minutes, the supernatant collected, a biotrin urine stabilizer added to the supernatant at a ratio of 1 to 4, with subsequent aliquoting into 1.5 mL labeled tubes and freezing at –80°C. Separate sample aliquots were sent to Pacific Biomarkers (Seattle, WA, USA), Rules Based Medicine (Austin, TX, USA) and a central laboratory at AstraZeneca for analysis. At the central laboratory, urine was thawed and analyzed using single analyte assays for creatinine, NAG, osmolality, potassium, sodium, and total protein. At Pacific Biomarkers, urine was thawed and evaluated using single analyte assays for B2M, creatinine, cystatin C, KIM-1, NAG, and NGAL. At Rules Based Medicine, urine was thawed and analyzed using

Table 1 Current status and rationale of biomarkers included in the current study

| Parameter | Preclinical | Clinical |
|--------------|---|--|
| AIM | | Increased in AKI patients post cardiopulmonary bypass ³⁵ |
| B2M | Increased with gentamicin sulfate or cisplatin treatment in rats ³⁶ | Increased with fumaric acid treatment for psoriasis ³⁷ |
| Calbindin | Increased with gentamicin sulfate treatment in rats ³⁶ Increased with a triple reuptake inhibitor in monkeys ³⁸ | Increased with cisplatin treatment ³⁴ |
| Clusterin | Increased with gentamicin sulfate, 2-bromoethylamine, or cisplatin treatment in rats ³⁶ Increased with a triple reuptake inhibitor in monkeys ³⁸ Increased with n-phenylanthranilic acid ³⁹ Had a higher diagnostic value than BUN and serum creatinine ⁴⁰ | |
| CTGF | Increased in diabetic rats ⁴¹ | Increased in diabetics with nephropathy and a small portion of diabetics without nephropathy ⁴¹ |
| Cystatin C | Increased with gentamicin sulfate, 2-bromoethylamine, or cisplatin treatment in rats ³⁶ Had a higher diagnostic value than BUN and serum creatinine ⁴⁰ Increased with paraquat treatment ⁴² | Increased in patients with AKI compared with healthy volunteers ³³ Predicted AKI in sepsis critically ill patients ⁴³ |
| GST α | Increased with gentamicin sulfate, 2-bromoethylamine, or cisplatin treatment in rats ³⁶ Increased with cisplatin treatment ⁴⁴ | Increased in subjects treated with amphotericin B desoxycholate ⁴⁵ |
| KIM-1 | Increased with gentamicin sulfate, 2-bromoethylamine, or cisplatin treatment in rats ³⁶ Increased with paraquat treatment ⁴² Increased with cisplatin treatment ⁴⁴ | Increased in patients with AKI compared with healthy volunteers ³³ |
| Microalbumin | Increased with cisplatin treatment ⁴⁴ | Increased in AKI patients post cardiopulmonary bypass ³⁵ |
| NAG | Increased with cisplatin treatment ⁴⁴ | Increased in patients with AKI compared to healthy volunteers ³³ Increased in AKI patients post cardiopulmonary bypass ³⁵ |
| NGAL | Increased with gentamicin sulfate, 2-bromoethylamine, or cisplatin treatment in rats ³⁶ | Increased in patients with AKI compared with healthy volunteers ³³ Increased in AKI patients post cardiopulmonary bypass ³⁵ |
| Osteopontin | Increased with gentamicin sulfate, 2-bromoethylamine, or cisplatin treatment in rats ³⁶ Increased with cisplatin treatment ⁴⁴ | Subjects with AKI had higher values than subjects without AKI ⁴⁶ |
| THP | | Neonates with AKI had lower levels of THP (uromodulin) ⁴⁷ |
| TIMP-1 | | Is higher in patient with renal disease than controls ⁴⁸ |
| TFF3 | Decreased in rats treated with cisplatin ⁴⁹ | Higher levels may indicate ongoing repair in the kidney ⁵⁰ |
| VEGF | | Increased in patients with AKI compared with healthy volunteers ³³ |

Abbreviations: AIM, alpha-1-microglobulin; AKI, acute kidney injury; B2M, β 2-microglobulin; BUN, blood urea nitrogen; CTGF, connective tissue growth factor; GST α , glutathione S-transferase alpha; KIM-1, kidney injury marker-1; NAG, N-acetyl- β -glucosaminidase; NGAL, neutrophil gelatinase-associated lipocalin; THP, Tamm-Horsfall urinary glycoprotein; TIMP-1, tissue inhibitor of metalloproteinase 1; TFF3, trefoil factor 3; VEGF, vascular endothelial growth factor.

multiplex assay(s) for A1M, B2M, calbindin, clusterin, CTGF, creatinine, cystatin C, GST α , KIM-1, microalbumin, NGAL, osteopontin, THP, TIMP-1, TFF3, and VEGF.

At approximately 8 am on the days of urine collection, 2.7 mL of blood was collected in a lithium-heparin anti-coagulated tube, processed to obtain plasma, and frozen until analyzed. At the central laboratory, plasma was thawed

and analyzed for creatinine, potassium, sodium, osmolality, and total protein.

Data analysis

Summary statistics were assessed for biomarkers in each corresponding database (biomarker concentration and concentration normalized to creatinine). Robust regression, based on *M* esti-

mation,¹⁷ was utilized to examine potential trends across days (visits). Subsequently, potential outliers were identified and omitted in the summary statistics provided. A nonparametric approach using sample quantiles was used to calculate 95% CIs for each biomarker.¹⁸ In addition, the intersubject coefficient of variation (%CV) was calculated to assess the consistency of biomarker values across subjects, using samples collected on three consecutive days.¹⁹ Similarly, the intrasubject %CV was calculated to assess the consistency of each biomarker across visits (within each subject), using the samples collected on three consecutive days. *P*-values were adjusted for multiplicity due to the number of biomarkers using the Benjamini and Hochberg false discovery rate method.²⁰

Effect of sex was calculated using the *t*-test. *P*-values were adjusted for multiplicity due to the number of biomarkers using the Benjamini and Hochberg false discovery rate method.²⁰ The coefficient of determination (r^2) using only day 1 samples was determined by linear regression to assess the correlation between two assays for the same biomarker. All analyses were carried out using SAS® version 9.3 (SAS Inc, Cary, NC, USA).

Results

Demographics

Except for age and inclusion of women of child-bearing potential, routine healthy volunteer inclusion and exclusion criteria were used for this study, that included 19 female and 20 male subjects with mean values of 43.9 years and 24.4 kg/m² for age and body mass index, respectively (Table 2). The ethnicity of 38 of the 39 study subjects was Caucasian, while one subject was of mixed Caucasian/Arab ethnicity.

Summary statistics

Confidence intervals as well as intersubject and intrasubject variability were calculated for each of the biomarkers using

Table 2 Study demographics

| Baseline characteristic | Biomarker population (n=39) |
|--------------------------------------|-----------------------------|
| Age (years) | |
| Mean, (SD) median | 43.9 (12.37) 43 |
| Body mass index (kg/m ²) | |
| Mean (SD) | 24.4 (2.65) |
| Sex, n (%) | |
| Female | 19 (49) |
| Male | 20 (51) |
| Race | |
| Non-Hispanic Caucasian | 38 (97.4) |
| Other | 1 (2.6) |

Abbreviation: SD, standard deviation.

the samples collected on days 1, 2, and 3 for routine plasma and urine kidney biomarkers (Table 3) and the urine biomarkers under characterization (Table 4).

Urinary biomarker analysis was evaluated using concentration as well as the concentration value normalized to creatinine. In many cases, normalization decreased intersubject and/or intrasubject variability, but in a few cases the variability was increased (eg, for CTGF and microalbumin). There was a wide range of intersubject variability, from 38% to 299% for biomarker concentration values and 24% to 331% for concentration normalized to creatinine. The intrasubject variability ranged from 29% to 82% for biomarker concentration values and from 10.0% to 67% for concentration normalized to creatinine (Table 4).

Effect of food intake

The influence of fed versus fasted state was based on the subjects being fed (eating within 2 hours of blood/urine collection) for the samples collected on days 1 and 2 and then fasted (not eating within approximately 8 hours) prior to sample collection on day 3. There were no significant differences ($P>0.05$) in any of the biomarker values between the fed and fasted states. The lack of a food intake effect supports combining the day 1, 2, and 3 data to calculate intersubject and intrasubject variability.

Effect of sex

Given that sex is known to influence reference intervals for the routine kidney biomarker creatinine, with male intervals being higher than females, the urine biomarkers under characterization were evaluated for effect of sex (Table 5). There were two biomarkers under characterization in this study that did not have a sex influence using concentration or concentration normalized to creatinine, ie, microalbumin and TIMP-1. The biomarker A1M had an effect of sex using concentration but not when using concentration normalized to creatinine. Most biomarkers had higher values in males for concentration data and higher values in females for concentration normalized to creatinine data. Exceptions included GST α , NGAL, and TFF3, which had higher values in females with both data sets, and calbindin, which had higher values in males with both data sets.

Assay correlations (single analyte versus multiplex assay)

It is also important to compare assays between vendors (Figure 1A). The three creatinine assays showed very good correlations ($r^2>0.98$). In spite of the good creatinine

Table 3 Summary statistics of plasma and urine samples analyzed with routine kidney biomarkers in the current study

| Parameter | Plasma | | Urine | |
|---------------|---------------------|--|---------------------|--|
| | Confidence interval | %CV _{inter} / ^o %CV _{intra} | Confidence interval | %CV _{inter} / ^o %CV _{intra} |
| Creatinine | 73.5–112.5 μmol/L | 11/6 | 1.09–12.46 mmol/L | 73/38 |
| Osmolality | 287–307 mOs/kg | 1/2 | 79.28–793.94 mOs/kg | 60/37 |
| Potassium | 3.49–4.71 mmol/L | 4/7 | 5.77–63.26 mmol/L | 68/46 |
| Sodium | 140.6–145.4 mmol/L | 1/2 | 10.97–99.41 mmol/L | 78/42 |
| Total protein | 66.4–80.8 g/L | 3/5 | 3.32–58.74 mg/L | 147/56 |

Abbreviations: %CV_{inter}, intersubject coefficient of variation; %CV_{intra}, intrasubject coefficient of variation.

correlations, the other biomarkers had better correlation when comparing concentration values rather than concentration normalized to creatinine. The single analyte NAG assays had a raw value correlation of $r^2 \geq 0.700$ with concentration data, but the correlation was only $r^2 = 0.224$ with creatinine-normalized data. Assay comparisons for B2M, cystatin C, KIM-1, and NGAL were based on a single analyte and multiplex assay. Cystatin C, KIM-1, and NGAL had correlations of $r^2 \geq 0.700$ with concentration data, but only KIM-1 and NGAL had $r^2 \geq 0.700$ with concentration

normalized to creatinine. Plots of each of the biomarkers that had correlations of $r^2 \geq 0.700$ are shown in Figure 1B–I.

Discussion

Safety biomarker qualification for clinical drug development is a fit-for-purpose, stepwise process from candidate biomarker identification to acceptance by health authorities.⁸ This study was designed to address several questions within the early healthy volunteer qualification step by characterizing the candidate clinical DIKI biomarkers in a

Table 4 Summary statistics of urine samples analyzed with renal biomarkers under characterization within the current study

| Parameter | Concentration | | Concentration normalized to creatinine | |
|---|---------------------|--|--|--|
| | Confidence interval | %CV _{inter} / ^o %CV _{intra} | Confidence interval | %CV _{inter} / ^o %CV _{intra} |
| Single analyte assays (central laboratory) | | | | |
| Creatinine | 1.09–12.46 mmol/L | 73/38 | NC | NC |
| NAG | 0.05–2.24 U/L | 117/62 | NC | NC |
| Single analyte assays (Pacific Biomarkers) | | | | |
| B2M | 20–83 μg/L | 50/29 | 0.04–0.15 μg/mg | 43/29 |
| Creatinine | 0.06–1.29 mg/mL | 74/44 | NC | NC |
| Cystatin C | 0.003–0.042 mg/L | 68/40 | 21.3–77.2 ng/mg | 33/10 |
| KIM-1 | 15–249 pg/mL | 65/43 | 50.3–539.7 pg/mg | 50/17 |
| NAG | 0.34–2.4 U/L | 62/37 | 0.001–0.003 U/mg | 32/17 |
| NGAL | 0.3–150.3 ng/mL | 299/50 | 1.9–580.3 ng/mg | 325/38 |
| Multiplex assays (Rules Based Medicine) | | | | |
| AIM | 0.12–3.00 μg/mL | 109/43 | 0.45–4.58 μg/mg | 63/23 |
| B2M | 0.01–0.09 μg/mL | 51/43 | 0.03–0.18 μg/mg | 40/21 |
| Calbindin | 0.31–114.0 ng/mL | 158/82 | 1.6–133.3 ng/mg | 107/60 |
| Clusterin | 0.00–0.07 μg/mL | 74/54 | 0.01–0.15 μg/mg | 75/24 |
| CTGF | 0.20–1.00 ng/mL | 41/33 | 0.19–2.08 ng/mg | 58/32 |
| Creatinine | 9.9–144.0 mg/dL | 58/40 | NC | NC |
| Cystatin-C | 3.2–75.0 ng/mL | 65/50 | 17.5–70.4 ng/mg | 27/19 |
| GSTα | 1.1–64.0 ng/mL | 78/61 | 1.8–169.0 ng/mg | 94/56 |
| KIM-1 | 0.05–0.48 ng/mL | 49/39 | 0.11–0.90 ng/mg | 42/15 |
| Microalbumin | 0.4–62.0 μg/mL | 260/47 | 1.9–258.3 μg/mg | 331/33 |
| NGAL | 3–254 ng/mL | 233/48 | 6.4–271.2 ng/mg | 264/45 |
| Osteopontin | 66–1,230 ng/mL | 66/50 | 236–1,459 ng/mg | 36/18 |
| THP | 0.23–1.90 μg/mL | 38/40 | 0.75–4.17 μg/mg | 38/20 |
| TIMP-1 | 0.3–10.0 ng/mL | 108/72 | 0.34–13.09 ng/mg | 114/67 |
| TFF3 | 0.03–3.10 μg/mL | 205/43 | 0.09–9.06 μg/mg | 230/32 |
| VEGF | 105–879 pg/mL | 52/35 | 456–1,286 pg/mg | 24/15 |

Abbreviations: NC, not calculated; AIM, alpha-1-microglobulin; B2M, β-2-microglobulin; CTGF, connective tissue growth factor; GSTα, glutathione S-transferase alpha; KIM-1, kidney injury marker-1; NAG, N-acetyl-β-glucosaminidase; NGAL, neutrophil gelatinase-associated lipocalin; THP, Tamm-Horsfall urinary glycoprotein; TIMP-1, tissue inhibitor of metalloproteinase 1; TFF3, trefoil factor 3; VEGF, vascular endothelial growth factor; %CV_{inter}, intersubject coefficient of variation; %CV_{intra}, intrasubject coefficient of variation.

Table 5 Effect of sex (20 males and 19 females) on the renal biomarkers under characterization within this study

| Concentration | | Concentration normalized to creatinine | |
|---------------|-----------------------|--|-----------------------|
| Parameter | Sex with higher value | Parameter | Sex with higher value |
| AIM | Male | | |
| B2M | Male | B2M | Female |
| Calbindin | Male | Calbindin | Male |
| Clusterin | Male | Clusterin | Female |
| CTGF | Male | CTGF | Female |
| Creatinine | Male | | |
| Cystatin C | Male | Cystatin-C | Female |
| GST α | Female | GST α | Female |
| KIM-1 | Male | KIM-1 | Female |
| NGAL | Female | NGAL | Female |
| Osteopontin | Male | Osteopontin | Female |
| TFF3 | Female | TFF3 | Female |
| THP | Male | THP | Female |
| VEGF | Male | VEGF | Female |

Abbreviations: AIM, alpha-I-microglobulin; B2M, β -2-microglobulin; CTGF, connective tissue growth factor; GST α , glutathione S-transferase alpha; KIM-1, kidney injury marker-1; NAG, N-acetyl- β -glucosaminidase; NGAL, neutrophil gelatinase-associated lipocalin; THP, Tamm-Horsfall urinary glycoprotein; TFF3, trefoil factor 3; VEGF, vascular endothelial growth factor.

typical Phase I healthy volunteer population to determine: CIs (pilot reference intervals), intersubject and intrasubject variabilities, effect of food intake, effect of sex, and vendor assay comparison. This early step in the biomarker clinical qualification process is designed to begin translation of the preclinical biomarker qualification data into the clinical setting, but care must be taken not to make too many assumptions concerning the candidate biomarkers when developing the study plan, such as subject population, data analysis/reporting, and potential applicability of different assay platforms (single analyte versus multiplex).

The healthy volunteer step in the qualification process is designed to do the initial characterization of the biomarkers as well as begin to understand conditions that can influence biomarker levels, such as food and sex. The subjects in the current study were deemed appropriate for qualification of the biomarkers in this step, based on the similarity of the study demographics, inclusion criteria, and exclusion criteria to those of typical clinical Phase I healthy volunteer studies.²¹

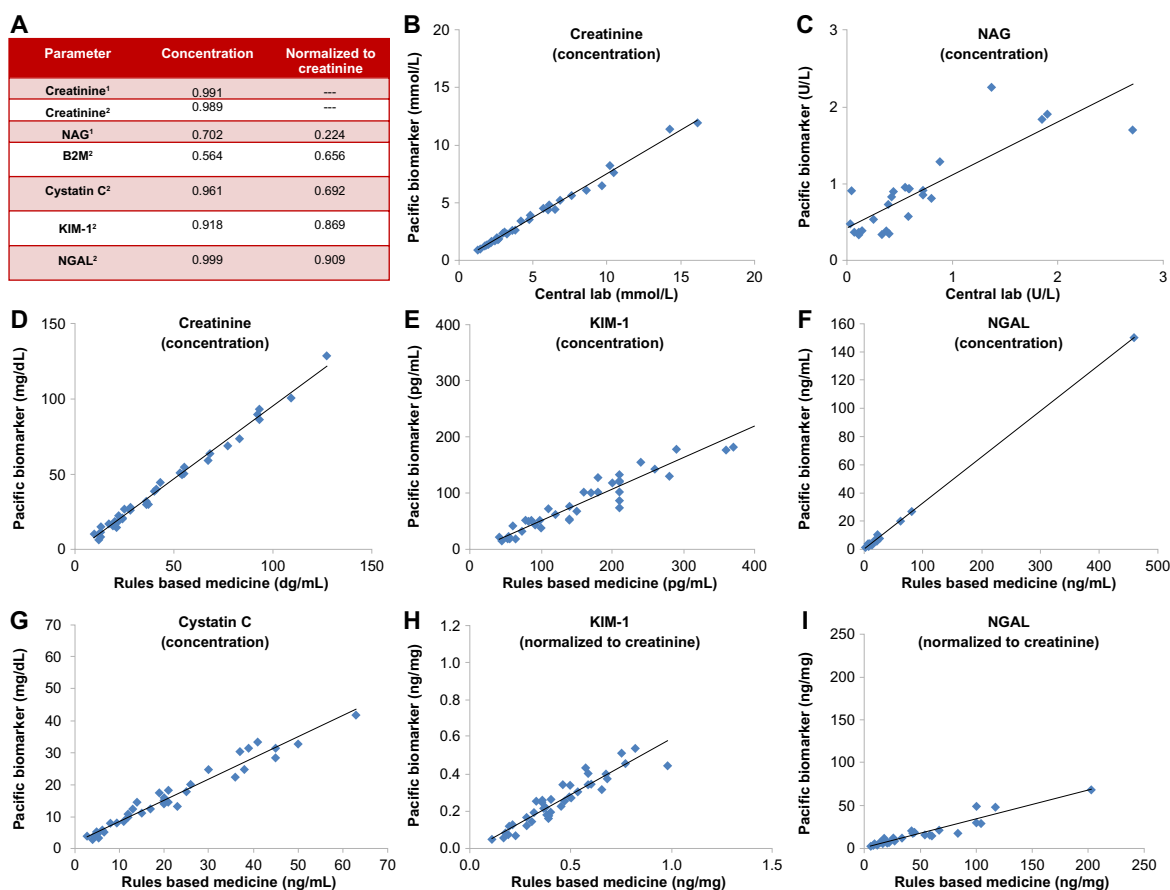


Figure 1 Plots of vendor assay correlations. (A) Coefficient of determination (r^2) for each biomarker with assays from two vendors. (B–I) are biomarkers with $r^2 \geq 0.700$ (B) Creatinine concentration, (C) NAG concentration, (D) creatinine concentration, (E) KIM-1 concentration, (F) NGAL concentration, (G) cystatin C concentration, (H) KIM-1 concentration normalized to creatinine, and (I) NGAL concentration normalized to creatinine.

Notes: ¹Comparison between Central and Pacific Biomarker assays, both single analyte assays; ²comparison between Pacific Biomarker (single analyte) and Rules Based Medicine (multiplex) assays.

Abbreviations: KIM-1, kidney injury marker-1; NAG, N-acetyl- β -glucosaminidase; NGAL, neutrophil gelatinase-associated lipocalin.

In addition, the routine kidney biomarker values in this study were similar to the Mass General Hospital reference intervals,²² assuming a urine excretion of approximately 2 L per day.

Urinalysis data can be reported in several ways, ie, total amount excreted over 24 hours, concentration, concentration normalized to urinary creatinine concentration, and/or concentration normalized to urinary specific gravity.^{23–27} When developing this study protocol, we needed to consider the feasibility of different sampling methods for application in typical clinical studies. Typical Phase I clinical studies evaluate spot samples. The most common reporting methods used for spot urine parameters in clinical trials are concentration and concentration normalized to creatinine, with the reporting method being dependent on the mechanism(s) of biomarker excretion.²⁸ Summary statistics for both of these reporting methods were determined in this study since the mechanism(s) of excretion was not known for each biomarker. Limitations of reporting normalized to creatinine values include creatinine excretion being age-dependent and sex-dependent, so separate age and sex reference ranges could be needed due to creatinine normalization and not due to the candidate biomarker. Limitations of reporting concentration values include altered water excretion and water intake; for example, biomarker concentration values will be higher with a compound that induces oliguria and lower with a compound that induces polyuria, even though the compound did not alter biomarker excretion.

This study determined both concentration and concentration normalized to creatinine CIs and biological variability for the candidate renal A1M, B2M, calbindin, clusterin, CTGF, cystatin C, GST α , KIM-1, microalbumin, NAG, NGAL, osteopontin, THP, TIMP-1, TFF3, VEGF, and NAG biomarkers. The CIs are considered pilot reference intervals due to the study design being fit-for-purpose. Bona fide reference intervals for use in the hospital setting and clinical development require between 120 and 400 subjects per group,^{29,30} but the data from this study support the early steps in biomarker qualification, ie, determining if the candidate biomarker is acceptable for and will aid interpretation of the next biomarker qualification steps, which include evaluation in subjects treated with known nephrotoxicants and disease populations. The CIs will be used to understand the expected predose values for the studies evaluating known nephrotoxicants in subjects, and the intersubject and intrasubject variability from this study will help determine how much change is required before a compound would be deemed a nephrotoxicant, as well as

provide a basis for power calculations for determining the appropriate size of these studies.

Some of the biomarkers in this study (A1M, KIM-1, NGAL, cystatin C, NAG, and VEGF) have been evaluated in various hospital populations including normal controls.^{16,23,26,28,31–33} Normal controls in the hospital setting are different from the subjects used in typical Phase I clinical trials. For example, normal control subjects used in the study reported by Vaidya et al³³ had three exclusion criteria, ie, recent hospitalization, diagnosis of chronic kidney disease, and treatment with nephrotoxic medications. Phase I clinical studies would not limit the exclusion criteria to only recent nephrotoxic medications, but would extend it to most medications. The exclusion criteria in Phase I clinical studies also entail alcohol and drug abuse as well as use of most nonprescription medications. Since clinical Phase I healthy volunteers are a subset of the hospital normal controls, the current study should have similar or lower intersubject variability (narrower CIs) than in the literature.³² Of the five biomarkers evaluated in the current study and reported in the literature, KIM-1 and VEGF had similar CIs, while NGAL, cystatin C, and NAG had narrower CIs in the current study.^{16,23,26,32,33} The narrower CIs (smaller intersubject variability) should aid in detecting DIKI, if the biomarker(s) continue to pass the qualification process.

The intrasubject variability determined in this study will aid data interpretation in the future steps of the biomarker qualification process, such as during evaluation of biomarkers in subjects treated with known nephrotoxicants. For example, KIM-1 and calbindin had intrasubject variability of 15% and 60%, respectively. Therefore, a smaller difference between predose and postdose values of KIM-1 would be needed to classify a compound as nephrotoxic compared with calbindin. However, there was an approximately 30-fold difference in calbindin levels (normalized to creatinine) between healthy volunteers and subjects treated with cisplatin,³⁴ that is well above the 60% intrasubject variability. Thus, none of the biomarkers should be excluded from progressing into the next steps of biomarker qualification based on intrasubject variability.

There was no significant difference in any of the parameters (urinary creatinine or candidate biomarkers) between fasted and fed subjects. This suggests that care may not be required when spot urine is collected in relationship to eating. However, this study was designed to determine if there was a major effect of food intake on the candidate biomarkers, and the breakfast food eaten by the subjects was not regulated. Further studies should be done later in the biomarker qualification process to determine if large amounts of food, high

protein meals (lunch or dinner), and/or specific foods will affect the candidate biomarkers.

Most of the biomarkers under characterization were influenced by sex, with values being higher in males for the concentration data sets and higher in females for concentration normalized to creatinine data sets. Exceptions included: microalbumin and TIMP-1 not being affected by sex; AIM being higher in males for the concentration data set and not influenced by sex in the normalized to creatinine data set; calbindin being higher in males in both data sets; and GST α , NGAL, and TFF3 being higher in females in both data sets. Pennemans et al²⁶ evaluated the effect of sex (199 women and 139 men) for cystatin C, KIM-1, and NGAL. In the current study and the literature,²⁶ NGAL had higher values in females in both the concentration data and normalized to creatinine data. However, in the literature,²⁶ cystatin C and KIM-1 were not influenced by sex when evaluating normalized to creatinine data. This is different from the current study, and is probably due to the lower subject number in the current study. Therefore, additional data will be needed later in the qualification process to determine the effect of sex on the biomarkers under characterization.

The stepwise approach to biomarker qualification is easiest when one assay (vendor) is consistently used for each biomarker, but this approach does raise some concerns: what if the antibody clones used by the vendor become unavailable either during or after the biomarker qualification process, and is the stepwise process qualifying the biomarker or is it more specifically qualifying the assay? In order to ensure qualification of the biomarker, it is important to at least periodically compare different vendor assays. The current study compared assays between vendors for creatinine, B2M, cystatin C, KIM-1, NAG, and NGAL. This study not only compared vendor assays, but also compared single analyte and multiplex vendor assays for B2M, cystatin C, KIM-1, and NGAL. It was no surprise that single analyte urine creatinine assays correlated between vendors because creatinine is a routine clinical chemistry parameter that has been used for many years within the hospital setting as well as during drug development. B2M was the only biomarker with poor correlation between the single analyte and multiplex assays for both the concentration and normalized to creatinine values, which could be due to different antibodies, protocols, and techniques used for the two assays and/or a low signal-to-noise ratio within both assays. Little can be done to improve the correlations if the poor correlation was due to antibodies, protocols, and/or techniques, but the next steps of qualification should increase the correlation if the differences were

due to the signal-to-noise ratio of the assays. For example, the B2M values are near the lower end of the standard curve for both the single analyte and multiplex assays, where the variation would be highest (lower signal-to-noise ratio) than for samples with medium or high B2M values. Adding values higher on the standard curve to the already collected samples could increase the correlation between assays. Therefore, further investigation between the B2M assays is needed prior to final qualification, but this can be addressed in the next qualification steps, provided that both assays are evaluated in the next qualification steps.

Of interest was the finding that concentration values for each assay typically had better correlation than the creatinine normalized values. The decreased correlation when using creatinine-normalized data could be due to variability of the normalizing biomarker (ie, creatinine), the fact that creatinine normalization typically narrows the confidence interval, or the multiplex platform contributing to the lack of correlation.

Creatinine normalization narrowing the CI is exemplified by the CI for cystatin C going from a 14-fold difference between the high and low CI values for the concentration data to a four-fold difference for the creatinine-normalized data. All of the creatinine-normalized parameters that had poor assay correlations (NAG, B2M, and cystatin C) also had a less than a five-fold difference between the high and low CI values in the single analyte Pacific Biomarker assay. Further assay comparison of values outside of the current CIs (eg, subjects treated with known nephrotoxicants) may help determine the cause of decreased assay correlations when using the creatinine-normalized data, but until that data are generated, it may be best to use the concentration data for primary analysis of assay comparisons and the normalized data for secondary analysis.

There is the potential for nonspecific binding or interference of analyte detection when assays are multiplexed, and this would decrease the correlation between single analyte and multiplex assays. This possibility cannot be ruled out for B2M, where there was poor correlation for both the concentration and normalized to creatinine values. However, this is probably not true when there is an acceptable correlation ($r^2 \geq 0.700$) with concentration values and a poor correlation with creatinine-normalized values, such as with NAG and cystatin C. Both NAG assays were single analyte assays, whereas the cystatin C correlation was between a single analyte and a multiplex assay.

A finding of note from the vendor assay comparisons was that the concentration values for cystatin C, KIM-1,

and NGAL multiplex assays correlated with the single analyte assays. There continues to be debate as to the value of multiplex versus single analyte assays. For many, there is a concern using multiplex assays when only one or two single analyte assays would be needed in a study. In this scenario, the multiplex assay would produce unneeded data. In spite of the debate, this study determined that multiplex and single analyte assays had similar intersubject and intrasubject variability, so both assay platforms pass the healthy volunteer step of biomarker qualification and both assay platforms should be further considered in future steps of biomarker qualification.

If the multiplex assays continue to pass the various biomarker qualification steps then the debate may ultimately be influenced by the translational relevance of the preclinical safety data. For example, if the preclinical biomarker data consistently predict which biomarker to use in clinical studies, then the specific single analyte assays could be deemed optimal. However, if preclinical studies only predict potential clinical nephrotoxicity, but not which specific biomarker to use in clinical studies, then multiplex assays could be advantageous in clinical studies.

Even if the future qualification studies determine that single analyte assays are preferred for clinical safety studies, multiplex assays may have value for use in some efficacy studies; an example could be to monitor if a compound delays or even reverses diabetes-related nephropathy.

Potential limitations of the current study include the number of healthy volunteers in the study and preanalytical variability of the samples, but this study was designed based on the fit-for-purpose stepwise qualification process, and therefore the study was meant to give initial findings for each of the biomarkers in the healthy volunteer population and not give final conclusive data for each candidate biomarker. This study was designed with only 39 subjects (20 male and 19 female), which is too low to yield true reference ranges and even too low to have split the analysis to determine which candidate biomarkers are affected by age. However, three samples for each of the 39 subjects enabled determination of intrasubject variability, which will help in the interpretation of results in the next steps of biomarker qualification. The CIs determined by this study are deemed only as pilot, due to the low number of subjects, but give a basis for understanding what would be deemed “normal” in the next qualification steps. The final conclusive results for each candidate biomarker, including setting more definitive reference intervals, will be generated at a later point in the stepwise qualification process, after/if the biomarkers pass

the qualification steps with patient disease populations and subjects treated with known nephrotoxicants.

Preanalytical variability is another potential limitation of this study. These candidate biomarkers are in the initial steps of qualification, so there are many unknowns concerning these biomarkers, such as stability, appropriate storage conditions, and/or if a stabilizer needs to be added to samples before storage. Each of these preanalytical variabilities need to be evaluated during the qualification process and can impact the data, but this study was designed to give an initial readout for each of these potential biomarkers and to give an understanding of what to expect from samples collected with stabilizer and stored using standard clinical study procedures. This is important because previous and current clinical studies do not prospectively plan to evaluate these kidney biomarkers, but will add these if there are clinical indications of potential kidney injury. It is still unknown how any candidate biomarker that passes the qualification process will be utilized. For example, there will be some instances where the qualified biomarkers will be prospectively put into the study plan and the samples would be stored under optimal conditions if different from standard storage conditions in clinical studies. However, there will also be times when a study unexpectedly has results suggesting potential kidney injury. This may lead the project team to retrospectively evaluate samples from the current and previous clinical studies. Retrospective samples will not have included the biomarkers in the study plan, and the samples would have been stored using standard procedures, so the project team will need to understand how to interpret the data. Therefore, future qualification studies will need to determine if samples stored using standard procedures: yield numerical values similar to samples stored under optimal conditions; yield different numerical values that would be interpretable within a given study; or cannot be used due to giving erroneous and uninterpretable values. Separate reference intervals may be necessary if the second option is correct, but if the third option is correct, then this would result in the loss of a lot of beneficial samples and hinder drug development due to the inability to investigate a compound with no preclinical DIKI signal and a slight potential signal in a late-stage clinical trial.

In conclusion, this study characterized renal biomarker candidates in the healthy volunteer step of the biomarker qualification process. Renal biomarker concentration and concentration normalized to creatinine CIs (pilot reference intervals) as well as intersubject and intrasubject variability were determined using urine from healthy volunteers.

No major effect of food intake was observed for any of the biomarkers. Many of the biomarkers under characterization in this study were affected by sex. Single analyte and multiplex vendor assay comparisons determined that both assay platforms were acceptable and should be considered for further evaluation in future biomarker qualification steps. In summary, these renal biomarkers/assays should be considered for future biomarker qualification steps. However, lack of correlation between the B2M assays should be better understood before completing qualification.

Acknowledgment

We thank Drs Raj Chetty, Tim Carlson, and Amar Sethi for useful discussions during the work and/or critical comments on the manuscript.

Disclosure

The authors report no conflicts of interest in this work.

References

- Star RA. Treatment of acute renal failure. *Kidney Int.* 1998;54(6):1817–1831.
- Uchino S. Creatinine. *Curr Opin Crit Care.* 2010;16(6):562–567.
- Dieterle F, Sistare F, Goodsaid F, et al. Renal biomarker qualification submission: a dialog between the FDA-EMEA and predictive safety testing consortium. *Nat Biotechnol.* 2010;28(5):455–462.
- Harpur E, Ennulat D, Hoffman D, et al. Biological qualification of biomarkers of chemical-induced renal toxicity in two strains of male rat. *Toxicol Sci.* 2011;122(2):235–252.
- Ozer JS, Dieterle F, Troth S, et al. A panel of urinary biomarkers to monitor reversibility of renal injury and a serum marker with improved potential to assess renal function. *Nat Biotechnol.* 2010;28(5):486–496.
- Guinee DG Jr, van Zee B, Houghton DC. Clinically silent progressive renal tubulointerstitial disease during cisplatin chemotherapy. *Cancer.* 1993;71(12):4050–4054.
- Shemesh O, Golbetz H, Kriss JP, Myers BD. Limitations of creatinine as a filtration marker in glomerulopathic patients. *Kidney Int.* 1985;28(5):830–838.
- Matheis K, Laurie D, Andriamandroso C, et al. A generic operational strategy to qualify translational safety biomarkers. *Drug Discov Today.* 2011;16(13–14):600–608.
- Bellomo R, Kellum JA, Ronco C. Acute kidney injury. *Lancet.* 2012;380(9843):756–766.
- Bentley ML, Corwin HL, Dasta J. Drug-induced acute kidney injury in the critically ill adult: recognition and prevention strategies. *Crit Care Med.* 2010;38 Suppl 6:S169–S174.
- Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annett JL. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA.* 2006;296(15):1858–1866.
- Buelow MW, Dall A, Regner K, et al. Urinary interleukin-18 and urinary neutrophil gelatinase-associated lipocalin predict acute kidney injury following pulmonary valve replacement prior to serum creatinine. *Congenit Heart Dis.* 2012;7(5):441–447.
- De Geus HR, Betjes MG, Bakker J. Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges. *Clin Kidney J.* 2012;5(2):102–108.
- Devarajan P, Krawczeski CD, Nguyen MT, Kathman T, Wang Z, Parikh CR. Proteomic identification of early biomarkers of acute kidney injury after cardiac surgery in children. *Am J Kidney Dis.* 2010;56(4):632–642.
- Vanmassenhove J, Vanholder R, Nagler E, Van Biesen W. Urinary and serum biomarkers for the diagnosis of acute kidney injury: an in-depth review of the literature. *Nephrol Dial Transplant.* 2013;28(2):254–273.
- Zhang X, Gibson B Jr, Mori R, et al. Analytical and biological validation of a multiplex immunoassay for acute kidney injury biomarkers. *Clin Chim Acta.* 2013;415:88–93.
- Huber PJ. *Robust Statistics.* New York, NY: John Wiley and Sons; 1981.
- Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem.* 1998;44(3):622–631.
- Salimetrics. Inter and intraassay coefficients of variability. Available from: <http://www.salimetrics.com/documents/spit-tips/publications/Inter%20and%20Intra%20Assay%20Coefficients%20of%20Variability.pdf>. Accessed September 1, 2013.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Statist Soc B.* 1995;57(1):289–300.
- Wind RD, Tolboom H, Klare I, Huys G, Knol J. Tolerance and safety of the potentially probiotic strain lactobacillus rhamnosus PRSF-L477: a randomized, double-blind placebo-controlled trial in healthy volunteers. *Br J Nutr.* 2010;104(12):1806–1816.
- Kratz A, Ferraro M, Sluss PM, Lewandrowski KB. Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Laboratory reference values. *N Engl J Med.* 2004;351(15):1548–1563.
- Delanaye P, Rozet E, Krzesinski JM, Cavalier E. Urinary NGAL measurement: biological variation and ratio to creatinine. *Clin Chem Acta.* 2011;412(3–4):390.
- Goldstein SL. Urinary kidney injury biomarkers and urine creatinine normalization: a false premise or not? *Kidney Int.* 2010;78(5):433–435.
- Heavner DL, Morgan WT, Sears SB, Richardson JD, Byrd GD, Ogden MW. Effect of creatinine and specific gravity normalization techniques on xenobiotic biomarkers in smokers' spot and 24-h urines. *J Pharm Biomed Anal.* 2006;40(4):928–942.
- Pennemans V, Rigo JM, Faes C, Reynders C, Penders J, Swennen Q. Establishment of reference values for novel urinary biomarkers for renal damage in the healthy population: are age and gender an issue? *Clin Chem Lab Med.* 2013;51(9):1795–1802.
- Waikar SS, Sabbisetti VS, Bonventre JV. Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. *Kidney Int.* 2010;78(5):486–494.
- Ralib AM, Pickering JW, Shaw GM, et al. Test characteristics of urinary biomarkers depend on quantitation method in acute kidney injury. *J Am Soc Nephrol.* 2012;23(2):322–333.
- Solberg HE, Petitclerc C. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert Panel on Theory of Reference Values. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *J Clin Chem Clin Biochem.* 1988;26(9):593–598.
- Ichihara K, Boyd JC. An appraisal of statistical procedures used in derivation of reference intervals. *Clin Chem Lab Med.* 2010;48(11):1537–1551.
- Grenier FC, Ali S, Syed H, et al. Evaluation of the ARCHITECT urine NGAL assay: assay performance, specimen handling requirements and biological variability. *Clin Biochem.* 2010;43(6):615–620.
- Han WK, Waikar SS, Johnson A, et al. Urinary biomarkers in early diagnosis of acute kidney injury. *Kidney Int.* 2008;73(7):863–869.

33. Vaidya VS, Waikar SS, Ferguson MA, et al. Urinary biomarkers for sensitive and specific detection of acute kidney injury in humans. *Clin Transl Sci*. 2008;1(3):200–208.
34. Takashi M, Zhu Y, Miyake K, Kato K. Urinary 28-kD calbindin-D as a new marker for damage to distal renal tubules caused by cisplatin-based chemotherapy. *Urol Int*. 1996;56(3):174–179.
35. Zheng J, Xiao Y, Yao Y, et al. Comparison of urinary biomarkers for early detection of acute kidney injury after cardiopulmonary bypass surgery in infants and young children. *Pediatr Cardiol*. 2013;34(4):880–886.
36. Sasaki D, Yamada A, Umeno H, et al. Comparison of the course of biomarker changes and kidney injury in a rat model of drug-induced acute kidney injury. *Biomarkers*. 2011;16(7):553–566.
37. Häring N, Mähr HS, Mündle M, Strohal R, Lhotta K. Early detection of renal damage caused by fumaric acid ester therapy by determination of urinary β 2-microglobulin. *Br J Dermatol*. 2011;164(3):648–651.
38. Guha M, Heier A, Price S, et al. Assessment of biomarkers of drug-induced kidney injury in cynomolgus monkeys treated with a triple reuptake inhibitor. *Toxicol Sci*. 2011;120(2):269–283.
39. Betton GR, Ennulat D, Hoffman D, Gautier JC, Harpur E, Pettit S. Biomarkers of collecting duct injury in Han-Wistar and Sprague-Dawley rats treated with N-phenylanthranilic acid. *Toxicol Pathol*. 2012;40(4):682–694.
40. Dieterle F, Perentes E, Cordier A, et al. Urinary clusterin, cystatin C, beta2-microglobulin and total protein as markers to detect drug-induced kidney injury. *Nat Biotechnol*. 2010;28(5):463–469.
41. Riser BL, Cortes P, DeNichilo M, et al. Urinary CCN2 (CTGF) as a possible predictor of diabetic nephropathy: preliminary report. *Kidney Int*. 2003;64(2):451–458.
42. Wunnapak K, Liu X, Peake P, et al. Renal biomarkers predict nephrotoxicity after paraquat. *Toxicol Lett*. 2013;222(3):280–288.
43. Aydođdu M, Gürsel G, Sancak B, et al. The use of plasma and urine neutrophil gelatinase associated lipocalin (NGAL) and cystatin C in early diagnosis of septic acute kidney injury in critically ill patients. *Dis Markers*. 2013;34(4):237–246.
44. Pinches M, Betts C, Bickerton S, et al. Evaluation of novel renal biomarkers with a cisplatin model of kidney injury: gender and dosage differences. *Toxicol Pathol*. 2012;40(3):522–533.
45. Pai MP, Norenberg JP, Telepak RA, Sidney DS, Yang S. Assessment of effective renal plasma flow, enzymuria, and cytokine release in healthy volunteers receiving a single dose of amphotericin B desoxycholate. *Antimicrob Agents Chemother*. 2005;49(9):3784–3788.
46. Askenazi DJ, Montesanti A, Hunley H, et al. Urine biomarkers predict acute kidney injury and mortality in very low birth weight infants. *J Pediatr*. 2011;159(6):907–912.
47. Askenazi DJ, Koralkar R, Hundley HE, et al. Urine biomarkers predict acute kidney injury in newborns. *J Pediatr*. 2012;161(2):270–275.
48. Hörstrup JH, Gehrman M, Schneider B, et al. Elevation of serum and urine levels of TIMP-1 and tenascin in patients with renal disease. *Nephrol Dial Transplant*. 2002;17(6):1005–1013.
49. Yu Y, Jin H, Holder D, et al. Urinary biomarkers trefoil factor 3 and albumin enable early detection of kidney tubular injury. *Nat Biotechnol*. 2010;38(5):470–477.
50. Astor BC, Köttgen A, Hwang SJ, Bhavsar N, Fox CS, Coresh J. Trefoil factor 3 predicts incident chronic kidney disease: a case-control study nested within the atherosclerosis risk in communities (ARIC) study. *Am J Nephrol*. 2011;34(4):291–297.

Drug Design, Development and Therapy

Publish your work in this journal

Drug Design, Development and Therapy is an international, peer-reviewed open-access journal that spans the spectrum of drug design and development through to clinical applications. Clinical outcomes, patient safety, and programs for the development and effective, safe, and sustained use of medicines are a feature of the journal, which

Submit your manuscript here: <http://www.dovepress.com/drug-design-development-and-therapy-journal>

Dovepress

has also been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.