QUALIFICATION DECISION AND EXECUTIVE SUMMARY

The FDA is issuing this Qualification Decision and Executive Summary to the Critical Path Institute's Predictive Safety Testing Consortium Nephrotoxicity Working Group (CPATH PSTC-NWG), and Foundation for the National Institutes of Health's Biomarker Consortium Kidney Safety Biomarker Project Team (FNIH BC-KSP) [herein referenced as "Submitter"], in response to your biomarker Full Qualification Package submitted to the Center for Drug Evaluation and Research (CDER) Biomarker Qualification Program (BQP). We have completed our review of your Full Qualification Package[1] submission, and have concluded to **Qualify** this biomarker panel for the Context of Use (COU) as described below. In addition, this document includes a summary of the discipline-specific reviews and recommendations by the members of the Biomarker Qualification Review Team (BQRT).

This biomarker qualification represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. This biomarker can be used by drug developers for the qualified COU in submissions of investigational new drug applications (INDs), new drug applications (NDAs), and biologics license applications (BLAs) without the need to resubmit the biomarker information or rereview by the relevant CDER disciplines.

Biomarker Panel

The safety biomarker panel is interpreted via a Composite Measure (CM) of the following six urinary biomarkers: clusterin (CLU), Cystatin-C (CysC), Kidney Injury Molecule-1 (KIM-1), N-acetyl-beta-D-glucosaminidase (NAG), Neutrophil Gelatinase-Associated Lipocalin (NGAL), and osteopontin (OPN) ("the biomarker panel"). The CM is a geometric mean (GM) of the fold changes from baseline of the urine creatinine (uCr)-normalized six urine biomarkers.

[1] Since the Full Qualification Package was submitted and under review prior to the enactment of the 21st Century Cures legislative 507 process, this qualification determination was made following the legacy process.

Acronym	Name (Unique ID (Uniprot))	Description
CLU	Urinary Clusterin (P10909)	A heterodimeric highly conserved secreted glycoprotein expressed in the proximal and distal tubules, glomerulus and collecting duct.
CysC	Cystatin-C (P01034)	A small serum protein produced by all nucleated cells and found in most tissues and body fluids. CysC is freely filtered by the glomerulus and completely reabsorbed and catabolized in healthy renal tubular epithelium.
KIM-1	Kidney Injury Molecule -1 (Q96D42)	A type I transmembrane glycoprotein containing an ectodomain consisting of an immunoglobulin-like domain and a mucin domain that is strongly induced by ischemic and toxic insults to the kidney.
NAG	N-acetyl-beta-D-glucosaminidase (O60502)	A large lysosomal enzyme with two isoforms and is mainly expressed in proximal tubules.
NGAL	Neutrophil gelatinase-associated lipocalin (P80188)	Expressed in various tissues at low levels with upregulated transcription in tubuloepithelial cells following ischemic and nephrotoxic kidney injuries.
OPN	Osteopontin (P10451)	A highly acidic glycoprotein expressed by many tissues that acts as a macrophage adhesion and chemotactic molecule.

Context of Use (COU)

"A safety composite biomarker panel to be used in conjunction with traditional measures to aid in the detection of kidney tubular injury in phase 1 trials in healthy volunteers when there is an a priori concern that a drug may cause renal tubular injury in humans."

The following section lists considerations when using the biomarker for this COU:

General Considerations

- The Composite Measure (CM) is not intended to replace standard measures (such as serum creatinine, BUN, urinalysis, urine albumin, and urine total protein) used to monitor for drug-induced renal injury in clinical trials, but is intended to be used in addition to these standard measures. There should be a plasma drug exposure margin relative to the anticipated clinical exposure range, such that the likelihood of kidney injury is considered low at the doses proposed for clinical investigation. Some items to consider when evaluating the appropriateness of relying on the CM to help guide safety monitoring in clinical trials: human exposure margins, prior

knowledge of drug class nephrotoxicity or absence thereof, and whether animal species other than the rat were shown to be have exposure-related nephrotoxicity.

- The CM values provided in Table 2 were derived from a normal healthy volunteer (NHV) sample. After calculating the CM for a new cohort of subjects, if the geometric mean (or ratio of the geometric means for two samples) is greater than the CM value from the normal healthy volunteer sample, then one can conclude that data from this new cohort of subjects is inconsistent with the NHV sample. One reason for this inconsistency may be potential nephrotoxicity as that is the purported intent of the biomarkers that comprise the composite measure.

- The CM may be used for adjunctive safety monitoring in clinical trials when nonclinical toxicology studies with a study drug demonstrate evidence of reversible histologic renal tubule damage that is associated with an elevation in any one of the six urine biomarkers.

- Biomarker levels should be measured at baseline, prior to exposure to the study drug, and post-baseline in study subjects. Sample collection times may be informed by animal toxicology data for the study drug.

- The CM is likely to be more reliable in populations that closely mirror the population from which the CM was derived.

- The CM is calculated for a cohort of subjects. Estimates based on a larger number of subjects are expected to be more reliable; however marked deviations from normal in a single biomarker or subject should also prompt further evaluation (see next bullet).

- Because elevations in the CM may reflect a non-renal etiology, elevations of the CM should prompt further evaluation and investigation for renal as well as non-renal etiologies.

- The CM is not qualified for individual patient safety monitoring.

Background

The current standard safety biomarkers used to monitor kidney function and toxicity include sCr, BUN, and urinary protein. These safety biomarkers have drawbacks and weaknesses in clinical trials. Currently, there are no biomarkers qualified by the FDA Biomarker Qualification Program to identify the presence of drug-induced renal injury in humans. When evaluating

drugs that have nonclinical signals of kidney injury in early clinical studies, it is critical to ensure the safety of volunteers, particularly if the volunteers are healthy volunteers, since healthy volunteers have no prospect of benefit from participation in the study. To mitigate risk to subjects, researchers often attempt to maintain a sufficient safety margin to the dose/exposure at which renal toxicity was seen in animals; however, this may prevent development programs from evaluating doses/concentrations that are needed to achieve efficacy. Biomarkers that may be more sensitive indicators of renal injury than current standard measures will aid in monitoring acute and sub-acute drug-induced tubular injury in clinical trials so that renal injury may be detected at an early and potentially reversible stage. For description of how this CM is used in a clinical trial setting, please see figure (1) below.

Sources of Data and Major Findings

In support of the context of use, the submitter provided data from two observational studies, one conducted in normal healthy volunteers ("the PSTC Normal Healthy Volunteer Study") and one conducted in patients with mesothelioma undergoing treatment with chemotherapy or surgery for their disease. The submission also references (1) biomarker data obtained in preclinical species showing a correlation between some of the biomarkers in the composite (i.e., CLU, CysC, Kim-1, NGAL and OPN) and histo-morphologic kidney damage; and (2) information gleaned from the published literature on the sensitivity and specificity of each of the biomarkers that make up the CM.

Minor discrepancies were noted in the analysis plan, written results provided and datasets provided. Information requests were made and clarifications received from the submitter.

Analytical Considerations:

The analytical validation for each assay used to evaluate individual biomarkers for this study included the following performance characteristics testing: linearity, recovery (for accuracy), precision, limits of detection/quantitation, sample stability and handling, and interference. Since the samples in the clinical and analytical studies were stored and handled under different conditions in this study, the submitter also included a bridging study to try to demonstrate those differences had no effect

CLU, CysC, KIM-1, NAG, NGAL, and OPN were measured with assays labeled "For Research Use Only" (RUO). Urinary creatinine (uCr) was measured using FDA-cleared

assays per manufacturer instructions. The recommendations of this executive summary do not alter the labeling recommendations of the assays used in this study.

Table 1 Assay Performance characteristics

Biomarker	CLU	CysC	KIM-1	NGAL	NAG	OPN
Assay method	ELISA	ELISA	ELISA	ELISA	Colorimetric	ELISA
Manufacturer	R&D System	R&D System	R&D System	BioPorto Diagnostics	Roche Diagnostics	R&D Systems
Reference interval (normalized to uCr)	35-383 ng/mg	10.4-58.0 ng/mL	<1.19 ng/mg	<41.8 ng/mg	<.78 U/mmol	495-2029 ng/mL
Recovery Range	67.8-105.2%	83.8-104.2%	96.9-118%	93.3-109.4%	99.1-104.5%	97.9-101.5%
Dilutional Range	<11-fold	<64-fold	<32 fold	<64 fold	<40 fold	<32 fold (pre-diluted)
Procedural Dilution	4			100		440
LLOQ	3.13 ng/mL	1.31 ng/mL	11.6 pg/mL	0.4 ng/mL	0.31 U/L	0.1 ng/mL (44 ng/mL after adjusting for 440 fold dilution)
ULOQ	800 ng/mL	100 ng/mL	2000 pg/mL	100 ng/mL	55.25 U/L	8800 ng/mL

Pre-analytical processes: Urine samples were collected in preservative free urine collection cups and were centrifuged at room temperature at 2000 x g for 10 minutes, aliquoted into cryotubes and frozen at -70°C within 3 to 4 hours. Samples were shipped on dry ice to a central storage facility and were maintained at -70°C. Frozen samples were shipped on dry ice to the laboratory for testing. Samples were desalted before testing for clusterin, Cystatin C, and KIM-1.

Urine samples were analyzed locally for creatinine, and total urine volume was measured and recorded. Urine samples were aliquoted and frozen and were analyzed later for biomarkers that included but were not limited to, albumin, total protein, clusterin, cystatin C, and KIM-1 levels. Blood samples were analyzed for serum levels of BUN and creatinine and possibly other potential renal biomarkers. An individual's biomarker values were normalized to urine creatinine levels before analysis.

Precision, Stability and Interference Assessments

Within-run precision was evaluated using low, medium and high in-house control urine samples assayed 16 to 20 times in one analytical run. Between-run precision was evaluated using the same control samples used for within-run precision in three separate analytical runs. Please refer to the analytical review for more information on the precision of the different assays.

Linearity data was provided for each biomarker assay type. The accepted percent recovery was within 80-120% for all concentrations tested.

Long-term and short-term stability studies were provided for each analyte. Please refer to the CDRH Analytical review for stability information for each analyte.

Interference studies showed that high albumin, hematuria and high hemoglobin levels may significantly interfere with some of the assays. Samples with high albumin and visible blood should be excluded from analysis.

Clinical Summary

The PSTC NHV Study was a prospective observational biomarker study conducted in healthy volunteers. The stated primary objectives of the study were to characterize the mean values, normal range, and inter- and intra-subject variability of renal biomarkers (including, but not limited to, urine albumin, total protein, clusterin, cystatin C, beta2-microglobulin, trefoil factor 3, and kidney injury molecule-1 [KIM-1]) in healthy subjects. The secondary objectives were to evaluate correlations among biomarkers, establish assay performance criteria, collect blood for future exploratory studies correlating genomic patterns with biomarker expression and create a well-annotated sample set for evaluation of other biomarkers submitted by the Predictive Safety Testing Consortium in the future. The PSTC NHV Study was conducted at a single site and enrolled 89 subjects who were mostly Caucasian (non-Hispanic or Latino) and overweight. From these 89 subjects enrolled, 76 subjects with biomarker samples collected at baseline (day 1) and post baseline (day 20 ±2 days) were included in the analysis. Subjects in the NHV study did not receive a pharmacological intervention. The study was intended to assess "normal" biomarker levels in a NHV population.

The mesothelioma study (MS) was a phase 1 single-center observational study that was not specifically designed to assess the performance of the CM. Its main objective was to determine the maximum tolerated dose (MTD) of intracavitary heated chemotherapy using a lavage of cisplatin and gemcitabine after extrapleural pneumonectomy (EPP), after pleurectomy/ decortication (P/DC), or after Tumor Debulking +/- Intrapleural Pneumonectomy (TD +/- IPP) with intravenous amifostine and sodium thiosulfate cytoprotection. Thirty-nine of the patients in the study who had no evidence of chronic kidney disease at baseline, and had evaluable specimens were used in the analysis. Longitudinal sample collection occurred prior to surgery or cisplatin treatment, during surgery and after surgery and up to 6 post-op days but the collection timepoints for urine and serum samples were not consistent or precisely timed. The number of sCr measurements for each subject was also highly variable with a median of 19 (range = 9 to 35). Three subgroups were defined including Mesothelioma Surgery (N = 4 surgical

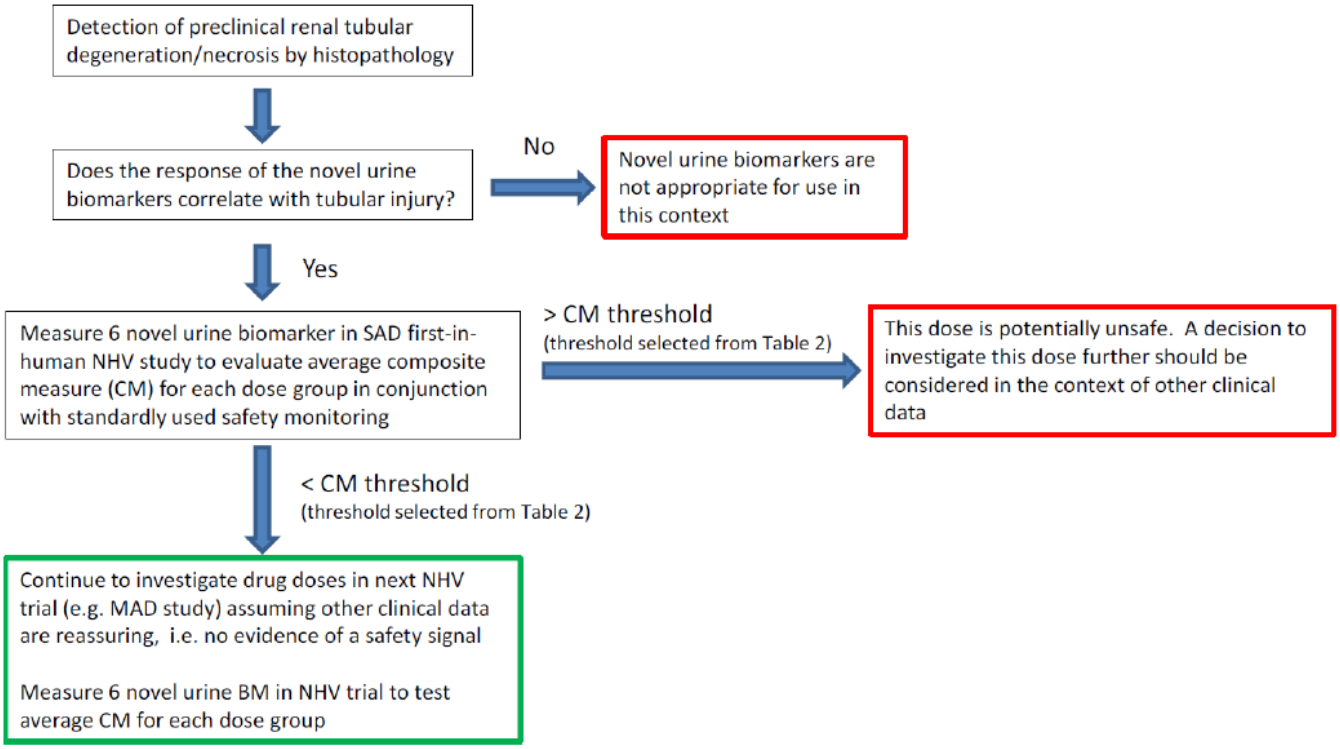
control patients without exposure to cisplatin), Meso Controls (N = 22 patients exposed to cisplatin without clinical manifestation of treatment related renal injury), and Meso Cases (N = 13 patients exposed to cisplatin with clinical manifestation of treatment related renal injury). The CM was calculated for each of the three mesothelioma subgroups and was used to assess whether it could distinguish the three subgroups from one another, or if it could distinguish the mesothelioma cohort from the cohort of normal healthy volunteers. The mesothelioma cohort baseline dataset fell outside of the normal range of the CM identified in the normal healthy volunteer study, even at baseline. The difference in CM between the mesothelioma cohort and the healthy volunteer study supports limiting this qualification to phase 1 studies in healthy volunteers.

There are important limitations to the submitted data:

- The reliability of the derived thresholds and associated probabilities have not been validated using another dataset (Table 2).
- The submission does not address potential intra-subject variability due to factors such as diurnal variation, diet, hydration, or other factors, and contains limited information on intrinsic or extrinsic factors that might affect variability.
- There are a few factors that limits the of the utility of such a measure given potential differences among biomarkers in terms of their sources of and rapidity of observable change in response to injury, and include potential interference, compromise and/or compensatory mechanisms that may influence values in urine. These factors should be considered during the drug development use of this biomarker.

This submission does not propose a set CM threshold to define injury; rather the submission includes tables that provide information on the probability of obtaining a value greater than or equal to a particular value in a cohort of normal healthy volunteers of a particular sample size.

Figure 1: Example decision tree for clinical use of qualified CM in Phase 1 NHV trials



Conclusions:

The CM can be qualified as a safety biomarker for the purpose of identifying a dose cohort that deviates from normal variability in a phase 1 study in normal healthy volunteers.

The rationale for the proposed limited qualification is as follows:

- Nonclinical and clinical data suggest that the component biomarkers have value for detecting acute renal tubule injury.
- Using the CM in phase 1 NHV trials of drugs that are suspected to be nephrotoxicants concurrently with other tests and patient monitoring may inform decision making.
- The risks to study subjects associated with using the CM is minimal because it will be used in conjunction with standard renal safety biomarkers.
- The risk of reaching a false conclusion can be minimized by qualifying the CM as an adjunctive tool with traditional measures for a limited context of use (COU) and specifying appropriate conditions of use.

Single Arm New Cohort of Normal Healthy Volunteers

The steps below define how to apply the CM to a single arm new cohort of healthy subjects, of size m, exposed to a new investigational treatment.

1. For each subject, calculate the uCr-normalized fold-change from baseline for each biomarker. Define this as FC_{ij} for subject i and biomarker j where j = 1, 2,…., 6. The biomarker concentration at a given timepoint is calculated as the concentration of biomarker at that timepoint divided by the concentration of uCr at the same timepoint. The biomarker fold change from baseline at a given timepoint is calculated as the biomarker concentration at a given timepoint divided by the biomarker concentration at baseline.

2. For each subject i, calculate the Composite Measure:

$$CM_i = exp\left\{\sum_{j=1}^{6} \frac{1}{6} log\left(FC_{ij}\right)\right\}$$

3. Calculate the geometric mean of the Composite Measure for the cohort of m subjects:

$$\overline{CM} = exp\left\{\sum_{i=1}^{m} log(CM_i)/m\right\}$$

4. Compare $\overline{CM}$ to the threshold T in tables provided in the Statistical review using the row (m) that corresponds to the sample size of the new cohort and the desired value of P. A sample table is provided as Table 2 below.

5. If $\overline{CM}$ > T, this indicates that the new cohort of normal healthy volunteers is not consistent with the NHV data.

Two Arm New Cohort of Normal Healthy Volunteers

Whereas the above section was based on a single arm cohort of new normal healthy volunteers, the steps below define how to apply the CM using the ratio for a two arm cohort where m subjects are exposed to an investigational product and n subjects are controls, where $6 \leq m \leq 20$ and $m \geq n$.

1. For each subject, calculate the uCr-normalized fold-change from baseline for each biomarker. Define this as FC_{ij} for subject i and biomarker j where j = 1, 2,…., 6.

2. For each subject i, calculate the Composite Measure:

$$CM_i = exp\left\{\sum_{j=1}^{6} \frac{1}{6} log\left(FC_{ij}\right)\right\}$$

3. Calculate the geometric mean of the Composite Measure for cohort k (k = Drug, Control):

$$\overline{CM_k} = exp\left\{\sum_{i=1}^{m} log(CM_i)/m\right\}$$

4. Calculate the ratio of the geometric means for the two cohorts:

$$GM_{ratio} = \overline{CM}_{Drug}/\overline{CM}_{Control}$$

5. Compare GM_{ratio} to the threshold T in the tables provided in the Statistical review corresponding to m drug exposed subjects using the row (n) that corresponds to the sample size of the control subjects and the desired value of P. A sample table is provided as Table 2 below.

6. If GM_{ratio} > T this indicates that the new cohort of normal healthy volunteers is not consistent with the NHV data.

Note that the above steps are based on a single time point. If multiple time points exist, repeat the above steps at each time point.

Table 2 ***Observed cohort CM thresholds using equal weights to calculate the individual CM, and based on varying probabilities of consistency within an NHV population for sample sizes of n = 6 to 20 per group***

Sample Size (m/group)	Thresholds based on NHV population: observed measures greater than or equal to thresholds are expected to be observed with less than P% probability in a NHV population									
	Threshold when assessing a drug study group alone (GM CM)					Threshold when assessing a drug study group relative to a comparator/placebo (ratio of GM CMs)				
	P = 50%	P = 20%	P = 10%	P = 5%	P = 1%	P = 50%	P = 20%	P = 10%	P = 5%	P = 1%
6	1.06	1.17	1.23	1.27	1.36	1.00	1.14	1.23	1.31	1.48
8	1.06	1.15	1.20	1.24	1.32	1.00	1.12	1.20	1.26	1.40
10	1.06	1.14	1.19	1.23	1.29	1.00	1.11	1.18	1.23	1.35
12	1.06	1.13	1.18	1.21	1.27	1.00	1.10	1.16	1.21	1.32
14	1.06	1.13	1.17	1.20	1.25	1.00	1.09	1.15	1.19	1.29
16	1.06	1.13	1.16	1.19	1.24	1.00	1.09	1.14	1.18	1.26
18	1.06	1.12	1.15	1.18	1.23	1.00	1.08	1.13	1.17	1.25
20	1.06	1.12	1.15	1.17	1.22	1.00	1.08	1.12	1.16	1.23

GM = geometric mean
CM = composite measure of the fold change from baseline of urine KIM-1, CLU, OPN, CysC, NAG, and NGAL normalized to uCr

Because the CM is a linear combination of all six biomarkers, the effective use of this method would require each subject in a dose cohort to have a recorded uCr-normalized fold-change from baseline for all six biomarkers. How missing values would be handled was not addressed in the study and it is not clear what the most appropriate imputation approach should be. One approach is to include only those patients in the CM who have results for all baseline and follow-up biomarkers.

Considerations:

Below, are important considerations and sensible practices related to the use of the CM:

- The optimal sampling time to evaluate these biomarkers relative to exposure to a drug nephrotoxin has not been resolved and introduces a potential risk of false negative findings.
- The submitter recommends samples are collected at the same time of day due to potential changes in the analyte throughout the day.
- In general, the timing of biomarker measurements should be informed by the findings in animal studies and, if the concern for toxicity is based on the experience with other members of the pharmacologic class, an understanding of the time course of toxicity for these other members should be considered in determining a schedule for sample collection.
- Including placebo-treated subjects in future studies may aid in the evaluation of the significance of any elevations in biomarkers or the CM.
- Following the biomarker components of the CM in real-time and minimizing missing values of the biomarker components across timepoints may maximize the monitoring utility of biomarker testing. Markedly high biomarker values in an individual subject should also prompt further investigation.
- Accurate interpretation of CM includes evaluation of a urinalysis, and potentially other tests and patient monitoring to ensure changes are nephrotoxic drug effects and not related to other intrinsic or extrinsic factors.
- The analytical performance characteristics of the assay used to develop the CM normal range was defined by a particular assay and CM may perform differently when used with assays that have different performance characteristics.