EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

30 April 2010
EMA/283298/2010
Committee for Medicinal Products for Human Use

Consultation on the Qualification Opinion ILSI/HESI Submission of Novel Renal Biomarkers for Toxicity

Agreed by Scientific Advice Working Party	February 2010
Adoption by CHMP for release for consultation	18 March 2010
End of consultation (deadline for comments)	31 July 2010

Comments should be provided using this template. The completed comments form should be sent to SAWPsecretariat@ema.europa.eu

Keywords	*Non-clinical, renal biomarkers, nephrotoxicity*[1]

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8416
E-mail info@ema.europa.eu **Website** www.ema.europa.eu

QUALIFICATION OPINION

ILSI/HESI Submission of Novel Renal Biomarkers for Toxicity

On 07 April 2008 the Applicant ILSI Health and Environmental Sciences Institute (HESI) requested qualification opinion for the biomarkers α-GST, RPA-1 and clusterin.

HESI proposes that there is sufficient evidence to support the voluntary use of α-GST, RPA-1 and clusterin along with currently used methods to gain further insight into renal injury when it is seen in preclinical safety assessment studies in rats.

Prof. Beatriz Silva Lima was appointed as coordinator. The EMA Scientific Administrator for the procedure was Mr Efthymios Manolis. The procedure started during the SAWP meeting held on 06 - 07 May 2008.

On 23 June 2008, the Qualification Team participated in a teleconference with FDA to discuss a common list of issues. On 26 June 2008, the Applicant was provided with the EMA/FDA common list of issues to be addressed in writing and during the discussion meeting.

The discussion meeting with the Applicant and FDA took place on 02 July 2008.

On 08 December 2008, the Applicant submitted the written responses to the list of issues including additional documentation.

On 05 February 2009 EMA and the FDA sent additional combined questions to be addressed by the Applicant in writing.

On 30 April 2009, the Applicant submitted the written responses to the EMA/FDA combined additional questions including additional documentation.

During its meeting held on 22 - 24 February 2010, the SAWP agreed on the opinion to be given to the Applicant.

During its meeting held on 15 - 18 March 2010, the CHMP adopted the opinion to be given to the Applicant. This opinion is annexed to this letter.

The opinion given by CHMP is based on the claims and supporting documentation submitted by the Applicant, considered in the light of the current state-of-the-art in the relevant scientific fields.

London, 18 March 2010

On behalf of the CHMP

Dr Eric Abadie, Chairperson

Background

The European Medicines Agency qualification process is a new, voluntary, scientific pathway leading to either a CHMP opinion or a Scientific Advice on innovative methods or drug development tools. It includes qualification of biomarkers developed by consortia, networks, public/private partnerships, learned societies or pharmaceutical industry for a specific intended use in pharmaceuticals R&D.

The HESI study evaluated four novel biomarkers (BMs) of nephrotoxicity (α-GST, μ-GST, RPA-1 and clusterin) and compared their performance against the more traditional measurements for diagnosis of nephrotoxicity.

The data presented in this report were all generated in single and repeated dose studies conducted in male rats of two strains (Sprague-Dawley and Wistar) that are commonly used in preclinical toxicity studies. The information obtained from these studies demonstrates the potential utility of these BMs for use in rodent studies conducted to evaluate the potential target organ toxicity of compounds as part of the preclinical safety assessment of candidate medicines.

Scope

HESI proposes that there is sufficient evidence to support the voluntary use of α-GST, RPA-1 and clusterin along with currently used methods to gain further insight into renal injury when it is seen in preclinical safety assessment studies in rats.

Specifically, HESI claims the following:

"The novel BMs evaluated have been shown not only to have utility for the detection of tubular injury but some also provide useful information on the tubular site of injury

- Clusterin is confirmed to have utility for detection of tubular injury without additional insight as to location. The superiority of clusterin (compared with the reference BMs: BUN and sCr) was evident when regeneration was present.
- α-GST was shown to be superior to all of the reference markers for detection of injury to the proximal tubule.
- RPA-1 was shown to be superior to all of the reference markers for detection of injury to the collecting duct."

These claims are the subject of the CHMP qualification process described below. Measurements were made for μ-GST (Yb1); however according to the Applicant, these data did not support a claim.

HESI Overall Strategy General Principles

- Select compounds causing damage to specific portions of the nephron. HESI stated that correlation of biomarker values with a defined histopathological phenotype is central to evaluation of the utility of biomarkers to report nephrotoxic injury. The use of different compounds and different dosage regimens (both unitary doses and duration of dosing) serves to produce varied pathology (different sites and patterns of injury). Histopathology was the gold standard used to assess biomarker performance. Immunohistochemistry was used to confirm location of specific lesions.
- Use male (HW and SD) rats as the initial test species.
- Select candidate biomarkers with promise based on literature and previous HESI programs.

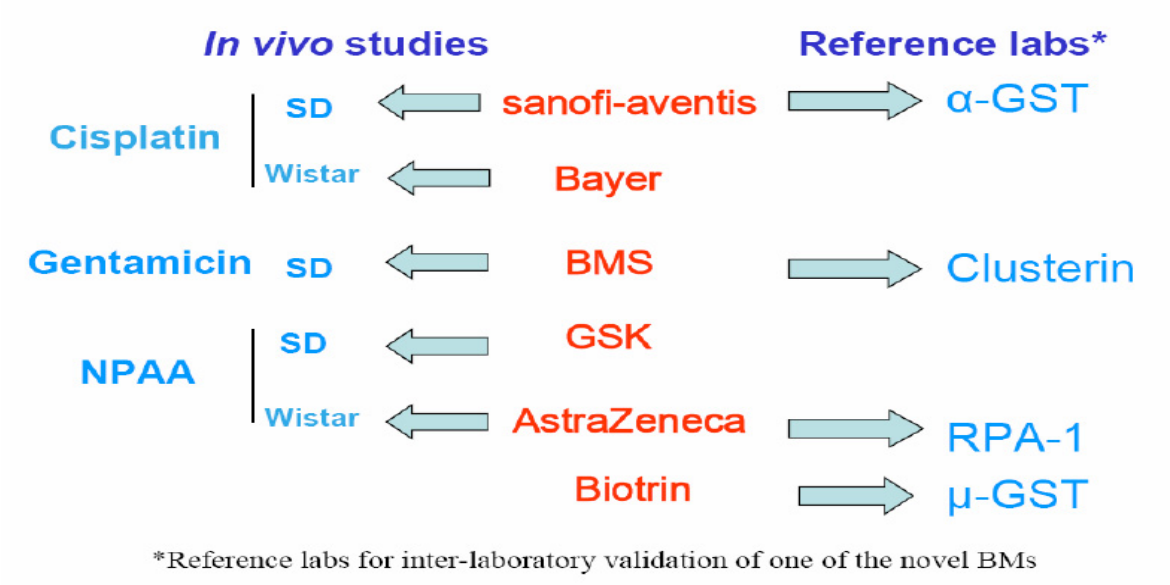
- Conduct pilot studies to define doses ranging from no observable effect dose to markedly toxic (i.e. using dose-response information to obviate, in the first instance, the need for "negative" controls other than vehicle).
- Define procedures to be used for sample collection by each participating laboratory to ensure optimal preservation of biomarkers and other analytes.
- Develop biomarker assays where none exist and comprehensively validate the biomarker assays.
- Design and execute single definitive studies for each selected nephrotoxin and rat strain and exchange samples between six labs as part of a robust assay validation process.
- Conduct according to GLP standards with the following exceptions:

  - There would be no independent audit although each participating laboratory would be responsible for conducting thorough quality control of their data.

  - No formulation analysis or exposure assessment would be done since development of validated assays would require disproportionate resource and time.
- Where sample volumes permit, conduct auxiliary studies using 'omics' technologies to identify additional biomarkers.

HESI Methodology

Distribution of work



*Reference labs for inter-laboratory validation of one of the novel BMs

Animals

Sprague Dawley or Han Wistar male rats

Test compounds and anticipated principal site of lesion

- Cisplatin (single dose) – proximal tubule
- Gentamicin (14 daily doses) – proximal tubule
- N-phenylanthranilic acid (14 daily doses) – papilla

Minimum biochemistry parameters for comparison

Urine: volume, quantitative measurement of protein, enzymes (NAG and GGT), creatinine, glucose, microscopy.

Plasma (or serum): creatinine, urea nitrogen, total protein.

Necropsy & histopathology

On days specified in individual protocols, approximately 24 hours post-dose and following completion of the urine collection, animals were killed, using the approved procedure in the facility conducting the study, and necropsied.

Both kidneys were removed and weighed, fixed in neutral buffered formalin followed by trimming, embedding, sectioning and staining to local practices. The standard stain used was haematoxylin and eosin; additional stains used are documented in the individual study reports. In some studies, standard liver portions were fixed in neutral buffered formalin and processed at the discretion of the study pathologist. In no study was any injury to the liver found so this is not discussed further.

Definition of pathology nomenclature (lexicon) and harmonisation of diagnostic terms

Slides and histopathology data from all five HESI studies were initially reviewed by a Pathology Working Group (PWG) to assess morphologic diagnoses and consistency in grading of finding severity, to identify key treatment-related findings for each toxicant, and to derive a common lexicon of morphologic diagnoses. Key agreements from this peer review were subsequently combined with the nascent kidney histopathology lexicon of the C-Path Nephrotoxicity Working Group, and a unified renal histopathology lexicon was drafted following consensus between the HESI and C-Path NWG in October of 2006.

Processing of histopathology data for the ROC analysis

Subsequent to the PWG and prior to performing the receiver operating characteristic analysis, morphologic diagnoses from the five HESI studies were converted to conform to the common lexicon resulting in identification of fourteen morphologic diagnoses. These diagnoses were further assessed by members of the PWG for relationship to treatment and prioritized for ROC analysis based on the significance of each finding relevant to the expected or observed toxicity. Because one of the main objectives of the performance evaluation was to assess the relationship between changes in urinary markers and injury to specific segments of the rat nephron, histopathology data were further processed to remove redundancies and ensure that each animal had only one histopathology diagnosis per pathologic process (i.e. degeneration/necrosis or regeneration/basophilia) per segment. A diagram summarizing the processing of the histopathology data is provided below:

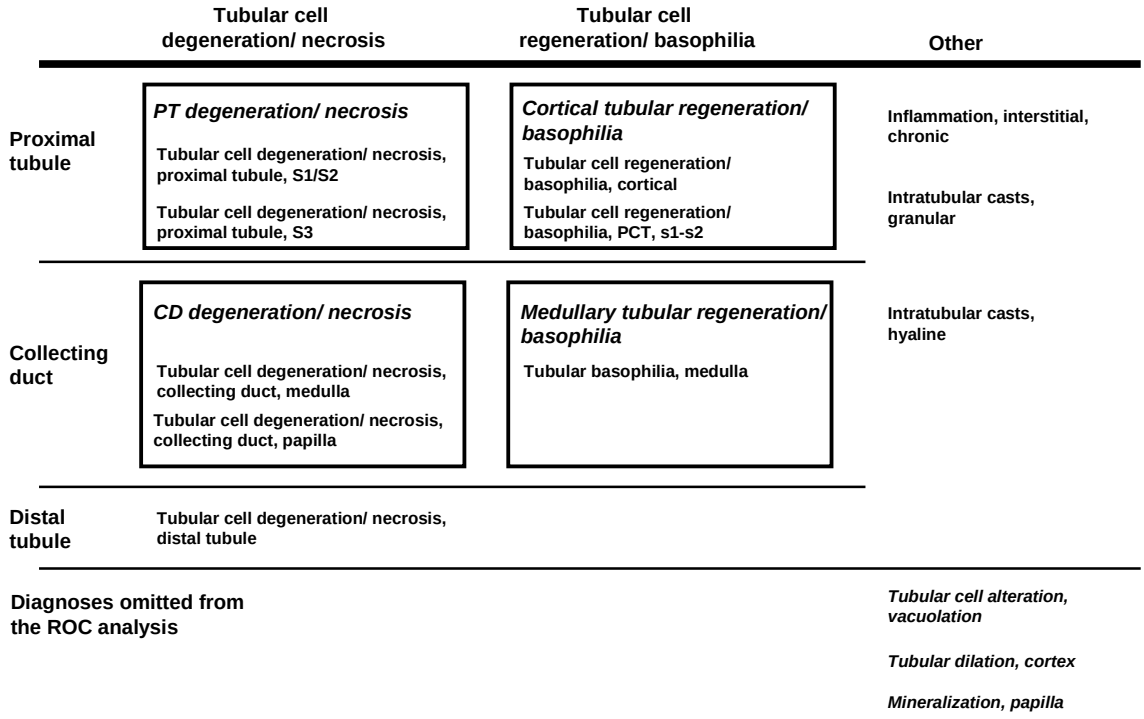


Figure - Histopathology data summary

Immunohistochemistry

Immunohistochemistry was performed to confirm nephron segment of injury on representative sections for each nephrotoxicant.

Dose ranging finding studies

Dose range finding (DRF) studies were conducted. Initial selection of dosages and study design was based on information from the literature and prior experience in the participating laboratories. Nevertheless, the DRF studies were judged to be necessary to confirm the dosage selection and time-points for assessment of nephrotoxicity in the definitive studies. Additional compounds, either having no known evidence of nephrotoxic potential or affecting parts of the nephron other than those of primary interest, were included in the DRF studies for evaluation as potential negative controls. However, in order to keep the size of the definitive studies manageable, it was subsequently decided to use doses ranging from sub-threshold (for renal injury) to clearly nephrotoxic, to eliminate the need for inclusion of negative controls (other than vehicle).

Definitive studies

HESI detailed summary of studies

Sponsor's Table - Dose groups, compound administration and numbers of animals in gentamicin study (Sprague-Dawley rats)

Group	Dose of gentamicin (mg/kg/day)[a]	Dose volume (mL/kg)	Total No. of Animals	No. of animals per necropsy (Day of necropsy)
1	0 (vehicle [b])	1.0	20	10 (Days 8 or15)
2	5	1.0	10	10 (Day 15)
3	25	1.0	10	10 (Day 15)
4	50	1.0	10	10 (Day 15)
5	100	1.0	20	10 (Day 8 or 15[c])

[a] Doses are expressed in terms of gentamicin base and dose volumes were based on the most recent body weight.
[b] 0.9% saline
[c] No high dose animals survived to study day 15

O/N Urine collected Day 3/4, Day 7/8 and 14/15 in all doses groups

Dose groups, compound administration and numbers of animals in cisplatin studies (Han-Wistar and Sprague-Dawley rats)

Group Number	Dose of Cisplatin (mg/kg)[a]	Dose volume (mL/kg)	Total No. of Animals	No. of animals per necropsy (Day of necropsy)
1	0(vehicle[b])	5	30	10 (Days 2, 3 or 5)
2	0.3	5	30	10 (Days 2, 3 or 5)
3	1	5	30	10 (Days 2, 3 or 5)
4	3	5	30	10 (Days 2, 3 or 5)

[a] All doses are expressed in terms of the pure parent compound, and dose volume was based on the most recent body weight.
[b] 0.9% saline

O/N urine was collected for biomarker analysis on Day 1/2; Day 2/3 and Day 4/5

Dose groups, compound administration and numbers of animals in NPAA studies (Han-Wistar and Sprague-Dawley rats)

Group	Strain of rat	Dose of NPAA(mg/kg/day)[a]	Dose volume (mL/kg)	Total No. of Animals	No. of animals per necropsy (Day of necropsy)
1	Han-Wistar	0 (vehicle[b])	10	30	15 (Days 8 or 15)
	Sprague-Dawley			20	10 (Days 8 or 15)
2	Han-Wistar	50	10	15	15 (Day 15)
	Sprague-Dawley			10	10 (Day 15)
3	Han-Wistar	350	10	15	15 (Day 15)
	Sprague-Dawley	400		30	15 (Day 8 or 15)
4	Han-Wistar	700[c]	10	30	15 (Day 8 or 15)

[a] Doses are expressed in terms of pure parent compound and dose volumes were based on the most recent body weight.
[b] Vehicle was 1.25% carboxymethylcellulose in both strains
[c] Dose reduced from Day 4 to 350 mg/kg for animals sacrificed on Day 8 and to 500 mg/kg for animals sacrificed on Day 14

O/N urine was collected for biomarker analysis on Day 3/4 (both strains, but only Control and high dose in Wistar); Day 7/8 and Day 14/15

Statistical methods

Biomarker assay validation

The following aspects of the performance of the assays for each of the novel biomarkers under study were assessed:

- Repeatability (intra-run precision)
- Intermediate precision (intra-lab precision)
- Reproducibility (inter-lab precision)

Biomarker Qualification

Data Normalization

For all urinary markers, analyte concentrations for all animals were first normalised by dividing by the corresponding urine creatinine concentration. All individual animal marker values (normalised to creatinine in the case of urinary markers) were divided by the mean of the values in the concurrent control (i.e. vehicle-dosed) animals. Thus, all marker values were expressed as a fold-change versus the time-matched control group mean.

Assessment of variability of the novel BMs among control animals

Repeat urinary measurements for both Sprague-Dawley and Han-Wistar control animals were used to estimate intra-animal and inter-animal variability for novel markers α-GST, μ-GST, RPA-1, and clusterin. The statistical analysis was done, as described above, after normalization by division of the analyte concentration by the corresponding urine creatinine concentration.

For each rat strain separately, creatinine-normalized analyte concentrations were analyzed by a one-way random effects analysis of variance (ANOVA). Intra-animal and inter-animal variances were calculated by equating observed and expected mean squares from the one-way random effects ANOVA. Two-sided 95% confidence intervals were calculated using the standard Chi-squared method for intra-animal variance and the modified large sample method for inter-animal variance. Negative estimates and confidence bounds for the inter-animal variance were reported as zero.

Variance estimates and confidence intervals were converted to coefficient of variation (%CV) by applying the square root transformation and dividing by the observed concentration mean.

Histopathology

Histopathology grades were assessed on a scale of 0 (no observed pathology) to 4 (severe pathology). For each path, animals were defined as 'Negative' or 'Positive' as follows:

'Negative': Animals which were dosed with either vehicle or toxicant and with histopathology score = 0.

'Positive': Animals which were dosed with either vehicle or toxicant and with histopathology score > 0.

Data Exclusion

All animals from the high dose group (100 mg/kg) of the gentamicin study were excluded from all statistical analyses. In this group, there were unscheduled sacrifices due to the poor clinical condition of some animals (1 animal on day 6, 11 animals with controls on day 7 and the remaining 4 animals with controls on days 8 and 10). Thus no high dose animals survived to day 14. It was therefore

considered that, despite the selection of the high dose based on a DRF study, this dosage exceeded a maximum tolerated dose in this study.

Additionally, animals for which any individual biomarker result was missing were excluded from all statistical analyses.

Descriptive Statistics

The number of animals classified as 'Negative' and 'Positive' was determined for each pathology, and stratified by rat strain and toxicant.

For each marker and pathology, summary statistics (mean and standard deviation) were calculated for 'Negative' and 'Positive' animals, and stratified by rat strain and toxicant.

Receiver Operating Characteristic Curves

The discriminatory accuracy of each marker was assessed using receiver operating characteristic (ROC) curve methods. The area under the ROC curve (AUCROC), a commonly used index of diagnostic accuracy, was used to compare the performance of each marker.

For each marker and pathology, nonparametric point estimates and standard errors of the AUCROC were calculated by rat strain and across both rat strains.

Using pooled data (across both rat strains), pairwise comparisons of AUCROC values for each novel marker (α-GST, μ-GST, RPA-1, clusterin) versus each reference marker (BUN, SCr, NAG, protein) were performed for each pathology and the corresponding two-sided p-value calculated.

For each novel marker separately, raw p-values were adjusted for multiple testing (i.e. pathologies and reference markers) via the Hochberg procedure. No adjustment was made for the number of novel markers considered.

Incremental Diagnostic Value

The incremental diagnostic value of each novel marker (α-GST, μ-GST, RPA-1, clusterin), when used in conjunction with reference markers, was also assessed using ROC curves.

Two reference marker sets were considered:

1) BUN and serum creatinine

2) NAG and protein

For each pathology and reference marker set, a logistic regression model was fit:

logit (P) = α + β1*X1 + β 2*X2 where P denotes the probability that an animal is 'Positive' and {X1, X2} denote the reference marker values. The AUCROC of the linear score (β 1*X1 + β 2*X2) was calculated. Denote this by AUCROC(X1,X2). For each novel marker separately, the logistic regression models above were then re-fit to include a term for the novel marker (denoted by X3). The AUCROC of the linear score (β 1*X1 + β 2*X2 + β 3*X3) was calculated and denoted by AUCROC(X1, X2, X3).

For each novel marker and reference marker set, pairwise comparisons of AUCROC(X1,X2) versus AUCROC(X1,X2,X3) were performed for each pathology and the corresponding two-sided p-value calculated . Raw p-values were adjusted for multiple testing (i.e. pathologies and reference marker sets) via the Hochberg procedure, for each novel marker separately.

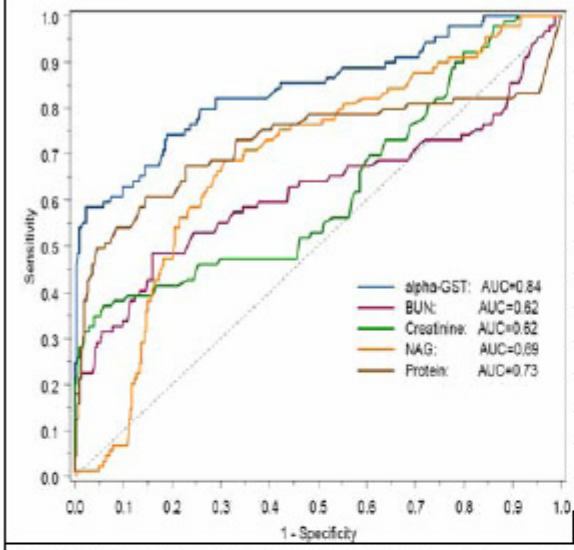
Note that the linear scores derived above via logistic regression do not necessarily yield optimal discriminatory accuracy. Other approaches are possible. However, logistic regression is commonly utilized for deriving marker combinations and provides a practical and reasonable framework for assessing the incremental diagnostic value of the novel markers under consideration.

RESULTS – Diagnostic Performance

The submission includes data from a total of five studies using three compounds (cisplatin, gentamicin and NPAA). Data submitted between April 2008 and April 2009. The HESI data on the analytical validation, inter-animal and intra-animal variability and immunochemistry support the qualification exercise with the caveats identified in the GAPS section below. The tables below (A,B,C,D) compiled by the assessors focus on the overall ROC curves, the number of animals used for this calculation and the AUCROC values vs. histopathology range used for subset. Tables 9, 10 and 11 present the AUCROC values for different pathologies. The full documentation on the statistical analysis, conducted according to the methodology part, is not included in order to keep the document concise.

A. PT degeneration or necrosis - α-GST versus BUN, sCr, NAG, and urine protein

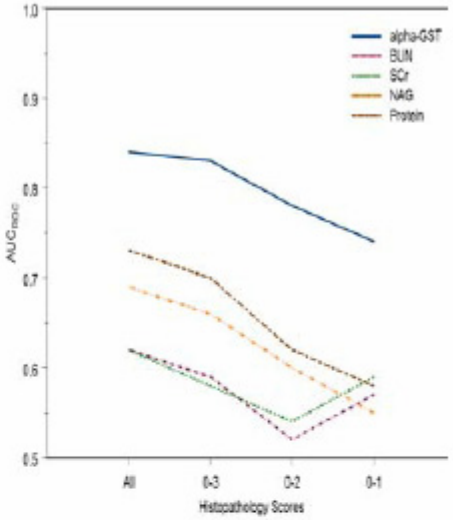
Overall ROC curves - PT degeneration or necrosis



Animals used in calculation

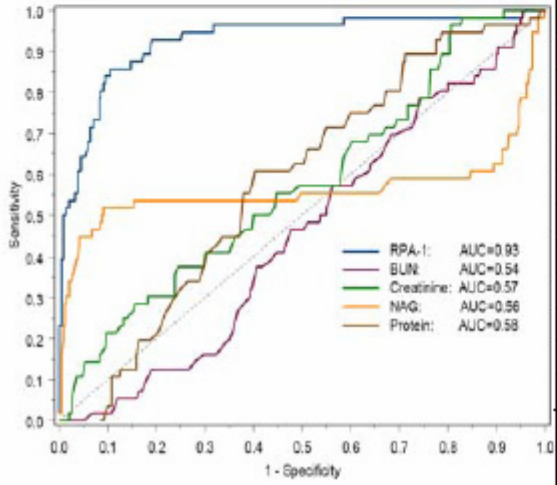
	SD	HW	Tot.
Overall			
Neg	170	137	307
Pos	34	55	89
Total	204	192	396
%	16.7	28.6	22.5
Cisplatin			
Neg	83	64	147
Pos	11	55	66
Total	94	119	213
%	11.7	46.2	30.9
Gentamicin			
Neg	27	ND	27
Pos	23	ND	23
Total	50	ND	50
%	46	ND	46

AUC_{ROC} values vs. histopathology range used for subset.



B. CD degeneration or necrosis - RPA-1 versus BUN, sCr, NAG, and urine protein

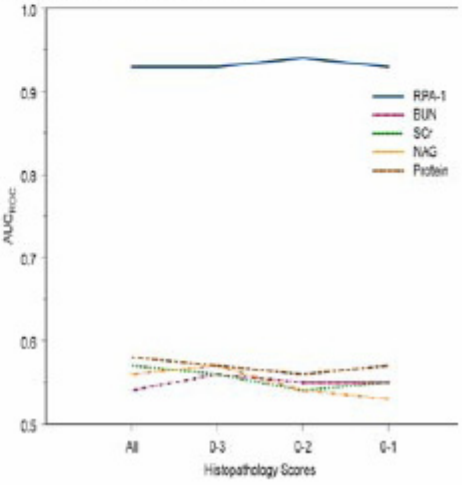
Overall ROC curves - CD degeneration or necrosis.



Animals used in calculation

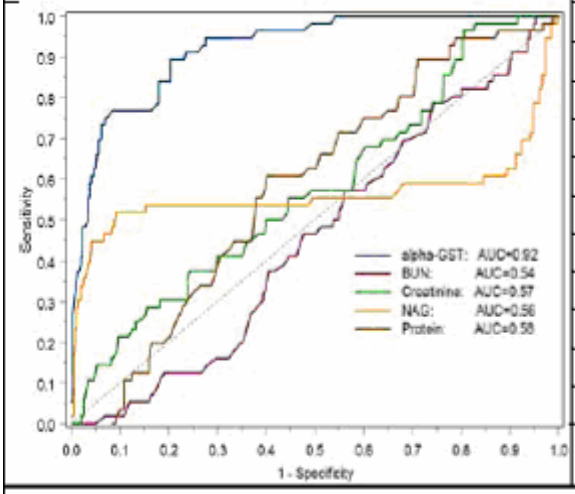
	SD	HW	Tot.
Overall			
Neg	179	161	340
Pos	25	31	56
Total	204	192	396
%	12.3	16.1	14.1
NPAA			
Neg	35	42	77
Pos	25	31	56
Total	60	73	133
%	41.6	42.5	42.1

AUC_{ROC} values vs. histopathology range used for subset.



C. CD degeneration or necrosis - α-GST versus BUN, sCr, NAG, and urine protein

Overall ROC curves - CD degeneration or necrosis.



Animals used in calculation

	SD	HW	Tot.
Overall			
Neg	179	161	340
Pos	25	31	56
Total	204	192	396
%	12.3	16.1	14.1
NPAA			
Neg	35	42	77
Pos	25	31	56
Total	60	73	133
%	41.6	42.5	42.1

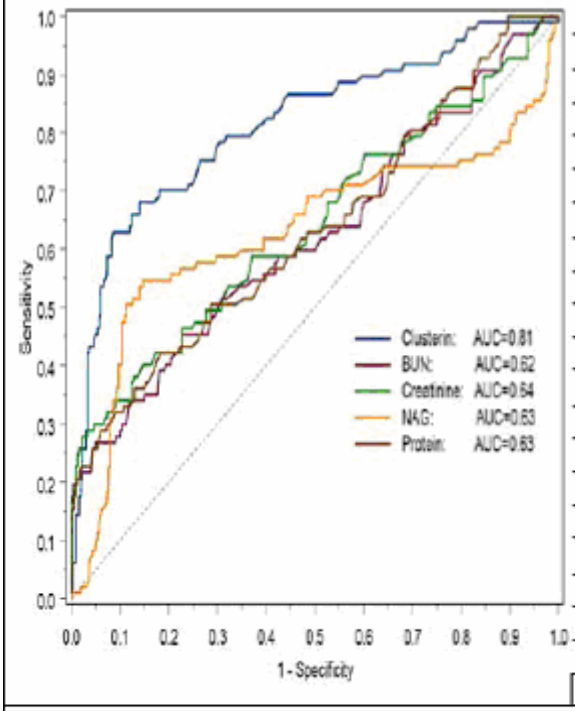
AUC_{ROC} values vs. histopathology range used for subset.

ROC analysis where range of histopathology score is limited was not provided

[Note: In contrast to PT pathology where α-GST values increased, α-GST values decreased in response to increased CD pathology alone.]

D. Cortical tubular regeneration/basophilia - clusterin versus BUN, sCr, NAG, and urine protein

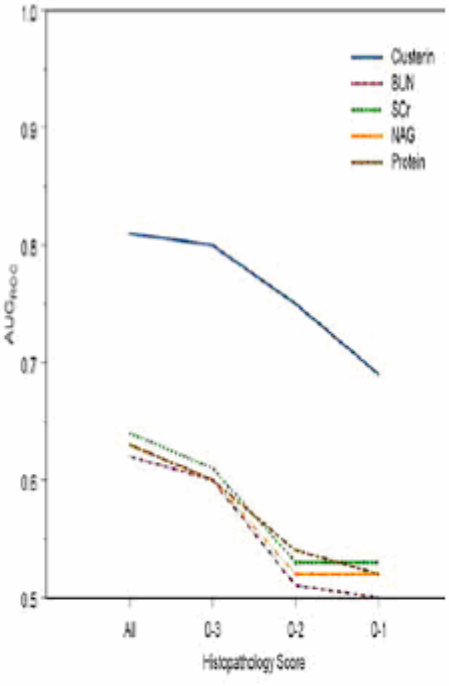
Overall ROC curves – Cortical tubular regeneration /basophilia.



Animals used in calculation

	SD	HW	Tot.
Overall			
Neg	153	146	299
Pos	51	46	97
Total	204	192	396
%	25.0	24.0	24.4
Cisplatin			
Neg	89	99	188
Pos	5	20	25
Total	94	119	213
%	5.3	16.8	11.7
NPAA			
Neg	37	47	84
Pos	23	26	49
Total	60	73	133
%	38	35.6	36.8
Gentamicin			
Neg	27	ND	27
Pos	23	ND	23
Total	50	ND	50
%	46	ND	46

AUC_{ROC} values vs. histopathology range used for subset.



SD = Sprague Dawley; HW = Han Wistar; Neg. = negative; Pos. = positive; ND = Not done

Table 9 - AUC_{ROC} estimates (standard error) for Sprague-Dawley animals

Pathology	BUN	SCr	NAG	Protein	α-GST	μ-GST	RPA	Clust
PT degeneration or necrosis	0.78 (0.05)	0.75 (0.06)	0.93 (0.02)	0.86 (0.04)	0.83 (0.04)	0.91 (0.03)	0.80 (0.04)	0.88 (0.04)
PT deg/nec with no regen	0.67 (0.09)	0.54 (0.12)	0.72 (0.05)	0.70 (0.08)	0.69 (0.11)	0.72 (0.07)	0.60 (0.05)	0.60 (0.08)
PT deg/nec with regen	0.79 (0.06)	0.84 (0.06)	0.97 (0.02)	0.88 (0.05)	0.85 (0.04)	0.94 (0.04)	0.84 (0.04)	0.95 (0.03)
Cortical tubular regeneration/basophilia	0.63 (0.05)	0.74 (0.04)	0.56 (0.06)	0.59 (0.05)	0.53 (0.06)	0.56 (0.06)	0.92 (0.03)	0.84 (0.04)
DT degeneration or necrosis	0.53 (0.07)	0.59 (0.05)	0.87 (0.04)	0.74 (0.05)	0.94 (0.03)	0.85 (0.04)	0.93 (0.02)	0.67 (0.06)
CD degeneration or necrosis	0.57 (0.06)	0.59 (0.05)	0.91 (0.03)	0.68 (0.06)	0.96 (0.02)	0.89 (0.03)	0.89 (0.03)	0.65 (0.05)
CD deg/nec with no regen	0.60 (0.09)	0.52 (0.11)	0.94 (0.02)	0.61 (0.14)	0.91 (0.02)	0.90 (0.02)	0.79 (0.09)	0.65 (0.09)
CD deg/nec with regen	0.63 (0.06)	0.62 (0.05)	0.87 (0.04)	0.69 (0.06)	0.94 (0.03)	0.86 (0.04)	0.90 (0.03)	0.64 (0.06)
Regeneration NOS with no degeneration	0.62 (0.11)	0.53 (0.08)	0.70 (0.11)	0.74 (0.07)	0.65 (0.13)	0.71 (0.11)	0.70 (0.12)	0.63 (0.09)
Intratubular casts, granular, cortex	0.93 (0.03)	0.77 (0.14)	0.98 (0.01)	0.89 (0.06)	0.92 (0.03)	0.96 (0.01)	0.86 (0.03)	0.96 (0.02)
Intratubular casts, hyaline, cortex	0.69 (0.09)	0.76 (0.08)	0.72 (0.09)	0.69 (0.08)	0.55 (0.10)	0.69 (0.09)	0.81 (0.07)	0.86 (0.05)
Inflammation, interstitial, chronic, cortex	0.64 (0.04)	0.62 (0.05)	0.70 (0.04)	0.68 (0.04)	0.65 (0.04)	0.72 (0.04)	0.63 (0.04)	0.65 (0.04)

Table 10 - AUC_{ROC} estimates (standard error) for Wistar animals

Pathology	BUN	SCr	NAG	Protein	α-GST	μ-GST	RPA	Clust
PT degeneration or necrosis	0.53 (0.05)	0.59 (0.05)	0.54 (0.04)	0.64 (0.05)	0.85 (0.03)	0.67 (0.05)	0.54 (0.04)	0.54 (0.05)
PT deg/nec with no regen	0.58 (0.06)	0.53 (0.05)	0.67 (0.04)	0.52 (0.06)	0.73 (0.05)	0.59 (0.05)	0.62 (0.04)	0.66 (0.04)
PT deg/nec with regen	0.76 (0.09)	0.79 (0.08)	0.74 (0.04)	0.92 (0.04)	0.93 (0.03)	0.75 (0.06)	0.65 (0.06)	0.93 (0.02)
Cortical tubular regeneration/basophilia	0.61 (0.05)	0.51 (0.06)	0.68 (0.05)	0.66 (0.05)	0.59 (0.05)	0.64 (0.04)	0.63 (0.05)	0.79 (0.04)
CD degeneration or necrosis	0.65 (0.05)	0.74 (0.05)	0.94 (0.03)	0.53 (0.05)	0.89 (0.03)	0.57 (0.06)	0.96 (0.03)	0.86 (0.03)
CD deg/nec with no regen	0.68 (0.05)	0.62 (0.09)	0.88 (0.07)	0.51 (0.07)	0.87 (0.04)	0.65 (0.09)	0.87 (0.07)	0.80 (0.07)
CD deg/nec with regen	0.60 (0.07)	0.79 (0.06)	0.92 (0.02)	0.55 (0.06)	0.84 (0.04)	0.50 (0.07)	0.95 (0.01)	0.84 (0.03)
Medullary tubular regeneration/basophilia	0.63 (0.07)	0.61 (0.08)	0.76 (0.08)	0.54 (0.06)	0.79 (0.05)	0.56 (0.07)	0.82 (0.07)	0.76 (0.05)
Regeneration NOS with no degeneration	0.54 (0.06)	0.56 (0.07)	0.58 (0.07)	0.52 (0.05)	0.51 (0.06)	0.50 (0.06)	0.63 (0.08)	0.51 (0.06)
Intratubular casts, granular, cortex	0.56 (0.07)	0.54 (0.07)	0.80 (0.06)	0.57 (0.11)	0.69 (0.10)	0.79 (0.06)	0.73 (0.09)	0.60 (0.11)
Intratubular casts, hyaline, cortex	0.88 (0.06)	0.88 (0.06)	0.66 (0.06)	0.88 (0.07)	0.84 (0.07)	0.86 (0.06)	0.62 (0.07)	0.81 (0.08)
Inflammation, interstitial, chronic, cortex	0.58 (0.18)	0.50 (0.16)	0.62 (0.10)	0.74 (0.14)	0.83 (0.08)	0.51 (0.12)	0.61 (0.14)	0.71 (0.12)

Table 11 - AUC_{ROC} estimates (standard error) for pooled data

Pathology	BUN	SCr	NAG	Protein	α-GST	μ-GST	RPA	Clust
PT degeneration or necrosis	0.62 (0.04)	0.62 (0.04)	0.69 (0.03)	0.73 (0.04)	0.84 (0.03)	0.77 (0.03)	0.59 (0.03)	0.69 (0.03)
PT deg/nec with no regen	0.56 (0.05)	0.58 (0.04)	0.52 (0.04)	0.53 (0.06)	0.74 (0.04)	0.62 (0.04)	0.57 (0.04)	0.57 (0.04)
PT deg/nec with regen	0.79 (0.05)	0.82 (0.05)	0.87 (0.02)	0.89 (0.03)	0.87 (0.03)	0.87 (0.03)	0.76 (0.04)	0.94 (0.02)
Cortical tubular regeneration/basophilia	0.62 (0.04)	0.64 (0.04)	0.63 (0.04)	0.63 (0.03)	0.52 (0.04)	0.59 (0.04)	0.77 (0.03)	0.81 (0.03)
DT degeneration or necrosis	0.52 (0.06)	0.67 (0.04)	0.89 (0.03)	0.73 (0.05)	0.94 (0.03)	0.87 (0.04)	0.85 (0.02)	0.63 (0.06)
CD degeneration or necrosis	0.54 (0.04)	0.57 (0.04)	0.56 (0.06)	0.58 (0.04)	0.92 (0.02)	0.72 (0.04)	0.93 (0.02)	0.76 (0.03)
CD deg/nec with no regen	0.64 (0.05)	0.60 (0.07)	0.63 (0.11)	0.52 (0.06)	0.88 (0.03)	0.72 (0.06)	0.85 (0.06)	0.76 (0.06)
CD deg/nec with regen	0.51 (0.05)	0.55 (0.05)	0.52 (0.07)	0.61 (0.04)	0.90 (0.02)	0.70 (0.05)	0.92 (0.02)	0.73 (0.04)
Medullary tubular regeneration/basophilia	0.59 (0.07)	0.66 (0.07)	0.81 (0.07)	0.51 (0.06)	0.77 (0.05)	0.54 (0.06)	0.84 (0.07)	0.77 (0.04)
Regeneration NOS with no degeneration	0.52 (0.05)	0.58 (0.05)	0.57 (0.06)	0.52 (0.05)	0.52 (0.06)	0.56 (0.05)	0.53 (0.07)	0.56 (0.05)
Intratubular casts, granular, cortex	0.62 (0.09)	0.59 (0.08)	0.54 (0.11)	0.71 (0.08)	0.79 (0.07)	0.56 (0.11)	0.51 (0.10)	0.64 (0.09)
Intratubular casts, hyaline, cortex	0.79 (0.06)	0.82 (0.05)	0.70 (0.06)	0.78 (0.05)	0.69 (0.07)	0.76 (0.06)	0.71 (0.05)	0.83 (0.05)
Inflammation, interstitial, chronic, cortex	0.63 (0.04)	0.64 (0.04)	0.59 (0.04)	0.63 (0.04)	0.62 (0.04)	0.67 (0.04)	0.56 (0.04)	0.61 (0.04)

Based on these results and additional pairwise statistical comparisons of AUCROC for the novel BMs vs. reference BMs the HESI concluded:

- These results indicate the diagnostic value of an increase of urinary α-GST as a BM for PT injury
- The data set is insufficient to support any conclusion about the diagnostic value of μ-GST.
- The results shown clearly indicate both the specificity of RPA-1 for CD and its superior performance over all of the reference BMs for detection of degeneration/necrosis in the CD, particularly when regeneration is also present.
- The diagnostic value of clusterin for tubular regeneration is evident.

The incremental diagnostic value of each novel marker (α-GST, μ-GST, RPA-1, clusterin), when used in conjunction with reference markers, was assessed. Despite finding that some combinations of the novel markers with traditional markers enhanced diagnostic performance of the traditional markers for a given diagnosis, HESI concluded that the magnitude of the added value was minimal. This conclusion was based on comparison between AUCROC value for the combination of reference markers with novel biomarker to the AUCROC value for the novel marker alone.

Regulatory Data Assessment

The FDA and the European Medicines Agency contributed to the evaluation via the ad hoc appointed Biomarkers Qualification Teams (QTs) providing (via written procedures and Joint Videoconferences/ meetings with the FDA and the HESI representatives) elements for gap analysis, questions on the statistical evaluations and drafting the conclusions.

Gaps identified by CHMP in the current qualification exercise

The QT assessed the data presented by the Applicant and identified some gaps in the qualification exercise. The Applicant is encouraged to address the gaps in future investigations.

Analytical methods

- Some of the results of interference testing are missing [Hb, bilirubin and high salt for clusterin assay and metals (mercury, cadmium, lead, lithium, gadolinium) for all assays].

- The impact of the criteria for repeatability, intermediate precision and reproducibility on the diagnostic performance of the biomarkers was not evaluated.

Histopathology

- Histopathological reading was not fully blinded. Knowledge of treatment assignment can bias the results.

- Assessment of the histopathology would be more reliable if multiple sections from each kidney were examined and if histopathology process was standardised across studies.

Limitations of the studies to address specificity of the biomarkers for injury at a particular site

- The specificity of these BMs to kidney injury needs to be further investigated since other possible target organs were not investigated (i.e. α-GST present in the liver also, clusterin in the cytoplasm of interstitial macrophages within stomach, skeletal muscle, heart, tongue, as well as macrophages within the medulla of thymus and lymph node of an untreated control rat). Liver toxicity could potentially interfere in the evaluation of the diagnostic performance of these novel biomarkers. In the current exercise the examination of liver was not standardised between studies and the statements that the liver is not affected by the three compounds cannot be supported. In this context the assessment of the liver should be standardised in future dose finding and definitive studies. Furthermore the testing of an additional intermediate (between clearly toxic and non toxic) doses in the future dose finding studies could help define a more appropriate control group and possibly increase the power of the definitive studies to identify specific biomarkers of renal vs. liver toxicity.

- The number and type of nephrotoxic compounds in the studies was limited.

- There were no studies conducted with non-nephrotoxic compounds (e.g. hepatotoxins).

Biomarkers Normalisation

For all urinary markers, analyte concentrations for all animals were first normalised by dividing by the corresponding urine creatinine concentration. All individual animal marker values (normalised to creatinine in the case of urinary markers) were divided by the mean of the values in the concurrent control (i.e. vehicle-dosed) animals. Thus, all marker values were expressed as a fold-change versus the time-matched control group mean. Urine creatinine normalisation of BMs values is a standard practice and is considered acceptable. However normalisation of the urinary BMs by the mean of the values in the concurrent control is not recommended. It is acknowledged that this is done to minimise the impact of inter-study variability in the BMs performance. However, the BMs should be normalised

to the individual baseline BMs values. Since urine baseline data was not collected in this experiment it is recommended to conduct this normalisation in future studies. The Applicant argues that intra-animal variability is greater than inter-animal, suggesting that baseline data may be of limited value. The QT does not concur since the suggested baseline normalisation is easier to interpret from a patho-physiological perspective and would be more informative for the dynamic range of each biomarker and the effects of age, gender, diet and circadian rhythm.

Reproducibility of experiments

The QT notices the inconsistency between dose-finding and definitive studies for gentamicin and NPAA which makes the interpretation difficult.

Difference between strains and inference

The possibility of strain dependent sensitivity to nephrotoxicants and differential BMs response should be further investigated. The QT notices differences in the histopathological finding between the two rat strains. Based on the descriptive statistics strain differences in the BMs response are observed. For α-GST the correlation between severity of histopathological findings and BM fold change is evident for Wistar but not for Sprague Dawley, likewise (though perhaps not as clearly) for RPA-1. For clusterin the correlation is more evident for Sprague Dawley. Inconsistency is also observed between strains in the AUCROC values. Consequently pooling together the results from the two strains is not considered optimal and complicates inference.

Extrapolation of findings to female rats is not possible

Only male animals were used which limits the scope of the qualification.

Unexpected findings

Consistent with the immunohistochemistry localisation of α-GST to the proximal tubule, increases in urinary α-GST were seen with PT injury in the absence of CD injury. However when isolated CD injury was induced by NPAA, α-GST values were consistently decreased in urine in both strains and α-GST was superior to all the reference BMs for the diagnosis of CD injury in the absence of PT injury. The opposing effects of the proximal and collecting duct injury on α-GST levels are not adequately understood and their impact on the diagnostic performance of the BM is not evaluated.

-It could be useful to revisit samples to understand elevation of biomarker levels in the absence of histopathological changes.

Replication of evidence

Associations identified in experiments can arise due to chance, in particular when multiple comparisons are made. In these experiments some adjustments for multiple testing were made. However, the conclusions drawn could be made more robust if replicated evidence was available from another, similar series of experiments.

CHMP Qualification Opinion

Clusterin was previously qualified by the FDA and the European Medicines agency after review of the PSTC submission. (published report:

http://www.emea.europa.eu/pdfs/human/sciadvice/67971908en.pdf):

"The urinary kidney BMs (Kim-1, Albumin, Total Protein, β2-Microglobulin, urinary clusterin, Urinary Trefoil Factor 3 and urinary Cystatin C) are considered acceptable in the context of non-clinical drug development for the detection of acute drug-induced nephrotoxicity, either tubular or glomerular with associated tubular involvement.

They provide additional and complementary information to BUN and serum creatinine to correlate with histo-pathological alterations considered to be the gold standard.

Additional data on the correlation between the BMs and the evolution and reversibility, of acute kidney injury are needed. Also, further knowledge on species-specificity is required."

The findings of the current HESI submission increase the level of evidence supporting the use of urinary clusterin. Clusterin is a biomarker that may be used by Applicants to detect acute drug-induced renal tubule alterations, particularly when regeneration is present, in male rats and can be included along with traditional clinical chemistry markers and histopathology in GLP toxicology studies which are used to support renal safety in clinical trials.

In addition the HESI data indicate that RPA-1 is a biomarker that may be used to detect acute drug-induced renal tubular alterations, particularly in the collecting duct, in male rats and can be included along with traditional clinical chemistry markers and histopathology in GLP toxicology studies which are used to support renal safety in clinical trials.

The QT acknowledges that the HESI data may support the use of α-GST in detecting proximal tubule injury in male rats. However the opposing effects of proximal and collecting duct injury on α-GST levels raise uncertainty about the usefulness of this biomarker for detecting early mild renal injury. Therefore before α-GST is qualified in this context further studies will be needed to evaluate the mechanistic basis and usefulness of this BM.

CHMP Recommendations towards future qualification experiments

Furthermore the QT agrees with HESI on the importance of the following future investigations:

- Prodromal claims (BM to detect injury prior to histopathology changes)

- Claims on the reversibility

- Claims following the chronic administration of nephrotoxicants

- BMs to report injury to the other parts of the nephron

- Extension of work to non rodent species

- Combinations of novel and/or conventional BMs to optimize diagnostic performance

HESI is encouraged to seek a qualification advice on these claims.

The extension of this exercise into the evaluation of use of novel BMs for renal injury in the translational and clinical context is of great importance and could be also the topic of a future qualification advice.