Follow-up: biomarker test case

By *Steve Usdin, Senior Editor*

A large-scale effort by a consortium of pharma companies to qualify seven well-characterized nephrotoxicity biomarkers in order to support regulatory applications has resulted in a green light from the **FDA** and **European Medicines Evaluation Agency** for using the biomarkers in preclinical testing and a yellow light for bridging between animal studies and Phase I studies.

A great deal of additional research will be required to make their use routine in clinical trials, and far more extensive testing would have to be performed in order for all or some of the biomarkers to help guide medical decisions.

The conclusions reached by the FDA and European Medicines Evaluation Agency (EMA) demonstrate the length and complexity of the path a possible biomarker will have to traverse from the pages of an academic journal to its use in clinical trials, appearance on drug labels and integration into clinical medicine.

The submission from the Predictive Safety Testing Consortium (PSTC), which is managed by the **Critical Path Institute** (C-Path), marked the first use of the new regulatory processes that U.S. and European regulators have established to qualify biomarkers for preclinical and clinical uses.

The biomarkers—kidney injury molecule-1 (KIM-1); albumin; total protein; β_2 -microglobulin (β_2M); urinary clusterin; urinary trefoil factor 3 (TFF3); and urinary cystatin C—were selected from among 23 candidates identified by PSTC members.¹

The ultimate goal of adding the new biomarkers is to increase the sensitivity and specificity of tests for kidney damage caused by compounds in preclinical development. Changes in serum creatinine and blood urea nitrogen (BUN) typically show up only after there is permanent organ damage.

As a result, when these markers show up in preclinical studies, compounds are often killed before reaching the clinic. Having biomarkers that provide earlier signals could make it possible to develop compounds that otherwise would have to be dropped, because patients could be monitored for adverse effects.

Janet Woodcock, director of the FDA's Center for Drug Evaluation and Research (CDER), and John Jenkins, director of the agency's Office of New Drugs, signed an April 14 letter to C-Path approving the qualification of the biomarkers. But the FDA asked the consortium to defer an announcement until the agency was ready to release a full package of information, Federico Goodsaid, associate director for operations in

genomics at CDER, told *SciBX* last week.

The EMA published a report on May 23 that outlined its approval of the PSTC's submission, pending a two-month public consultation period.²

According to Goodsaid, the agencies came to the same conclusions.

The EMA report states that the seven biomarkers "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."

It adds that the markers are complementary to the "gold standard" biomarkers serum creatinine and BUN.

The regulators also accepted the consortium's request to qualify five of the seven submitted biomarkers as bridging markers for early clinical studies on a case-by-case basis when concerns are generated by findings in good laboratory practice (GLP) animal toxicology studies. Clusterin and TFF3, which have not been studied clinically, were excluded from this qualification.

The FDA and EMA also were not persuaded to extend either the qualification of the biomarkers' ability to monitor progression of kidney injury or their use in showing that renal injury has been reversed.

The regulators provided feedback on additional work that could broaden the qualification of the biomarkers.

"FDA and EMA have qualified the biomarkers for use in conjunction with existing markers, BUN and serum creatinine, but the hope is that as we generate additional data, they will eventually allow us to replace those less-sensitive markers," Kevin Carl, senior associate director of drug regulatory affairs at **Novartis AG**, told *SciBX*.

Novartis and **Merck & Co. Inc.** took the lead among PSTC members in conducting studies of the analytical validity of the biomarkers.

Regulatory specificity

The EMA and FDA approvals are very specific about the uses for which the biomarkers have been authorized.

According to the EMA, the accepted renal biomarkers "provide additional and complementary information" to BUN and serum creatinine when used in nonclinical research.

William Mattes, C-Path's director of toxicology, predicted that this initial qualification will change the way companies conduct preclinical research. It "allows companies to run a GLP study and include one of the markers without worrying that it will be misinterpreted by the regulatory bodies," he said.

Indeed, most of the newly qualified biomarkers have been in the academic toolkit for a long time. But the lack of scientific and regulatory consensus on their sensitivity and specificity has been a barrier to their use in animal studies. This is largely because companies must turn over to the FDA and EMA all of the data they generate in GLP studies of investigational drugs. Companies are unlikely to include an endpoint in a study that could be misinterpreted, leading regulators to delay or reject approval of an IND.

"Some teams were hesitant to run these assays in GLP studies because

they didn't know how to interpret the data," Frank Dieterle, head of external affairs for investigative toxicology at Novartis, told *SciBX*. "Now we have opened the door. Before, nobody knew if a twofold increase in one of the markers was significant or not. Now we have rules that allow us to interpret the biomarkers."

Further downstream, the EMEA said use of the PSTC's "renal biomarkers in clinical trials is to be considered on a case-by-case basis in order to gather further data to qualify their usefulness in monitoring drug-induced renal toxicity in man."

Similarly, according to Goodson, the FDA's approval letter "said the new biomarkers, which are much more sensitive than BUN or creatinine, could be used on a case-by-case basis at the discretion of the clinical division in some Phase I trials."

Mattes suggested that even the preclinical qualification may offer a resolution to a frequent barrier to getting a compound into the clinic. Companies often face a "conundrum when they see something in pre-clinical safety studies that they don't have any way to measure in humans," he said.

Thus, the FDA's qualification will make it possible in some instances to gain approval for a Phase I trial if a sponsor can provide a good explanation for why a signal in animal data isn't likely to predict toxicity in humans.

Merck, which conducted 20 of the 34 rat studies submitted in support of the renal biomarkers, has already integrated the qualification into its drug development, said Frank Sistare, executive director of safety assessment.

"We have already to some extent applied these tools in a way that has allowed certain programs to proceed," he said.

Initially, Sistare said, the biomarkers will be proposed for use clinically "in carefully selected circumstances," for example if kidney toxicity is seen in one species but not another.

"You might not be certain that the human is more like the rat, where kidney toxicity may be seen, or the dog, where there is no kidney injury, and you have other data to suggest that kidney toxicity might not be a clinical problem," he said.

The qualified biomarkers "provide us with tools to be able to convince clinicians and patients that a compound can be tested safely," Sistare said. "If we didn't have these tools, we might have to wait to test a backup compound that is a year or two behind in development."

Data limitations

Comments by the EMEA and FDA on the limitations of the dataset, and their suggestions of steps that could lead to qualification of additional claims, are at least as important as the claims that were accepted.

The EMEA's report notes that regulators were not convinced the data support use of the biomarkers to track the progression of kidney injury. Showing that the markers can track progression is critical to gaining a prognostic claim.

Although the consortium "provided information on dose- and time-dependent changes in the biomarkers and the appearance of histopathological alterations during periods of dosing, there is insufficient data to establish a clear correlation between the [biomarkers] and the evolution

of the nephrotoxic alterations over time, as documented by histopathology," the EMEA said.

To achieve claims involving progression, the EMEA said the PSTC will have to submit studies that establish temporal correlations between kidney lesions and biomarkers.

The EMEA also concluded that the consortium had not documented the ability to show reversal of injury, stating that "reversibility of the biochemical changes is insufficiently correlated with kidney function recovery. Therefore the use of these [biomarkers] in 'monitoring' renal toxicity at this stage is not sufficiently demonstrated."

To demonstrate reversibility, the FDA's Goodson said a sponsor would have to submit data from animal experiments showing that the biomarker returns to normal levels as kidney damage goes away.

"Sponsors will have some responsibility to demonstrate reversibility for their particular compound in preclinical testing," if they want to use the biomarkers as a bridging tool to Phase I trials, said Carl. The sponsor will have to "demonstrate that it can manage the injury preclinically, that when dosing is stopped, the lesions resolve and the markers correlate accordingly."

"These tools have been developed in the rat. Tools are needed in dogs, nonhuman primates and mice. So there is still more work to do in the preclinical space."

—Frank Sistare, Merck & Co. Inc.

Next steps

Mattes said one of the consortium's key goals is to demonstrate that the biomarkers are sufficiently sensitive to detect drug-induced damage to the kidney at a stage when it is still reversible.

The PSTC also is working to develop data to support claims that the biomarkers have prognostic value. It is also developing plans to expand the preclinical claims and to use the

markers in clinical trials that could produce data to broaden their clinical use.

"Even in the preclinical situation, the biomarkers have been qualified in a narrow context," Sistare noted.

He said the consortium is considering expanding the "scope in the preclinical area to beyond tubular and glomerular injury to other types of kidney toxicity."

In addition, he said, "these tools have been developed in the rat. Tools are needed in dogs, nonhuman primates and mice. So there is still more work to do in the preclinical space."

As to use in clinical development, Sistare noted, "the health authorities have requested a more prospective trial design that will give them greater confidence in some of the details. A couple of clinical studies are being planned right now and hopefully will have some results within 12–18 months."

According to Mattes, such studies probably will involve adding the collection of data on the biomarkers to studies of known nephrotoxins, such as the antibiotic gentamycin and the chemotherapy cisplatin.

Lessons learned

The PSTC worked through several methodological issues with the FDA and EMEA, setting precedents that could speed future submissions.

For example, there was some uncertainty about whether statistical analysis of the submission should include a statistical penalty for multiplicity. A multiplicity penalty is often imposed to compensate for the

possibility that screening for a wide variety of associations, or data dredging, can produce false positives.

The FDA and EMEA agreed that the biomarker submissions were “not a screening exercise, so there is no need for multiplicity adjustment,” said Dieterle. “For each biomarker we already have evidence that it is a useful tool, and every biomarker is qualified on its own.”

In addition, according to Novartis and Merck, the FDA requested use of histopathology methods that are more rigorous than those typically used for drug applications.

The principal data supporting the application consisted of experiments in which rats were exposed to a drug that is known to cause nephrotoxicity. The values for the proposed biomarkers, as well as BUN and creatinine, were correlated with histopathology assessments of kidney samples.

According to Sistare, the consortium submitted data using the standard method specified by the **Society of Toxicologic Pathology**. This involved first examining the slides of control animals to get a feel for the background noise and to optimize sensitivity and then analyzing slides of high-dose animals to see if there were treatment-related differences. The next step, if a difference was detected, was to blind the slides, read and score them.

“The feedback from the FDA review group was that the whole thing should be blinded from beginning to end, for all the organs that are examined, and not just kidney,” he said. “They didn’t want any first look at the controls and the high dose; they wanted every single slide scored without first calibrating against controls.”

To address the FDA’s concerns, the PSTC conducted two sets of histopathology analyses.

“We ended up following the standard practice, blinding all slides after calibrating the eye against slides from controls and high-dose groups,” Sistare said. “After the FDA questioned whether there might be some bias introduced by doing it that way, they asked for another approach. We sent our slides to a pathologist at **SRI International** who reread all the slides, blinded. The results were very concordant.”

A disagreement over appropriate statistical methods led to different conclusions by the EMEA and the consortium, and apparently between the two regulatory bodies, about the performance of one of the biomarkers, TFF3.

The consortium claimed in its submission that all seven biomarkers were more sensitive than BUN and creatinine, but the EMEA concluded the data did not support that claim for TFF3.

The disagreement centered on application of receiver operating characteristic (ROC) statistical analysis, which is used to assess sensitivity and specificity by comparing true-positive and false-positive rates. According to Sistare, the EMEA focused on “inclusion” ROC whereas the consortium argued for “exclusion” ROC.

“The exclusion ROC analysis, which we favor, excludes data from the ROC curve for animals where a biomarker changes significantly at early time points before histopathology is present in that animal,” he said. “The

inclusion analysis includes that data as a false positive, so it hurts the apparent performance of the biomarker.”

Sistare said that although TFF3’s performance is not better than BUN or creatinine based on an inclusion analysis, “if you look at the exclusion analysis, TFF3 also will reach the threshold of showing itself to outperform BUN and serum creatinine.”

The FDA’s approval letter states that all seven biomarkers outperformed BUN and creatinine, which suggests that the agency came to a different conclusion than the EMEA, Sistare said.

More applications

The regulators have agreed to accept common applications and to collaborate closely on reviews.

According to the FDA’s Goodson, three other biomarker qualification applications have been submitted to the two agencies, which also have had discussions with several other groups that are planning to submit applications.

One of the pending applications, from the **ILSI Health and Environmental Sciences Institute** (HESI), seeks approval of three preclinical urinary nephrotoxicity markers that were not included in the PSTC submission, as well as clusterin.

HESI is a not-for-profit scientific organization that brings together academic, industrial and government scientists to reach consensus on scientific questions around human health, toxicology, risk assessment and the environment.

The application is for preclinical qualification of immunoassays for the urinary markers renal papillary antigen-1 (RPA-1), glutathione-S-transferase- α (GST- α), GST- μ and clusterin. It was submitted on May 1 and will be the subject of a July 2 meeting with the FDA

and EMEA, Syril Pettit, HESI senior scientific program manager, told *SciBX*.

Meanwhile, the European regulatory agency is seeking public comment on a proposed pathway for qualifying biomarkers and for providing scientific guidance on qualification protocols.³

REFERENCES

1. Usdin, S. *SciBX* 1(13), 1–5; April 24, 2008
2. European Medicines Agency. Final report on the pilot joint EMEA/FDA VXDS experience on qualification of nephrotoxicity biomarkers. <<http://www.emea.europa.eu/pdfs/human/biomarkers/25088508en.pdf>> (May 23, 2008)
3. European Medicines Agency. Biomarkers qualification: guidance to applicants (draft). <<http://www.emea.europa.eu/pdfs/human/biomarkers/7289408en.pdf>> (April 24, 2008)

COMPANIES AND INSTITUTIONS MENTIONED

Critical Path Institute, Tucson, Ariz. and Rockville, Md.
European Medicines Evaluation Agency, London, U.K.
FDA, Rockville, Md.
ILSI Health and Environmental Sciences Institute, Washington, D.C.
Merck & Co. Inc. (NYSE:MRK), Whitehouse Station, N.J.
Novartis AG (NYSE:NVS; SWX:NOVN), Basel, Switzerland
Society of Toxicologic Pathology, Reston, Va.
SRI International, Menlo Park, Calif.

“FDA and EMEA have qualified the biomarkers for use in conjunction with existing markers, BUN and serum creatinine, but the hope is that as we generate additional data, they will eventually allow us to replace those less-sensitive markers.”

—Kevin Carl, Novartis AG