Adaptive Bayesian Analysis of Serum Creatinine as a Marker for Drug-Induced Renal Impairment in an Early-Phase Clinical Trial

Pierre-Edouard Sottas,1, * Gordon F. Kapke,2 and Jean-Marc Leroux1
1Covance Central Laboratory Services, Meyrin, Switzerland; 2Covance Central Laboratory Services, Indianapolis, IN; *address correspondence to this author at: Covance CLS, rue Moïse-Marcinhes 7 Meyrin GE, 1217 Switzerland. Fax: +41-58-822-61-01; e-mail pesottas@gmail.com.

BACKGROUND: A concern with using creatinine for the identification of drug-induced renal impairment is that small changes in serum creatinine (SCr) that frequently are perceived as measurement bias or imprecision translate into important changes in the glomerular filtration rate. Important drug-generated changes in creatinine are difficult to detect because they are frequently observed within the reference interval. The design of a crossover drug protocol is an opportunity to use study participants as their own control to identify these small but important changes.

METHODS: Twenty individuals participating in a phase I clinical trial were evaluated for SCr changes beyond those expected for biological variation according to individual Z scores derived from an adaptive Bayesian model. After 2 screening tests, participants were administered either drug (n = 11) or placebo (n = 9) during the first dosing interval. A washout period followed, and drug was then administered to the group that initially received placebo, and vice versa (10 visits total per participant).

RESULTS: Although all creatinine values fell within the reference interval, 8 participants individually showed increased concentrations (Z scores >2.33). These 8 participants were confirmed at unblinding to have received the drug in the identified dosing period, with 1 exception.

CONCLUSIONS: The ability to identify a drug effect on an individual-participant basis in early-phase studies permits drug developers to recognize issues early in development and rapidly engage in risk–benefit analysis. These results suggest that SCr monitoring is able to detect early kidney dysfunction when individual-based reference intervals are used.

The traditional renal biomarker creatinine is viewed as being insensitive to renal impairment (1–3). Intraindividual biological variation for serum creatinine (SCr)3 (CV, 5%–6%) is small relative to group biological variation (CV, 15%) (4); hence, substantial renal impairment can occur with SCr results within the reference interval. Because of this small intraindividual variation, SCr has been proposed for serial monitoring of renal function in individuals with a stable muscle mass (5). SCr is inversely proportional to an individual’s glomerular filtration rate (GFR) because of first-order clearance. For instance, with GFR-estimating equations based on SCr, a 10% increase in SCr implies a 10%–12% decrease in the GFR.

Important questions are what the variation in GFR is and how SCr can be used to model it (6). SCr is a relatively inexpensive laboratory test, is readily available in the clinical laboratory, and is stable during sample transport from patient to laboratory. The challenge in drug development is to obtain actionable information more effectively from SCr as a biomarker of renal function. Given that interindividual variation in SCr is substantially greater than intraindividual variation, the sensitivity of SCr as a marker is improved by using individuals as their own reference. A large fraction of interindividual variation can be removed after stratification of reference intervals in terms of such heterogeneous factors as age, sex, and disease. In addition, the availability of previous measurements for an individual allows a decision maker to use an adaptive Bayesian framework to move progressively from population-based to patient-based reference intervals (7–9). In that context, the crossover design of a clinical trial protocol permits individuals—from the perspective of statistical analysis—to be their own control.

We have analyzed longitudinal SCr data obtained in a double-blind crossover phase I trial. The analysis is retrospective and is motivated by the fact that a renal impairment was observed in a large later-phase trial. In early-phase studies in which the number of patients is limited, the ability to identify statistically significant and medically important changes in SCr is difficult, owing to the analytical and biological variation. The goals are therefore twofold: first, to determine if an adaptive Bayesian framework can be used to generate better information for drug developers and, second, to see if individual-based reference intervals for SCr can be used to detect early kidney dysfunction.

Twenty healthy women 24 to 37 years of age participated in a regulated clinical phase I trial conducted in accordance with the Declaration of Helsinki. As part of a panel of drug safety biomarkers, SCr was measured in a

3 Nonstandard abbreviations: SCr, serum creatinine; GFR, glomerular filtration rate; BN, Bayesian network; RCV, relative change value.
clinical trial laboratory in accordance with good clinical laboratory practices. After 2 screening tests (SC1, SC2), participants were administered drug \( n_{H11005} 11 \) or placebo \( n_{H11005} 9 \) during the first dosing interval (3 visits: T1V1, T1V2, T1V3). A washout period (WO1) followed, and drug was administered to the group that initially received placebo, and vice versa (3 visits: T2V1, T2V2, T2V3). Finally, all participants were tested 4 weeks after the second round of drug administration (WO2), for a total of 10 tests per participant in 4 months.

We used a hierarchical Bayesian network (BN), which is described in detail elsewhere \((7–9)\). The BN is customized for biomarker and study participant. The first level of the BN integrates the heterogeneous factors of age and sex to generate stratified population distributions. For women 18 to 50 years of age, the stratified reference interval was 0.35–1.14 mg/dL (31–101 \( \mu \)mol/L). The second level of the BN includes hidden variables for adaptively moving in a continuous data stream from stratified prior population distributions to individual-specific posterior distributions. Prior distributions of intraindividual and interindividual components of variation were assumed to be normally distributed with an initial mean corresponding to CVs of 5.3% and 14.2%, respectively \((4)\). Individual screening values for SCr were integrated as hard evidence in the BN to yield posterior distributions of expected values. Contrary to other biomarkers, such as hemoglobin concentration \((7)\), how intraindividual variations in SCr change from one individual to another have been poorly studied. We allowed the adaptive Bayesian model to progressively learn the intraindividual variation for each individual. We then computed reference intervals for each individual, which were defined as the 0.5%–99.5% interval of the posterior distribution of SCr. Individual \( Z \) scores were then computed for all subsequent visits. Additionally, we generated the posterior distributions of expected sequences of length 3 to evaluate the series consisting of the 3 SCr values for the 2 dosing intervals. \( Z \) scores \( >2.33 \) were considered abnormal. For comparison purposes, we applied a 10% change from baseline as well as a reference change value (RCV) equal to 17.5% to the data. All statistical simulations were performed with MATLAB software (version 7.10.0) with Statistics Toolbox version 7.3 (MathWorks).

Fig. 1 presents the sequence of SCr values for study participant 15, together with the limits produced with
the adaptive Bayesian model. Table 1 tabulates the data for the 20 patients during the crossover protocol. Visits SC1 and SC2 returned no Z scores > 1.01. Visit WO1 is at the end of a 4-week washout period; the highest observed Z score is 1.89.

All SCr values fell within the stratified reference interval of 0.35–1.14 mg/dL (31–101 μmol/L). On the other hand, 12 values (8 participants) fell outside the individual reference intervals (Table 1). Nine of these values were obtained during drug treatment; 2 values (participants 3 and 10) obtained after a dosing interval may have been due to a residual drug effect, and 1 value (participant 9) was obtained when the individual was receiving placebo. Twelve participants did not present any significant increase in SCr, although 17 participants (85%) presented higher sequence Z score values during the dosing period than when receiving placebo.

For comparison purposes, the numbers of values 10% higher than the baseline were as follows: 18 (11 participants) during the dosing interval, 2 (2 participants) at WO1 after a dosing interval, 12 (6 participants) when receiving placebo, and 2 (2 participants) at WO2. Similarly, the numbers of values higher than the RCV were as follows: 5 (4 participants) during the dosing interval, 2 (2 participants) after a dosing interval, and 1 (1 participant) when receiving placebo.

As the P values in Table 1 show, we observed no significant change in raw SCr concentrations at any visit during the 2 dosing periods. On the other hand, individual Z scores were significantly higher in 5 of the 6 visits when the drug was administered. These results suggest that the use of population-based reference intervals for SCr do not detect early kidney dysfunction, whereas the opposite is true when individual-based reference intervals are used.

Participant 9 had unusual SCr sequences, with Z scores > 2.33 for both dosing intervals. In visits T1V1 and T2V3, this participant’s individual Z scores were > 2.33. At unblinding, the value of 1.07 mg/dL (94.7 μmol/L) (Z score = 2.93; 17% increase from baseline) measured at T1V1 appeared not to have been induced by the drug. If this abnormal value, which is based on an individual reference interval, would have triggered repeat analysis/re-collection, there would be no need to question the data retrospectively.

In the second dosing interval, 6 participants presented 1 individual Z score or sequence Z score > 2.33. The SCr concentration for participant 5 increased from a baseline mean of 0.87 mg/dL (76.5 μmol/L) to a 3-visit mean of 0.99 mg/dL (87.9 μmol/L)—a 15% increase. The SCr concentration for participant 6 increased from a baseline mean of 0.89 mg/dL (78.3 μmol/L) to a 3-visit mean of 1.02 mg/dL (89.9 μmol/L)—a 15% increase. The SCr concentration for participant 10 increased from a baseline mean of 0.85 mg/dL (75.2 μmol/L) to a 3-visit mean of 0.96 mg/dL (85.2 μmol/L), or a 13% increase. Owing to first-order clearance, the magnitude of the change in the participants’ GFR is close to the percentage increase in the SCr.

Two aspects of early-phase drug studies continue to challenge drug development. First, are there reliable biomarkers for identifying important safety issues? Second,

<table>
<thead>
<tr>
<th>SCrα</th>
<th>Individual Z scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD), Range,</td>
<td>0.06 (0.52), 0.82 – 1.01</td>
</tr>
<tr>
<td>mg/dL</td>
<td>mg/dL</td>
</tr>
<tr>
<td>SC1</td>
<td>0.80 (0.13), 0.63 – 1.04</td>
</tr>
<tr>
<td>SC2</td>
<td>0.81 (0.11), 0.65 – 1.09</td>
</tr>
<tr>
<td>T1V1</td>
<td>0.82 (0.12), 0.65 – 1.11</td>
</tr>
<tr>
<td>T1V2</td>
<td>0.84 (0.10), 0.66 – 1.05</td>
</tr>
<tr>
<td>T1V3</td>
<td>0.85 (0.10), 0.69 – 1.07</td>
</tr>
<tr>
<td>WO1</td>
<td>0.82 (0.12), 0.62 – 1.10</td>
</tr>
<tr>
<td>T2V1</td>
<td>0.84 (0.13), 0.64 – 1.00</td>
</tr>
<tr>
<td>T2V2</td>
<td>0.85 (0.13), 0.66 – 1.03</td>
</tr>
<tr>
<td>T2V3</td>
<td>0.84 (0.15), 0.61 – 1.10</td>
</tr>
<tr>
<td>WO2</td>
<td>0.82 (0.13), 0.62 – 1.06</td>
</tr>
</tbody>
</table>

α The factor for converting creatinine in milligrams per deciliter to micromoles per liter is: × 1003/H11003.

β Testing points for each individual: SC1, first screening test; SC2, second screening test; T1V1, first test during first dosing interval; T1V2, second test during first dosing interval; T1V3, third test during first dosing interval; WO1, test after first washout period; T2V1, first test during second dosing interval (after crossover); T2V2, second test during second dosing interval (after crossover); T2V3, third test during second dosing interval (after crossover); WO2, test 4 weeks after the second round of drug administration.

The SCr concentration for participant 6 increased from a baseline mean of 0.89 mg/dL (78.3 μmol/L)—a 15% increase from baseline) measured at T1V1 appeared not to have been induced by the drug. If this abnormal value, which is based on an individual reference interval, would have triggered repeat analysis/re-collection, there would be no need to question the data retrospectively.

In the second dosing interval, 6 participants presented 1 individual Z score or sequence Z score > 2.33. The SCr concentration for participant 5 increased from a baseline mean of 0.87 mg/dL (76.5 μmol/L) to a 3-visit mean of 0.99 mg/dL (87.9 μmol/L)—a 15% increase. The SCr concentration for participant 6 increased from a baseline mean of 0.89 mg/dL (78.3 μmol/L) to a 3-visit mean of 1.02 mg/dL (89.9 μmol/L)—a 15% increase. The SCr concentration for participant 10 increased from a baseline mean of 0.85 mg/dL (75.2 μmol/L) to a 3-visit mean of 0.96 mg/dL (85.2 μmol/L), or a 13% increase. Owing to first-order clearance, the magnitude of the change in the participants’ GFR is close to the percentage increase in the SCr.
can important safety issues be identified in small, early-phase populations? SCr is viewed as a marginal biomarker of renal function (1). With clever protocol design and use of individualized reference intervals, SCr should meet the criteria for being sensitive to drug-induced renal impairment in small, early-phase populations.

On unblinding the dosing information, we observed that the adaptive Bayesian approach was able to identify 8 of the 20 patients during the drug-administration period for a drug without prior documentation of renal impairment. This renal impairment was later observed in a large phase III protocol. The observed changes in SCr were in the range of 10% to 20%, which translates to absolute values of 0.11–0.22 mg/dL (10–20 μmol/L). More than 60% of the time, a change in SCr of 0.22 mg/dL (20 μmol/L) would maintain the participant’s SCr concentration within the population reference interval and therefore be interpreted as normal. Unsurprisingly, the more customized adaptive Bayesian approach led to a better relationship of sensitivity to specificity than a 10% increase from baseline and the use of an RCV. These latter 2 approaches should nevertheless be favored over the less sensitive population-based reference intervals.

Interestingly, SCr concentrations in 12 participants were not notably affected by administration of the drug, despite the use of sensitive individual reference intervals. These results suggest that drug-induced renal impairment was present in only about half of the participants. The personalization of a biomarker signal makes possible the evaluation of drug efficacy and safety on an individual basis and, in turn, facilitates the tailoring of drug therapy to a dosage that is most appropriate for the individual patient. Correlations with genotyping data for metabolic pathways involved in drug metabolism may be searched for further improvements in personalization.

The use of individual reference intervals as obtained with an adaptive Bayesian approach can identify drug-induced changes in an individual’s SCr values. In addition, from a laboratory perspective the use of single-visit individual Z scores can help identify unexpected results for an individual that may then be confirmed within the laboratory by repeat analysis and/or collection of additional samples in real time.

Although the Predictive Safety Testing Consortium has expended much effort in evaluating the next generation of renal biomarkers (10), our study suggests that a conventional biomarker like SCr is capable of detecting early kidney dysfunction when individual reference intervals are used. We encourage the use of methods based on intraindividual variation, such as a 10% increase from baseline, RCVs, and, ultimately, the adaptive Bayesian model. A drawback of the Bayesian model is the requirement of prior population data to narrow the limits of statistical uncertainty; however, the use of screening data permits not only computation of individual reference intervals but also verification of prior statistical assumptions at the population level. If prior statistical assumptions are incorrect, statistical analysis of the data will provide conclusions inconsistent with the study design. In addition, a higher number of baseline samples leads to higher personalization of the evaluation obtained with the Bayesian model.

A biological signal is not necessarily clinically relevant. Although our study suggests that the sensitivity of SCr can be improved substantially with the use of individual reference intervals, GFR presents some limitations as a biomarker of nephrotoxicity. Kidney disease increases the SCr concentration, but the opposite is not necessarily true. In prerenal azotemia, an SCr increase is not symptomatic of an established structural kidney injury. Therefore, we recommend that any call of an SCr increase as significant be accompanied by changes in other biomarkers of tubular or glomerular toxicity (10).

In the current highly competitive economic environment, the use of individual reference intervals is important because it reduces the study size for the same statistical power. For example, 600 individuals are required to detect an absolute change in the SCr concentration of 0.06 mg/dL (5 μmol/L) (<10% of the reference interval) in an unbalanced randomized study (2:1) with 2 screening visits at a significance level of 0.05 and a power of 99% with a population-based approach (unpaired t-test); 210 individuals are required for an analysis of covariance with baseline values as covariates; and 150 patients are required for an unpaired t-test applied to the standard scores derived from the posterior distributions returned by the adaptive Bayesian model. Clearly, the use of only a fourth of the individuals would allow a substantial reduction in study costs.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: G.F. Kapke, Covance; J.-M. Leroux, Covance.

Consultant or Advisory Role: P.-E. Sottas, Covance.


Honoraria: None declared.

Research Funding: None declared.

Expert Testimony: None declared.


Role of Sponsor: No sponsor was declared.
References


Previously published online at DOI: 10.1373/clinchem.2012.193698