*Research Article*

# Analysis of Imprecision in Incurred Sample Reanalysis for Small Molecules

Sriram Subramaniam,[1,4] Devvrat Patel,[1] Barbara M. Davit,[2] and Dale P. Conner[3]

**Abstract.** Over the years, incurred sample (IS) reanalysis (ISR) has become a tool to confirm the reliability of bioanalytical measurements. The recommendation for ISR acceptance criterion for small molecules is at least 67% of ISR samples that have reanalyzed concentrations within 20% of their original concentrations when normalized to their means. To understand the relevance of the ISR acceptance criterion and sample size requirements, simulated ISR studies evaluated the probability of ISR studies passing the acceptance criterion (ISR pass rate) as a function of IS imprecision and sample size. When IS imprecision (percent coefficient of variation: %CV) is low (≤10 or 1–10% CV), high ISR pass rate (≥99%) is attained with <50 samples. At intermediate IS imprecision (*e.g.*, 12% CV or 7–12% CV range), 80–160 samples are required for a high ISR pass rate. When IS imprecision is at the higher end of the acceptance limit, ISR pass rate decreases significantly, and increasing sample size fails to achieve high ISR pass rate. The effect of systematic bias (*e.g.*, instability, interconversion) on ISR pass rate is strongly dependent on sample size at intermediate IS imprecision. The results provide an understanding of the effect of IS imprecision on ISR pass rates and a framework for selection of ISR sample sizes.

**KEY WORDS:** bias; imprecision; incurred sample reanalysis; sample size; small molecules.

## INTRODUCTION

Bioanalytical methods are routinely used in bioequivalence (BE), pharmacokinetic (PK), and toxicokinetic (TK) studies to generate concentration data in biological matrices for exposure analysis. Meaningful interpretation of BE, PK, and TK results is directly dependent on the reliability of the bioanalytical methods. Consequently, bioanalytical methods in BE, PK, and TK studies have to be accurate, precise, and sensitive to measure the actual drug concentrations achieved in the body. The performance of bioanalytical methods is routinely monitored by assessing the performance of quality control (QC) samples. However, over the last few years, there have been several observations that indicate monitoring performance of QCs may not always guarantee assay performance during analysis of study samples. Broadly, several factors including assay conduct, drug characteristics, and matrix effects may contribute to the above discrepancy (1–4). The concept of incurred samples reanalysis (ISR) was discussed at the American Association of Pharmaceutical

Scientists (AAPS) workshop in Crystal City, VA in 2006 and other meetings as a tool to confirm the reliability of the bioanalytical methods during analysis of study samples (5). The recommendations for the implementation of ISR conduct were reached at the 2008 AAPS workshop in Crystal City, VA (2008 ISR Workshop) (6). These issues were further discussed at other meetings (7,8). Over the years, ISR has become an integral part of bioanalysis.

ISR is recommended for preclinical and clinical studies where PK assessment is the primary end point for all bioavailability (BA) and BE studies. Since the true concentrations of study samples are unknown, the reproducibility assessment in ISR serves as a surrogate to understand the accuracy of the study samples. The 2008 ISR Workshop recommendations for ISR conduct involve reanalysis of 5–10% of the study samples and assessment of reproducibility against their original concentrations. According to the 2008 ISR Workshop recommendations, ISR is acceptable for small molecules if at least two thirds of the reanalyzed concentrations are within 20% of their original concentrations [*i.e.*, (repeat−original)/(average of repeat and original)×100]. These ISR recommendations have been adopted by industry over the years and incorporated in guidance issued by the European Medicines Agency and Health Canada (9,10) and the recently issued draft guidance by the US Food and Drug Administration (FDA) (11). While the need for ISR has been accepted, there is discussion on ISR acceptance criterion and the sample size recommendations, specifically in terms of the ability to discriminate variability in the analysis of study samples. There have been reports that have assessed the above issues. Rocci *et al.* (2007) and Hoffman (2009)

investigated appropriate design and sample size requirements for ISR testing (12,13). Recently, Thway *et al.* (2011) investigated the effect of imprecision on ISR analysis for macromolecules (14).

This study attempts to address the relevance of the ISR acceptance criterion and sample size recommendations of the 2008 ISR Workshop for small molecules based on *in silico* analysis. In this paper, unless specified otherwise, the terms "incurred sample (IS) imprecision" and "IS coefficient of variation (CV)" refer to imprecision in analysis of ISs. IS imprecision may or may not reflect in-study assay imprecision which is routinely monitored by the performance of quality control (QC) samples. "ISR sample size" refers to the number of samples used for ISR, and "ISR acceptance criterion" refers to ISR acceptance criterion recommended at the 2008 ISR Workshop. Since IS imprecision is not routinely monitored in BE, PK, or TK studies, we simulated ISR studies to assess the sensitivity of the current ISR acceptance criterion to detect random errors as a function of IS imprecision and ISR sample size. Also, since systematic bias can be introduced in bioanalytical conduct for reasons, including instability, interconversion, or assay-related issues, we simulated ISR studies to understand the impact of systematic bias on the probability of ISR studies passing the ISR acceptance criterion. Finally, examples of ISR in *in vivo* BE studies are discussed to understand the relevance of the *in silico* results. The current analysis aims to provide a framework for the selection of sample size based on IS imprecision.

## EXPERIMENTAL

### Simulation of ISR Studies: IS Imprecision, Acceptance Criteria, and Sample Size

Simulations were done using a procedure similar to that reported by Thway *et al.* (14). Simulations were conducted using SAS, version 9.3 (SAS Institute, Inc., Cary, NC). Each simulated study includes a specified number of ISR samples (*i.e.*, sample size). Each ISR sample was randomly assigned a target value between 20 and 200 units, which represents the "true" sample value, and a fixed IS imprecision. A normal distribution was generated for each sample based on mean equal to its true value and a standard deviation derived from the IS imprecision. Each ISR sample included observed original and reanalyzed concentrations (*i.e.*, sample pairs). The observed original and reanalyzed concentrations for a sample were randomly selected from the sample's normal distribution. The percent differences between the observed original and reanalyzed concentrations normalized to their average values (from hereon referred to as "percent difference" or "% difference") were derived for each sample using the following formula:

$$\% \text{ Difference} = (\text{Reanalyzed–Original})/0.5$$
$$\times (\text{Original} + \text{Reanalyzed}) \times 100$$

Each simulated ISR study has a sample size between 20 and 300 paired samples (20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 250, 300) and IS CV between 5 and 20% (5, 10, 11, 12,

13, 14, 14.5, 15, and 20% CV). Ten thousand studies were simulated at each IS CV and ISR sample size combination. As an example, for an IS CV of 5% and ISR sample size of 20 sample pairs, 10,000 ISR studies were simulated. A total of 1,080,000 ISR studies were simulated (12 sample sizes×9 IS imprecision×10,000 studies). For each study, an evaluation was performed on whether the ISR study meets the 2008 ISR Workshop's ISR acceptance criterion recommendation for small molecules (*i.e.*, % difference for at least 67% of the ISR sample pairs is within ±20%). The percentage of ISR studies that met the ISR acceptance criterion (*i.e.*, ISR pass rate) was assessed at each IS imprecision and ISR sample size combination. While IS CVs of 5, 10, 15, and 20% were of primary interest, IS CVs of 11 to 14 and 14.5% were added to understand the influence of ISR sample size on the ISR pass rate.

### ISR Study Simulation

This simulation is designed to closely mimic an *in vivo* ISR study. The simulation study design is similar to that in the earlier section, except that ISR samples are randomly assigned imprecision from specified IS imprecision ranges. For each ISR sample, a target value (*i.e.*, true concentration) was randomly assigned between 20 and 200 units, and an IS CV was randomly assigned from a specified IS imprecision range. The observed original and reanalyzed values for each sample were then derived from the true value and the imprecision as detailed earlier. Extrapolating the imprecision limit of 15% CV for bioanalytical measurement, as described in the current FDA guidance (15), to IS CV, four specific acceptable IS CV ranges (1–15, 7–12, 10–15, and 12–15% CV) were selected. The 1–15% IS CV represents the total acceptance range, 7–12% IS CV represents intermediate acceptable range, whereas 10–15 and 12–15% IS CV represent the higher end of the acceptance range. At each IS CV range, ISR studies were simulated with ISR sample sizes between 20 and 300 paired samples (20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 250, 300). The IS CV ranges of 7–12 and 10–15% were selected based on the results of the earlier simulations with fixed IS CVs to better understand the influence of sample size on ISR pass rates. Ten thousand ISR studies were simulated at each IS imprecision range and sample size combination. A total of 480,000 ISR studies were simulated, with each study having a specified imprecision range and a specified sample size (12 sample sizes×4 imprecision ranges×10,000 studies). The ISR pass rate was estimated as detailed earlier.

In addition, since the acceptable precision at the lower limit of quantitation (LLOQ) as defined by the FDA guidance (15) is 20% CV, the simulations were performed as detailed above, except an IS CV range of 1–20% was used at the LLOQ range, and the assay range was set at 2 to 200 units which is representative of an assay range used in routine bioanalysis. The FDA guidance (15) recommends placement of low quality control (QC) level at 3× LLOQ. Therefore, for these simulations, the LLOQ range was set from 2 units to less than 6 units (*i.e.*, less than 3× LLOQ). In addition to setting the IS imprecision range to 1–20% CV at a LLOQ range (*i.e.*, 2–6 units), the IS CV ranges described above (*i.e.*, 1–14, 1–15, 7–12, 10–14, 10–15, and 12–15% CV) were used

for the rest of the assay range (*i.e.*, 6–200 units) in the simulation studies. A total of 720,000 ISR studies were simulated, with each study having a specified IS imprecision range and a specified sample size (12 sample sizes×6 imprecision ranges×10,000 studies).

## ISR Study Simulation: Systematic Bias

This simulation is designed to analyze the effect of bias on ISR pass rate of ISR studies. The bias in IS could stem from several factors, including instability, interconversion, and matrix effect. Initially, a constant bias was assumed in in-study measurements to understand its impact on ISR pass rate as a function of IS imprecision and sample size. Each ISR sample was analyzed twice: original and reanalyzed. Specifically, a sample was randomly assigned a true value between 20 and 200 units and a specified IS imprecision, and a concentration was randomly generated from the normal distribution based on the sample's true value and imprecision. This is the observed original concentration for the sample. For the same sample, the reanalyzed true value equal to the original true value plus a specified bias was assigned. From the normal distribution based on reanalyzed true value and fixed IS imprecision, the observed reanalyzed concentration was randomly generated. The % difference between the observed original and reanalyzed concentrations was estimated for each ISR sample. Each simulated study was assigned a specified bias (5, 7, 10, and 12%), sample size (20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 250, 300), and IS CV (5, 10, and 13% CV). Ten thousand studies were simulated at each bias, IS imprecision, and ISR sample size combination. A total of 1,440,000 ISR studies were simulated, with each study having a specified bias, IS CV, and sample size (12 sample sizes×3 imprecision×4 bias×10,000 studies).

Since the bias is not constant during bioanalytical conduct, additional simulations were performed to analyze the effect of variable bias on ISR pass rate as a function of an IS imprecision range and sample size. Similar to earlier ISR simulation studies, an assay range was set at 2–200 units, and systematic bias ranges of 0–10, 5–10, 10–15, and 0–20% were selected. Also, the IS CV at the LLOQ range (*i.e.*, >2<6 units) was set at 1–20% CV for all simulations, while IS CV ranges for the rest of the assay range (6–200 units) was set at 1–5, 1–10, 7–12, and 1–15% CV. The ISR sample sizes of 20, 40, 60, 80, 100, 125, 150, 160, 180, 200, 250, and 300 sample pairs were selected. A total of 1,920,000 ISR studies were simulated, with each study having a specified bias range, IS CV range, and sample size (12 sample sizes×4 imprecision range×4 bias range×10,000 studies).

## Analysis of ISR in *In Vivo* BE Studies

*In vivo* BE studies, that included ISR and contained a subset of samples that were analyzed at least three times (*i.e.*, three valid measurements: original and at least two reanalysis), were selected from submitted data. IS imprecision was estimated by calculating the %CV for each of the study samples analyzed at least three times. For ISR samples, the percent differences between the original and reanalyzed values were determined as stated earlier. The mean, standard deviation (SD), median, and range of the % difference of ISR

samples were estimated. Also, to interpret the ISR data, 1) probability plot was constructed for each ISR study, to determine the distribution of the % difference data, as detailed by Lytle *et al.* (2009) using Microsoft Excel (Microsoft Office Professional Plus 2010), and 2) Bland-Altman plots incorporating tolerance intervals (TLs) were generated as discussed by Lytle *et al.* (2009) (16). Bland-Altman analysis involved estimation of 95% confidence limits (CLs) and TLs (16). For the 95% CLs estimation, normal distribution value (for a mean of 0 and SD of 1: *z* score) of 0.967 was used. For TLs estimation, tolerance factors (K) for a normal distribution were obtained from Lytle *et al.* (2009) based on the number of ISR samples; 66.7% proportion of % difference data are contained within the TLs and 95% confidence. Also, assay imprecision was determined by estimating the inter-run CVs of low, medium, and high in-study QC data in the BE studies.

## RESULTS

### Simulation of ISR Studies: Effect of Imprecision

To understand the extent of the effect of IS imprecision on ISR acceptance criterion, ISR studies were simulated at a specified imprecision and sample size, and the number of studies passing the ISR acceptance criterion (*i.e.*, pass rate) per the 2008 ISR workshop was determined. Here, IS imprecision or IS CV denotes the imprecision in ISs, as opposed to in-study assay imprecision (usually based on QC data).

Figure 1 represents IS imprecision plotted as a function of ISR sample size and pass rate. When IS imprecision is ≤10%, the number of studies passing the ISR acceptance criterion is almost independent of sample sizes evaluated (Fig. 1) and requires a maximum of 40 samples to attain a 99% pass rate. When IS imprecision increases from 10 to 14% CV, the ISR pass rate decreases, and the effect of sample size on pass rate becomes more pronounced. Specifically, when IS imprecision is 11% CV, approximately 60 samples are sufficient for 99% of ISR studies to meet the ISR acceptance criterion. At an IS imprecision of 12% CV, at least 120 samples are required to attain a pass rate of 99%. When the IS imprecision increases to 13% CV, a sample size of 300 is required to produce a pass rate of 97%. In contrast, at IS imprecision ≥15% CV, the pass rate of ISR studies drops to below 50% with 20 samples and decreases with increase in sample size.

Since assay or IS imprecision during sample analysis is not constant, the effect of IS imprecision within a specified range on ISR pass rate was investigated. ISR studies were simulated at specified IS imprecision ranges (1–15, 7–12, 10–15, and 12–15% CV) and sample sizes, and the number of studies passing the ISR acceptance criteria were determined. As mentioned earlier, IS CV ranges of 1–15 and 12–15% CV, represents acceptable and relatively higher but acceptable CV ranges, respectively. The 7–12 and 10–15% IS CV ranges were selected based on Fig. 1, to understand the effect of other IS CV ranges on ISR pass rates and to identify the influence of sample sizes at these IS CV ranges. When IS imprecision is 1–15% CV, 90% of ISR studies pass with a sample size of 40, with no significant increase in pass rate at
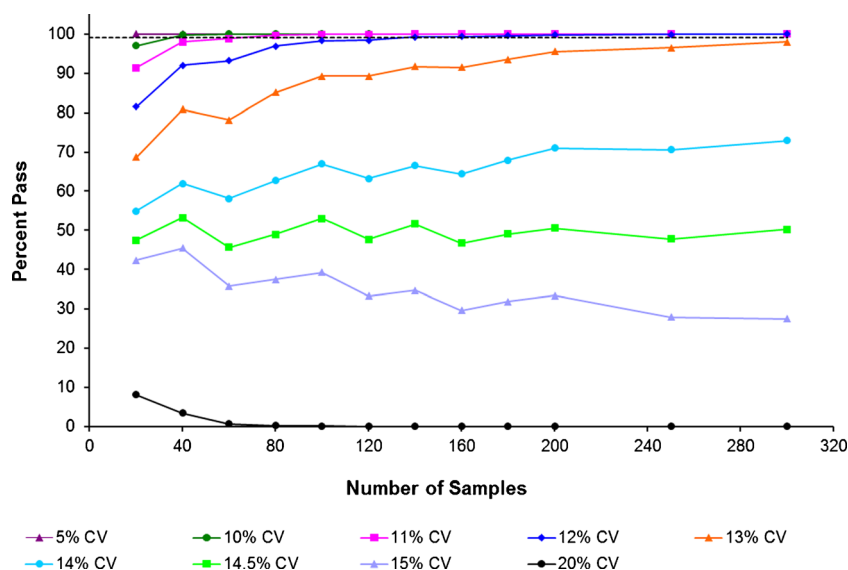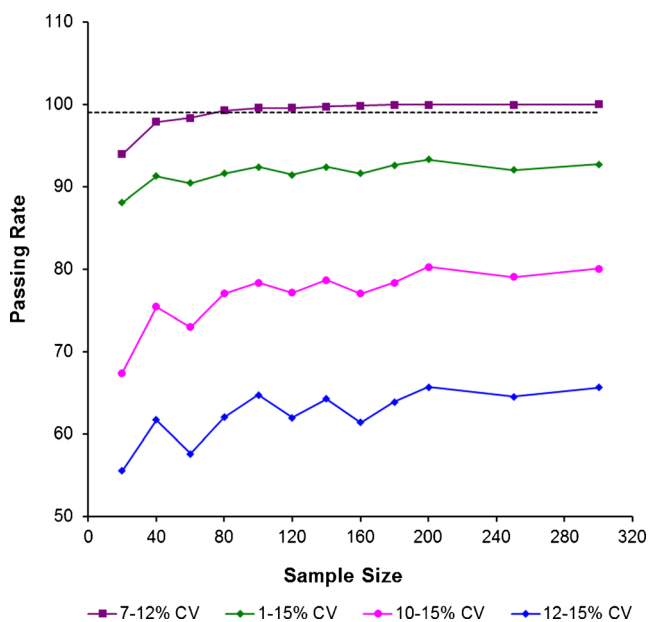
**Fig. 1.** Effect of incurred sample (IS) imprecision and ISR sample sizes on pass rate of ISR studies. For an ISR study to pass, at least two thirds of reanalyzed values should agree within 20% of their original values when normalized to their average values. The assay range for the simulation studies was set at 2–200 units. The probability of meeting the ISR acceptance criterion was examined at 5, 10, 11, 12, 13, 14, 14.5, 15, and 20% IS imprecision (%CV). The *dashed line* represents 99% of the ISR studies passing the ISR acceptance criterion

higher sample sizes evaluated (Fig. 2). At an IS imprecision range of 7–12% CV, ISR pass rate increases with sample size and a 99% pass rate is attained at a sample size of 80. Further increases in sample sizes evaluated produce negligible effects on pass rate of ISR studies. When the IS imprecision range is between 10% CV and 15% CV, a large increase in sample size (*i.e.*, 300) is required for 80% ISR pass rate. When an IS imprecision range is between 12 and 15% CV, the ISR pass rate is less than 70% regardless of the evaluated sample sizes. These results were mostly unchanged when an IS CV range of 1–20% CV was assigned at the LLOQ range, and the rest of the assay range was assigned an IS CV range of 1–15, 10–15, 7–12, or 12–15% (Fig. 3).

### Simulation of ISR Studies: Effect of Systematic Bias

To determine the impact of systematic bias during ISR, ISR studies were simulated at fixed systematic bias and IS imprecision and at specified bias and IS imprecision ranges. At an IS imprecision of 5% CV, a positive bias of up to 12% did not affect the high (≥99%) ISR pass rate (data not shown). Similar results were observed for a systematic bias range of 0–10% and an IS CV range of 1–5 or 1–10% (Fig. 4). At a positive bias of 12% and an IS imprecision of 10% CV, or a positive bias range of 10–15% and an IS CV range of 1–10%, at least 300 samples are required to attain a 95% ISR pass rate (Figs. 4 and 5a). Also, positive bias ranges of 0–10 or 5–10% at an IS CV range of 7–12% require at least 250 samples to attain a 98% ISR pass rate. When IS CV increases to 13% and the bias is 5%, 300 samples attain only a 90% ISR pass rate (Fig. 5b). Similar results were observed at an IS CV range of 1–15% and a bias range of 0–10% (Fig. 4). ISR pass rates are lower than 85% at bias ranges of 10–15 or 0–

20% and IS CV ranges of 1–15, 1–10, or 7–12% (data not shown). Similar result was obtained at an IS CV range of 1–15% and a bias range of 5–10% (data not shown).



**Fig. 2.** Effect of incurred sample (IS) imprecision ranges and ISR sample sizes on pass rate of ISR studies. For an ISR study to pass, at least two thirds of reanalyzed values should agree within 20% of their original values when normalized to their average values. The assay range for the simulation studies was set at 20–200 units. The probability of meeting the ISR acceptance criterion was examined at IS imprecision (% CV) ranges of 7–12, 1–15, 10–15, and 12–15%. The *dashed line* represents 99% of the ISR studies passing the ISR acceptance criterion
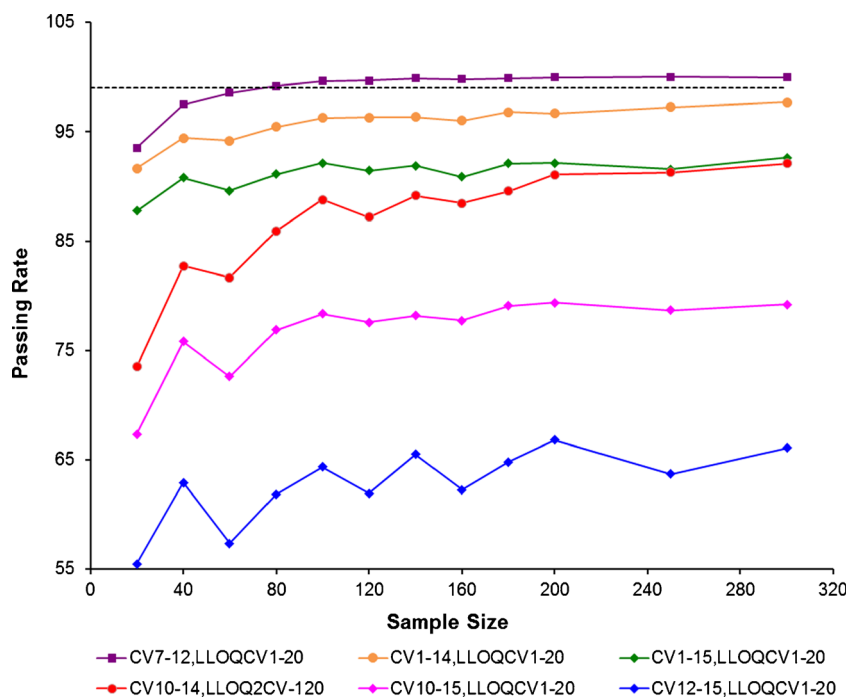
**Fig. 3.** Effect of incurred sample (IS) imprecision ranges and ISR sample sizes on pass rate of ISR studies. For an ISR study to pass, at least two thirds of reanalyzed values should agree within 20% of their original values when normalized to their average values. The assay range for the simulation studies was set at 2–200 units. The probability of meeting the ISR acceptance criterion was examined at IS imprecision (% CV) range of 1–20% at the LLOQ range (2–6 units) and IS imprecision (% CV) ranges of 7–12, 1–14, 1–15, 10–14, 10–15, and 12–15% for the rest of the assay range. The *dashed line* represents 99% of the ISR studies passing the ISR acceptance criterion

**Analysis of ISR in *In Vivo* BE Studies**

The three real *in vivo* BE studies with ISR (ISR #1, #2, and #3) were selected because, the studies also included 3 to 21 study samples that were each analyzed at least three times (all valid measurements). This facilitated estimation of IS imprecision. Also, the IS imprecision in the 3 studies corresponded to different IS CV ranges.

The IS CVs in BE studies for ISR #1 and ISR #2 were between 1–3 and 5–7%, respectively. The assay imprecision ranges in the BE studies for ISR #1 and #2 were 2–4% CV and 8–9% CV, respectively. ISR #1 and #2 passed the ISR acceptance criteria. In ISR #1, the % differences for all sample pairs were ≤20%, whereas in ISR #2, 90% of sample pairs had % differences ≤20% (Table I). The probability plots in Fig. 6a, b indicate that the % difference data in ISR #1 and ISR #2 approximates a normal distribution as the data points reasonably fit the 45° line. Also, the 95% CLs and TLs about the means using the Bland-Altman approach were estimated to assess systemic bias and random errors, respectively (16). The analysis indicates a low systematic negative bias in ISR #1 as the 95% CLs did not include 0 (Table II). However, the bias does not affect ISR acceptance as the % differences of all samples in ISR #1 are within 20%.

ISR #3 was conducted in six batches (A through E) during the course of the BE study. Each batch involved 10–32 ISR samples. While the % difference data for the majority of the ISR samples (80–93%) were within 20% in batches A, C, and D, only 19, 45, and 66% of ISR samples in batches B, E,

and F, respectively, had % difference values within 20% (Table III). The ISR samples in batch B were reanalyzed twice. The mean imprecision in the IS samples in batch B was found to be 22% CV (range 5–40% CV). Also, the overall assay imprecision in the BE study was relatively high (11–24% CV). Combining the data from all six batches, only 66.4% of the 143 ISR samples had % differences within 20%. The Bland-Altman plot of the combined data indicates a positive bias in ISR #3 (as the 95% CLs are greater than 0, Table II), with the reanalyzed concentrations for the majority (71%) of the ISR samples were greater than that of their original concentrations (data not shown). The probability plot for ISR #3 indicates that the ISR data are not normally distributed (Fig. 6c).

**DISCUSSION**

ISR is routinely performed to confirm the reliability of measurements during study sample analyses. The ISR acceptance criterion recommended for small molecules at the 2008 ISR workshop is that the % differences (*i.e.*, percent difference between original and reanalyzed concentrations when normalized to their means) for at least two thirds of ISR samples are within ±20% (6). Currently, this ISR criterion is routinely used in bioanalysis and has been adopted by regulatory agencies (9–11). There have been studies in the past that have compared the current acceptance criterion with alternate acceptance criteria (12–14). The current study attempts to assess the sensitivity of the ISR
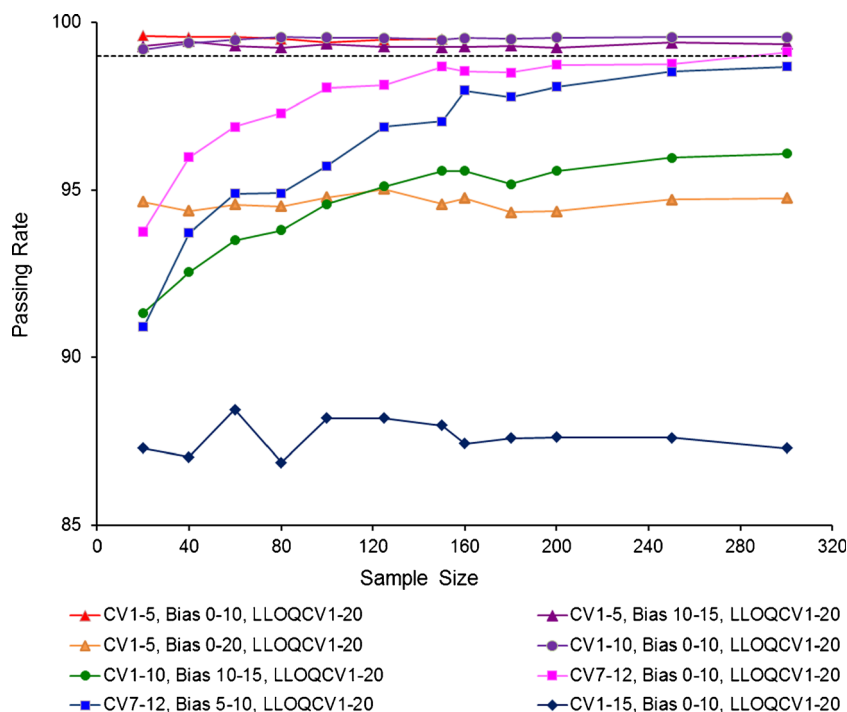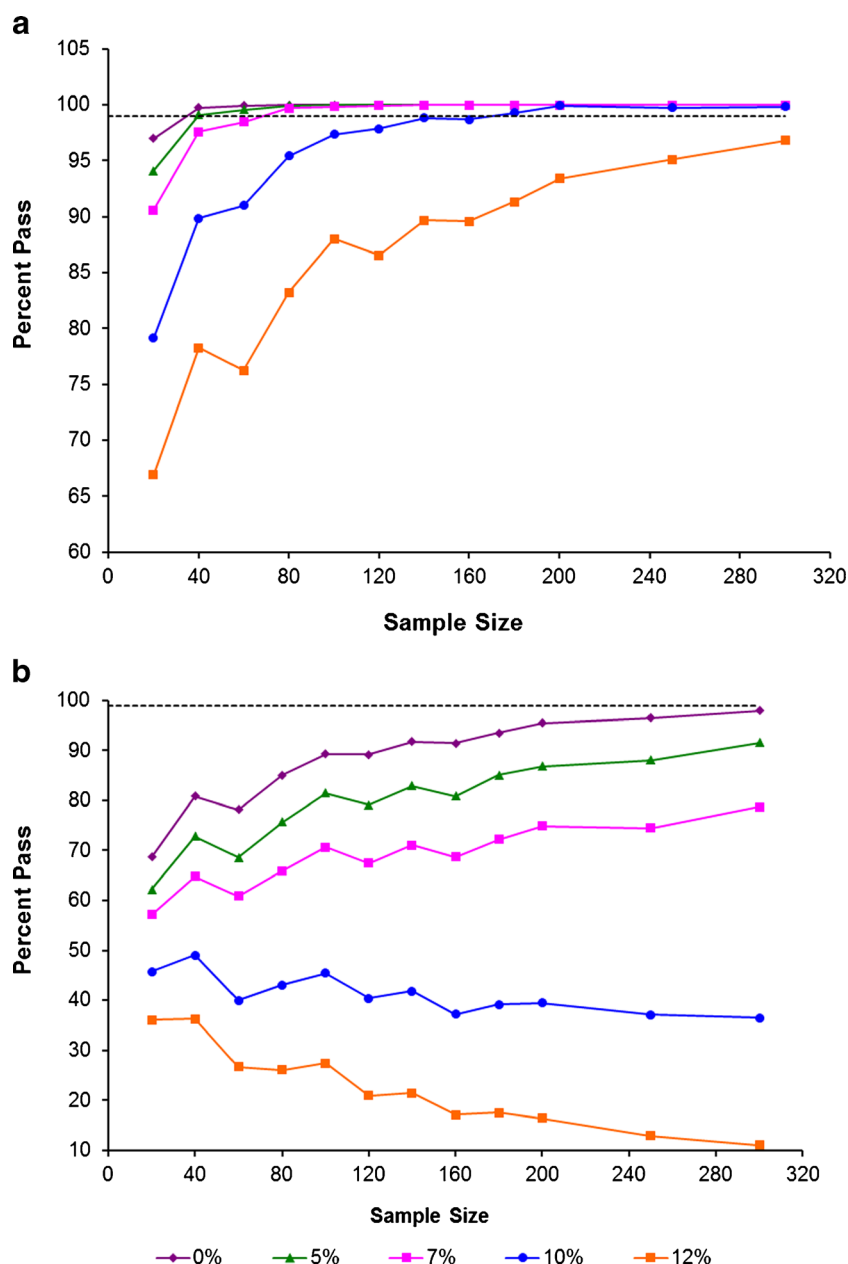
**Fig. 4.** Effect of systematic bias ranges, incurred sample (IS) imprecision ranges, and ISR sample size on pass rate of ISR studies. For an ISR study to pass, at least two thirds of reanalyzed values should agree within 20% of their original values when normalized to their average values. The assay range for the simulation studies was set at 2–200 units. In all the simulation studies, the IS imprecision at the LLOQ range (2–6 units) was set at 1–20% CV. The IS imprecision (% CV) ranges at the rest of the assay range were varied as 1–5, 1–10, 7–12, and 1–15% CV. The effect of systematic bias ranges of 0–10, 5–10, 10–15, and 0–20% CV and the above IS imprecision ranges on the probability of meeting the ISR acceptance criterion were examined. The *dashed line* represents 99% of the ISR studies passing the ISR acceptance criterion. Simulation results are not shown in this figure at bias ranges of 10–15 or 0–20% and IS CV ranges of 1–15, 1–10, or 7–12% and at a bias range of 5–10% and IS CV range of 1–15%, as the ISR pass rates were <85%

acceptance criterion, specifically its ability to discriminate between low and high IS imprecision during study sample analysis and address the effect of sample size on ISR study acceptance. Figures 1, 2, and 3 demonstrate that when IS imprecision is ≤10% CV, there is a low probability of ISR failure. However, when fixed IS imprecision is at 11, 12, or 13% CV or when IS CV range is 7–12 or 1–14%, the ISR pass rate decreases and becomes a function of sample size; for example, when fixed IS CV increases from 11 to 13%, the pass rate decreases from approximately 90 to 70% at a sample size of 20 and requires a fivefold increase (60 to 300) in a sample size to maintain a high pass rate (Fig. 1).

The results of the above simulation studies seem to be reasonable at IS CV ≤12%. In fact, the simulation results at IS CVs 10 to 12% are supported by literature (13). Also, the results are supported by IS results in the *in vivo* BE studies presented in Table I. Specifically, at an IS CV range of 1–4%, all ISR #1 samples have a % difference ≤20%. At IS imprecision range of 5–7% CV, 88% of ISR samples have a % differences ≤20%. Although the reliability of the estimated IS CVs in the BE studies are debatable due to the small sample pool, the IS CVs in the BE studies are supported by the in-study assay CVs. Further, the probability plots indicate normal distribution of the ISR data (Fig. 6a, b).

The current simulation studies and other *in silico* studies (13,16) indicate that at IS CVs of ≤12% CV or at IS CV range of 7–12% CV, ISR sample sizes of 40–160 are sufficient to attain high ISR pass rates. This approach may be sufficient to capture random errors (*e.g.*, errors in sample processing such as contamination and sample switching or sample mixing). However, the usefulness of such ISR sample sizes to readily identify isolated errors (*e.g.*, processing errors for selective batches, subjects, or analytical batches) can be limiting, especially in BE or PK studies with large sample sizes (*i.e.*, >2500 samples). Isolated errors in large studies are more likely, as the studies may involve large analytical batch sizes, a longer duration to complete analysis, multiple analysts, and/or multiple instruments. In addition, the isolated errors in study samples may not be reflected in the calibration standard and QC samples (*i.e.*, spiked samples), especially when the ratio of study samples to spiked samples are high. In fact, for the majority of ISR failures reported, retrospective review of spiked samples rarely indicates problems in spiked samples (3,4). In such cases, an increase in ISR sample size will be useful.

Also, the current simulations show that, at an IS CV of 13% or a broad IS CV range (*e.g.*, 1–14% CV), the sample size requirement dramatically increases to 300 to attain an

**Fig. 5.** Effect systematic bias and ISR sample size on pass rate of ISR studies: **a** at incurred sample (IS) imprecision (% CV) of 10% and **b** at IS CV of 13%. For an ISR study to pass, at least two thirds of reanalyzed values should agree within 20% of their original values when normalized to their average values. The assay range for the simulation studies was set at 20–200 units. The probability of meeting the ISR acceptance criterion was examined at IS imprecision of 5, 10, and 13% CV and bias of 0, 5, 7, 10, and 12%. The *dashed line* represents 99% of the ISR studies passing the ISR acceptance criterion. At 5% IS CV, a positive bias of up to 12% has no significant effect on passing rate (>99%); therefore, the data are not included

ISR pass rate of ≥98%. Lytle *et al.* (16), using TLs that include 66.7% of experimental data, showed that, at 14% CV, the sample sizes required to meet the TLs of ±20% at 80, 90, and 95% confidence were dramatically higher compared to those at 12% CV. For example, when imprecision increased from 12 to 14% CV, the sample size required to meet TLs of ±20% at 80% confidence increased from 19 to 233 ISR samples and at 95% confidence

increased from 49 to 785 ISR samples. These results are similar to those presented in Figs. 2 and 3.

In contrast, the *in silico* results at high IS imprecision (≥15% CV) or imprecision ranges (*e.g.*, 12–15% CV) are not entirely intuitive, specifically, the lack of effect or decrease in ISR pass rate with increasing sample sizes up to 300. Nonetheless, the results are supported by literature. Hoffman showed similar pass rates and higher probability of

**Table I.** ISR and Assay Data for ISR #1 and ISR #2 in BE Studies

| Parameters | ISR #1 | ISR #2 |
|---|---|---|
| Incurred sample CV[a](%) | | |
| Mean ($n$[b]) | 2.9% (8) | 6.0% (3) |
| Range (%) | 1.1–4.3 | 4.6–6.8 |
| ISR analysis[c]: % difference | | |
| Mean | −1.8 | 0.1 |
| SD | 5.7 | 12.1 |
| Median | −2.1 | −0.6 |
| Min | −15.2 | −24.2 |
| Max | 17.9 | 31.1 |
| $N$[d] | 114 | 80 |
| %>20%[e] | 0.0% | 12.5% |
| In-study assay CV range ($n$)[f] | 2.3–4.4% (20–30) | 7.8–8.5% (36–38) |

*ISR* incurred sample reanalysis, *CV* coefficient of variation, *SD* standard deviation

[a] Refers to % CV of study samples that were analyzed at least three times. These samples were independent of ISR samples

[b] Refers to number of study samples that were each analyzed at least three times to estimate incurred sample CV

[c] Refers to analysis of ISR samples. Analyzed twice, *i.e.*, original and ISR

[d] Refers to number of ISR samples

[e] Refers to the percentage of ISR samples with % differences >±20%

[f] Refers to %CV range of quality control (QC) samples in the BE studies. "n" refers to the range of number of QCs per level

ISR failure with an increase in sample size at 15.5, 17.5, and 20% CV (13). Also, the *in silico* results presented in the current paper are in line with those reported for macromolecules (14).

However, it is interesting to note that in the example of BE study for ISR #3 (Table III), the ISR data (combining results of all batches) almost passed (66.4% of samples have % differences ≤20%) the ISR acceptance criterion, although the mean IS CV in batch B (22%) and the in-study assay CVs (23, 11, and 16% at low, medium, and high QC levels, respectively) were high, and the ISR sample size was only 148. The probability plot shows that the ISR #3 data is not normally distributed, and the Bland-Altman data shows a positive systemic bias. In retrospect, had the samples in other batches been reanalyzed more than once, the correlation or the lack of correlation between IS CVs and assay CVs within batches could have been determined. This information would have been useful, considering the within-batch QC imprecision was consistently high for all batches (data not shown). It should be noted that for majority (>50%) of ISR samples in batch B, no two assay values (*i.e.*, original *vs* repeat 1, original *vs* repeat 2, or repeat 1 *vs* 2) were reproducible (data not shown). In such cases, even when the overall ISR results (*i.e.*, after combining results from multiple ISR batches) happen to meet the ISR acceptance criterion, the accuracy of data for the subjects, whose ISR samples consistently demonstrated lack of IS reproducibility, are questionable,

Although, the IS CVs and the assay CVs (based on QCs) are similar in the current ISR examples in BE studies, there are instances when IS imprecision do not mimic assay imprecision in BE or PK studies (3,4,17). This is the reason why imprecision in the current paper differentiates between ISR imprecision and assay imprecision. This does not mean that assay imprecision is not important; however, one should
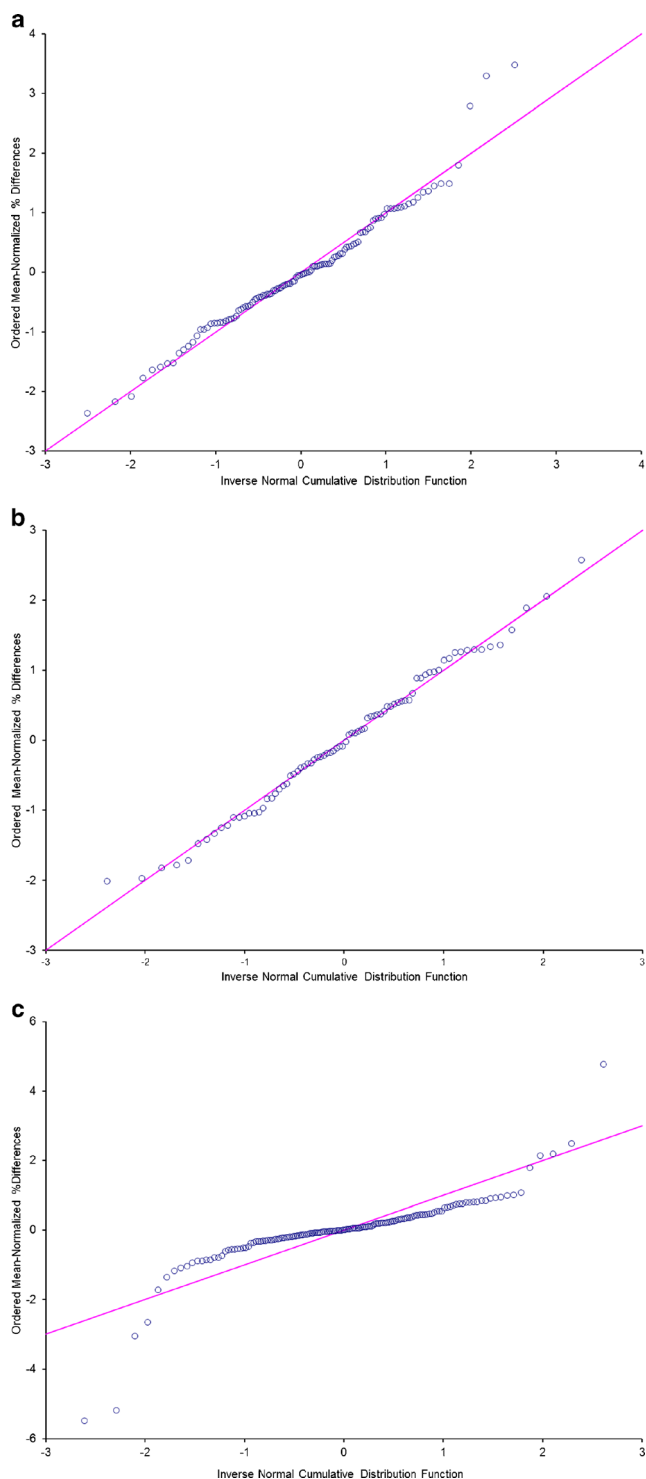


**Fig. 6.** Probability plots for **a** ISR #1, **b** ISR #2, and **c** ISR #3. The *x*-axis is the inverse normal cumulative distribution function, and *y*-axis is a sorted list of % differences (percent difference of original and reanalyzed values when normalized to their average values) normalized by its mean and standard deviation (*i.e.*, ordered mean-normalized % differences). The solid 45° line (*pink*) represents theoretical normal distribution with a mean of 0 and a standard deviation of 1

be aware that assay imprecision sometimes may not reflect imprecision in incurred samples.

**Table II.** Results of % Difference Data Analyzed Using the Bland-Altman Approach for Studies ISR #1, ISR #2, and ISR #3

| ISR # | Range of average values of sample pairs[a] | Mean % difference | Standard deviation of % difference | 95% confidence limits of % difference[b] | Tolerance limits of % difference[c] | Acceptance range (%) | N | Sample pairs outside acceptance range (%) | Sample pairs outside the tolerance limits (%) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.9–35.6 | −1.8 | 5.7 | Lower=−2.8 Upper=−0.8 | Lower=−7.3 Upper=3.7 | Min=−20 Max=20 | 114 | 0 | 26.3 |
| 2 | 375–3850 | 0.1 | 12.1 | Lower=−2.6 Upper=2.7 | Lower=−13.5 Upper=13.6 | Min=−20 Max=20 | 80 | 12.5 | 26.3 |
| 3 | 0.2–17.5 | 7.2 | 28.5 | Lower=2.6 Upper=11.9 | Lower=−20.4 Upper=34.8 | Min=−20 Max=20 | 143 | 33.6 | 16.1 |

[a] Refers to minimum and maximum averages of original and reanalyzed concentrations of ISR samples
[b] Estimated based mean and standard deviation (SD) of % differences and the number of samples used in the ISR study (N) and *z* score of 0.967
[c] Estimated using mean and standard deviation of % differences in the ISR study and tolerance factor (k) for a normal distribution with a proportion of 66.7 and 95% confidence limits

In addition to imprecision, the effect of systematic bias on pass rates were simulated since systematic bias can potentially be introduced in bioanalysis for reasons, including instability, interconversion, and conduct-related issues. The results of the simulation studies indicate a fixed systematic bias of <10% or a bias range of 0–10% do not significantly impact ISR pass rates when IS CV is fixed at ≤10% or the IS CV range is set at 1–10% (Figs. 4 and 5). However, when bias range and/or IS CV range increases beyond the levels stated above, ISR pass rate drops and becomes a function of sample size (Fig. 4). For example, when bias range increases to 10–15% at an IS CV range of 1–10 or 7–12%, ISR pass rate decreases and requires at least 300 samples to attain a pass rate of 95%. However, large IS imprecision range relative to systematic bias range or *vice versa* (*e.g.*, 0–10% bias at 1–15% IS CV or 0–20% bias at 1–5% IS CV) is likely to adversely impact ISR pass rates, independent of sample size (Fig. 4). The results of *in vivo* ISR studies presented (Table II and Fig. 6) seem to agree with the *in silico* results. In ISR #1, the negative bias at a low IS CV range (1–4% CV) did not affect

the acceptance of the ISR study. In contrast, in ISR #3, the high IS CV range (5–40% CV) and a relatively strong positive bias affected the acceptance of the ISR study.

**CONCLUSIONS**

Although limitations exist (13), we believe that the ISR acceptance criterion recommended at the 2008 ISR Workshop is relevant. The analysis shows that the ISR acceptance approach is able to discern variability in the analysis of ISs, although at high IS imprecision (≥15%), the results are not self-evident. The results also provide an understanding of sample size requirements as a function of ISR imprecision. Currently, sample size recommendations differ, as highlighted in literature and regulatory guidances (9,11–13). The current *in silico* analysis results predict that, to attain a high ISR pass rate (≥99%), 40–50 samples are sufficient at low IS imprecision range (1–10% CV) and approximately 160 samples at intermediate IS imprecision range (*e.g.*, 12% CV or 7–12% CV). In addition to random errors, the *in silico* results

**Table III.** ISR and Assay Data for ISR #3 in a BE Study

| ISR #3 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Incurred sample CV[a] (%) | | | | | | |
| Mean (n[b]) | – | 22.0% (21) | – | – | – | – |
| Range | | 4.7–39.5% | | | | |
| ISR analysis[c]: % difference | | | | | | |
| Mean | 3.1 | 28.1 | 3.1 | 4.1 | 5.1 | 15.9 |
| SD | 10.7 | 14.8 | 38.8 | 10.4 | 23.5 | 33.3 |
| Median | 1.8 | 34.0 | 2.6 | 5.1 | −2.0 | 12.7 |
| Min | −13.0 | −5.7 | −109 | −27.3 | −40.0 | −25.2 |
| Max | 25.5 | 45.3 | 97.8 | 20.9 | 39.9 | 183 |
| N[d] | 10 | 21 | 30 | 29 | 21 | 32 |
| %>20%[e] | 10.0% | 81.0% | 20.0% | 6.9% | 52.4% | 34.4 |
| In-study assay CV range (n)[f] | 11.2–24.0% (151) | | | | | |

*ISR* incurred sample reanalysis, *CV* coefficient of variation, *SD* standard deviation
[a] Refers to %CV of study samples that were each analyzed at least three times. Includes ISR samples
[b] Refers to number of study samples that were each analyzed at least three times to estimate incurred sample CV
[b] Refers to analysis of ISR samples. Analyzed twice, *i.e.*, original and ISR
[d] Refers to number of ISR samples
[e] Refers to the percentage of ISR samples with % differences >±20%
[f] Refers to %CV range of quality control (QC) samples in the BE study. "n" refers to number of QCs per level

indicate that systematic bias ≥10 and <15% is likely to affect ISR pass rate at intermediate IS imprecision ranges and requires ≥250 samples to attain ≥95% ISR pass rate. Although IS imprecision ranges have not been extensively studied in *in vivo* BE or PK studies, the imprecision at the low and intermediate ranges stated above should cover the expected IS imprecision ranges in bioanalysis of small molecules in such studies, especially with the use of chromatographic systems coupled to tandem mass-spectrometric detectors (LC/MS/MS). Therefore, the sample sizes of 40–250 samples approximate 5–10% of total study samples in routine BE studies that involve a two-way cross-over design, with 25–50 subjects and 20–25 time points (which translates to 1000–2500 samples). Although the use of the 40–250 samples can be extrapolated to larger studies (>2500 samples) based on the *in silico* analysis, the utility of such sample sizes to identify isolated errors during study sample analysis (that affects a portion of the study data) may be limiting, as the propensity for such errors increase in large studies due to a longer duration to complete analysis, large batch sizes, and multiple analysts /instruments. The authors believe that a thorough investigation of ISR failures is critical to understanding the nature of the errors (what drug types, matrices, and/or assay conditions) and how the errors can be tightly controlled during study sample analysis to provide accurate and reliable results. Various tools to investigate ISR data have been reported in literature (12,16,18), and some of them have been used in this article to understand the *in vivo* ISR data.

## ACKNOWLEDGMENTS

**Disclaimer** The views expressed in this article are those of the authors and do not reflect official policy at the FDA. No official endorsement by the FDA is intended or should be inferred.

## REFERENCES

1. Brockman AH, Hatsis P, Paton M, Wu J-T. Impact of differential recovery in bioanalysis: the example of bortezomib in whole blood. AAPS J. 2007;79:1599–603.
2. Silvestero L, Gheorghe MC, Tarcomnicu I, Savu S, Savu SR, Iordachescu A, *et al*. Development and validation of an HPLC-MS/MS method to determine clopidogrel in human plasma. Use of incurred samples to test back-conversion. J Chromatogr B. 2010;878:3134–42.
3. Côté C, Lahaie M, Latour S, Bergeron M, Dicaire C, Savoie N, *et al*. Impact of methylation of acyl glucuronide metabolites on incurred sample reanalysis evaluation; ramiprilat case study. Bioanalysis. 2011;3(9):951–65.
4. Yadav M, Shrivatsav PS. Incurred sample reanalysis (ISR): a decisive tool in bioanalytical research. Bioanalysis. 2011;3 (9):1007–24.
5. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, *et al*. Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. AAPS J. 2007;9(1):E30–42.
6. Fast DM, Kelley M, Viswanathan CT, O'Shaughnessy J, King SP, Chaudhary A, *et al*. Workshop report and follow-up—AAPS workshop on current topics in GLP bioanalysis: assay reproducibility for incurred samples—implications of Crystal City recommendations. AAPS J. 2009;11(2):238–41.
7. Timmerman P, Luedtke S, van Amsterdam P, Brundy-Kloeppel M, Lausecker B, Fischmann S, *et al*. Incurred sample reproducibility; views and recommendations by the European Bioanalysis Forum. Bioanalysis. 2009;1(6):1049–56.
8. Savoie N, Booth BP, Bradley T, Garofolo F, Hughes NC, King SP, *et al*. The 2nd CVG workshop on recent issues in good laboratory practice bioanalysis. Bioanalysis. 2009;1(1):19–30.
9. Guideline on bioanalytical method validation. Committee for Medicinal Products for Human Use. European Medicine Agency. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/WC500109686.pdf (2011). Accessed September 2013
10. Notice to guidance document: conduct and analysis of comparative bioavailability studies. Health Canada. http://www.hc-sc.gc.ca/dhp-mps/alt_formats/pdf/prodpharma/applic-demande/guide-ld/bio/gd_cbs_ebc_ld-eng.pdf (2012). Accessed September 2013.
11. Draft guidance for industry: bioanalytical method validation. US Department of Health and Human Services, Food and Drug Administration. September 2013 http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM368107.pdf (2013). Accessed November 2013.
12. Rocci Jr ML, Devanarayan V, Haughey DB, Jardieu P. Confirmatory reanalysis of incurred bioanalytical samples. AAPS J. 2007;9(3):E336–43.
13. Hoffman D. Statistical considerations for assessment of bioanalytical incurred sample reproducibility. AAPS J. 2009;11 (3):570–80.
14. Thway TM, Eschenberg M, Calamba D, Macaraeg C, Ma M, DeSilva B. Assessment of incurred sample reanalysis for macromolecules to evaluate bioanalytical method robustness: effects from imprecision. AAPS J. 2011;13(2):291–8.
15. Guidance for industry: bioanalytical method validation. US Department of Health and Human Services, Food and Drug Administration. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070107.pdf (2001). Accessed September 2013.
16. Lytle FE, Julian RK, Tabert AM. Incurred sample reanalysis: enhancing the Bland-Altman approach with tolerance intervals. Bioanalysis. 2009;1(4):705–14.
17. Fu Y, Li W, Smith HT, Tse FLS. An investigation of incurred human urine sample reanalysis failure. Bioanalysis. 2011;3(9):967–72.
18. Voicu V, Gheorghe MC, Sora LD, Sârbu C, Medvedovici A. Incurred sample reanalysis: different evaluation approaches on data obtained for spironolactone and its active metabolite canrenone. Bioanalysis. 2011;3(12):1343–56.