The AAPS Journal 2007; 9 (3) Article 40 (http://www.aapsj.org).

Themed Issue: Bioanalytical Method Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays
Guest Editors - Mario L. Rocci Jr. (MRocci@prevalere.com), Vinod P. Shah (DDr.VPShah@Comcast.net), Mark J. Rose (marose@amgen.com), Jeffrey M. Sailstad (sailstad@aol.com)

Confirmatory Reanalysis of Incurred Bioanalytical Samples

Received: August 28, 2007; Final Revision Received: September 14, 2007; Accepted: September 14, 2007; Published: October 5, 2007

Mario L. Rocci, Jr.,[1] Viswanath Devanarayan,[2] David B. Haughey,[1] and Paula Jardieu[1]

[1]Departments of Bioanalytical Sciences and Product Development, Prevalere Life Sciences Inc., Whitesboro, NY
[2]Global Statistics & Data Management, Abbott Laboratories, Parsippany, NJ

Abstract

Bioanalytical methods used to support the drug development process are validated to ensure that they function in the manner in which they are intended. "Incurred" or study samples can vary in their composition when compared with the standards and quality control samples used to validate the method and analyze these samples. During the 3rd American Association of Pharmaceutical Scientists(AAPS)/Food and Drug Administration(FDA) Bioanalytical Workshop, it was suggested that the reproducibility in the analysis of incurred samples be evaluated in addition to the usual prestudy validation activities performed. This manuscript provides recommendations concerning the number and types of samples that should be analyzed in such an evaluation, as well as the manner in which the resultant data should be analyzed. Suggestions as to follow-up activities and data reporting are also discussed. This approach is at best a beginning and is offered as a platform for future discussion, comments, and revision.

Keywords: Bioanalytical, incurred samples, LC/MS/MS, ELISA, immunoassay, reproducibility

Introduction

Bioanalytical methods used to support toxicology and clinical studies are validated to establish that they function in the manner in which they are intended.[1] Guidelines for the validation of analytical methods have evolved for both small molecules and macromolecules over the last 15 years as discussed by Shah.[2] This evolution has most recently resulted in the publication of a consensus report from the 3rd AAPS/FDA Bioanalytical Workshop dealing with the validation and implementation of bioanalytical methods for both small molecules and macromolecules.[3] Both of these publications are part of this themed issue.

Corresponding Author: Mario L. Rocci, Jr., Prevalere Life Sciences Inc., 8282 Halsey Road, Whitesboro, NY 13492. Tel: (315) 736-3050 (ext) 2313; Fax: (315) 736-2460; E-mail: mrocci@prevalere.com

The purpose of this article is to address one of the summary opinions from this latest workshop suggesting that the reproducibility of bioanalytical methods be demonstrated using incurred samples. In many cases, reproducibility refers to the precision of an assay between laboratories. For the purposes of this manuscript, reproducibility is defined as the consistency of results derived from the analysis of incurred samples on 2 (or more) independent occasions in the same laboratory. This consistency is a function of both assay accuracy (systematic bias) as well as precision (random error). These parameters when taken together are a measure of the total error of the assay.

Much of the data generated during validation studies describe the precision and accuracy of the method, both within and between analytical runs as well as the reproducibility of the method under a variety of circumstances.[1-6] Under most conditions, validation studies employ standards and quality control samples (QCs) that are prepared in vitro in the matrix of interest. This preparation involves the exogenous addition of the analyte(s) being measured to these standards and QCs.

While every attempt is made to formulate standards and QCs to be as similar to the study samples being analyzed as possible, "incurred" or study samples can differ in a variety of ways. These differences are dependent in part on whether the analyte(s) in question are small molecules or macromolecules.

As one example, incurred samples from studies conducted on small molecules can have metabolites present that neither the standards or quality control samples contain. Should a small molecule drug metabolite revert in vitro to its parent, nonreproducible results could occur. If both the metabolite and the parent involved in this conversion process are being measured by the assay, traditional validation studies, if properly conducted, should reveal the problem. When metabolite(s) that are not being measured (ie, that are not present in the standards and QCs used during the validation studies) are involved, this reproducibility problem will not be observed during traditional method validation studies but may be present in incurred samples being analyzed.

In addition to this example, others have been presented elsewhere.[3]

Background or matrix interference can occur during the analysis of macromolecules that is not detected during assay validation. Validation studies typically employ single-donor or pooled plasma from normal volunteers to prepare calibration standards and quality control specimens. It is highly recommended that matrix dilution experiments, using several different lots of matrix, are performed during method development to assess whether potential selectivity issues exist with the assay. If this is not done, it may only be with the analysis of validation QCs prepared in multiple lots of matrix or "pre-dose" plasma specimens, in the first stages of actual sample analysis, that potential background problems will be revealed.

There are several reasons why the analysis of incurred study samples can vary when compared with the standards and QC samples employed to analyze macromolecules:

- Biotransformation of proteins/peptides can lead to metabolites that display different levels of cross-reactivity (and nonparallelism) vis-à-vis their parent.
- Elevated levels of proteases in disease states can lead to instability of macromolecules when compared with normal plasma.
- The existence of disease-specific pseudoreceptors and endogenous inhibitors in plasma can bind to the ligand of interest and alter its binding characteristics in the immunoassay.
- During the method development stages for noncompetitive ligand binding assays (LBAs), higher concentrations of analyte should be employed to test for and, if necessary, account for possible "prozone or hook" effects that will produce diminished assay responses at higher analyte concentrations. If this is not done, the analysis of samples with high levels of analyte may not be reproducible on reanalysis.

Any time- or storage-related instability in protease, pseudoreceptor or inhibitor activity can cause reproducibility problems in incurred sample results from one occasion to the next.

Despite these potential issues, few laboratories have traditionally performed routine repeat assays using incurred samples. Among 227 respondents to an AAPS survey conducted from November 2005 to February 2006, 89% of pharmaceutical scientists responded "no" to the following question: "Do you reanalyze any randomly selected study samples for confirmation?" Only 8% of respondents replied that they routinely reassayed between 10% and 15% of study samples for confirmation.[7] Indeed the Health Products and Food Branch of Canada recommended that 15% of the incurred samples for a study must be randomly selected and reassayed.[8,9] The requirement for 15% repeat analysis of incurred samples for bioavailability submissions was subsequently removed by the Canadian Ministry of Health in September 2003.[10]

Why don't most scientists perform replicate analysis of incurred study samples to assure assay reproducibility? There would seem to be at least the 4 following reasons:

- There is no consensus regarding the methodology and criteria to evaluate the level of agreement between original and repeated results.
- It is unclear what follow-up action may be required if repeat assay(s) do not agree with the original assay result. How does one handle situations where most repeated results agree with the original result while others do not agree?
- Guidelines as to the manner in which repeated results are handled, reported, and used need to be conceptualized.
- There may be added cost for the extra work involved.

Viswanathan et al[3] have suggested that an evaluation of the reproducibility in the analysis of incurred samples be performed on each species used for Good Laboratory Practices (GLP) toxicology assessments, as well as an appropriate evaluation of incurred sample reproducibility from clinical studies. This article will present one approach to the confirmatory reanalysis of incurred samples. In our opinion, confirmatory reanalysis is only worthwhile if the approach employed is scientifically and statistically sound and practical. This approach is at best a beginning and is offered as a platform for future discussion, comments, and revision.

Several practical and scientific questions arise when contemplating an evaluation of the reproducibility of incurred samples. The most fundamental of these questions are

- In what type of studies should samples be reanalyzed?
- How many and which samples should be reanalyzed to assure ourselves that the assay is reproducible?
- What constitutes acceptable reproducibility and in what manner should the data be analyzed to arrive at valid conclusions?
- What actions, if any, should be taken, once the analysis is completed, and the results have been evaluated?

Each of these questions will be considered below.

In What Types of Studies Should Samples Be Reanalyzed?

The main objective in performing confirmatory reanalysis of incurred samples is to demonstrate that the assay is

reproducible. Over the course of a drug development program, it is likely that a given assay will be validated and used to analyze samples from a variety of different matrices. In order to ensure that the assay is reproducible in each matrix, a subset of samples in each matrix should be reanalyzed. Once such an analysis is performed, it can be used to support the reproducibility of the assay in the same matrix from other studies. For example, if an incurred sample evaluation produces acceptable results for samples obtained from a normal, healthy volunteer study, then such an evaluation will not be necessary for other studies conducted in normal, healthy volunteers. Hence, it is unnecessary to reanalyze samples from each study performed.

In some cases, immunoassay laboratories analyze study specimens at multiple dilutions. This practice provides a rigorous evaluation of incurred sample reproducibility and further reanalysis of incurred samples is unwarranted. For laboratories that analyze samples at a single dilution, reanalysis of incurred samples may be necessary.

During preclinical toxicology assessments, studies are performed in several animal species to assess the safety of the drug. Based on the recommendations of Viswanathan et al,[3] a subset of samples from each matrix should be reanalyzed to ensure that the assay is reproducible. One practical issue that can complicate this assessment is that the volume of sample that can be taken from certain animal species may preclude a reanalysis. In this event, the reanalysis of pooled samples on more than one occasion is preferable to not conducting this evaluation at all.

For clinical studies, it has been proposed that "an appropriate evaluation of incurred sample reproducibility" be performed.[3] What constitutes an "appropriate evaluation" is highly dependent on each drug's individual characteristics and how much is known about the drug. Nonetheless, we offer the following general guidelines:

Normal, Healthy Volunteer Studies Involving Either Small Molecules or Macromolecules

The reanalysis of a subset of samples from a single study should be sufficient to support all of the studies conducted in normal healthy volunteers. In addition, if small molecule drug interaction trials are anticipated to raise the blood levels of drug metabolites (measured or otherwise), then it would make sense to reanalyze a subset of samples from a single study of this type in order to evaluate assay reproducibility. This reanalysis is only required if a significant positive drug interaction is observed.

Bioequivalence studies used to support the approval of generic versions of marketed drugs are considered part of this category. As a result, reanalysis of a subset of samples from one study supporting the application should be performed.

Studies Conducted in Various Patient Populations

Patient population studies need to be evaluated on a case-by-case basis. This evaluation should include an assessment of whether the disease being studied is likely to produce a sample matrix that is substantially different and variable from patient to patient when compared with the same type of matrix taken from normal, healthy volunteers. It should also include an evaluation of whether the pharmacokinetics of the drug in question is likely to be appreciably altered in the disease state. If the answer to both of these questions is "no," then assessing the reproducibility of the assay should not be necessary. In instances where the sample matrix is different but not highly variable among patients, partially validating the assay in pooled matrix obtained from patients should be considered.

If the answer to either or both of the above questions is "yes," then an evaluation of the assay's reproducibility should generally be performed for a single study for each disease being studied. For example, the plasma from patients with moderate-to-severe renal insufficiency is different and can be highly variable in its quantitative composition from patient to patient. Furthermore, this degree of renal disease is likely to alter the pharmacokinetics of small molecules and metabolites eliminated by the kidney. As a result, the reproducibility of the assay in this disease should be evaluated in a single study. For macromolecules, disease-specific instability issues might exist, such as elevations in protease levels secondary to disease that could potentially affect the reproducibility of assay results. In addition, other disease-related factors such as the presence of pseudoreceptors and endogenous inhibitors can affect the reliability and reproducibility of immunoassay results. Should such disease-related factors exist, reanalysis of a subset of samples from a single study conducted in patients with a particular disease should be performed, if the assay has not been validated in the disease-specific matrix.

How Many and What Types of Samples Should Be Reanalyzed to Confirm That the Assay Is Reproducible?

It is important to select enough samples for reanalysis, so that meaningful conclusions regarding the reproducibility of the assay can be drawn. Analyzing more samples than necessary can be burdensome and increase the overall costs associated with the analytical work. The number of samples used for the repeat testing of incurred samples should be statistically justified. We provide here an example of such a justification realizing that it may be necessary to further alter or refine this approach, depending on the variability of the particular assay under consideration and the purpose for which the results will be used.

Differences between assay results can be either (1) systematic—where the results of repeat assays are biased higher or lower than the original assay result, or (2) random—where the results of repeat assays do not agree with the original results but no bias exists. In the case of a systematic difference, we consider testing enough samples to detect a 20% or greater systematic difference in the repeat results a desirable objective (25% for macromolecules). To be on a statistically sound footing, we would like to detect this difference with 80% power and a 5% type I error (which is the probability of falsely claiming a difference when none exists).

In order to calculate the number of samples required for reanalysis, it is necessary to know what the intrasample coefficient of variation (CV) is for the assay, which can be defined as the variation observed when a single sample is reanalyzed over multiple runs. In the absence of prior biological data to estimate the intrasample variability, the intermediate precision of the assay determined during prestudy validation may be used to make a reasonable assumption on the degree of intrasample variability.

For small molecules, we usually require that samples at the assay's lower limit of quantitation (LLOQ) have a CV of 20% or less, while those above the LLOQ have a CV no greater than 15%. If we use a CV of 20% in the calculation, then the reanalysis of 12 samples would be sufficient to determine whether a 20% systematic difference exists. For LBAs, the CV at the LLOQ and the upper limit of quantitation (ULOQ) often used as an acceptance criterion is 25%. In this instance, approximately the same number of samples should be analyzed in order to detect a 25% difference.

In the event that the difference in assay results is random, the number of samples required would need to be greater than that required to detect a systematic difference. Generally speaking, reanalysis of ~20 samples for small molecule assays and LBAs should be sufficient to detect 20% to 25% differences in assay results, as will be illustrated in the examples below.

It is our opinion that selecting, for example, 2 samples/per subject for 10 subjects in a study would produce a more rigorous evaluation of incurred sample reproducibility when compared with analyzing a complete pharmacokinetic profile for 1 or 2 subjects because intersubject differences in matrix composition can exist, particularly in certain disease states. For small molecules, we suggest that a sample at or near the peak level be selected, as well as one later in the drug elimination phase, as these are most likely to have higher levels of drug metabolites present. A similar sample selection strategy should be used for macromolecules. In this case, evaluating peak levels and levels near the LLOQ of the assay provides a sufficient assessment of the reproducibility of the results.

In What Manner Should the Data Be Analyzed to Arrive at Valid Conclusions?

Appropriate statistical methods should be used to determine the level of reproducibility of incurred samples. The assessment of reproducibility should include the testing of systematic difference between the results. It should also include the characterization of the degree of agreement between the results, which should prove useful in detecting random differences between assay results.

These objectives can be met through the use of a Bland-Altman plot[11] and the estimation of what we term the 67% "limits of agreement" between the results, both of which will be explained in greater detail below.

We do not recommend that correlation coefficients such as the intraclass correlation, Pearson's correlation, or Spearman's correlation be used to evaluate differences in sample results because the correlation of such data are heavily influenced by the range of the data.[12] Data that are narrower in range showing similar agreement as other data with broader range may look inferior with respect to the correlation values.

The following determinations are recommended for evaluating the level of reproducibility of assay results using incurred samples. Although the definitions and criteria are different, the concepts proposed here are similar in spirit to earlier publications.[11,13]

1. Accuracy Assessment. The mean ratio of the sample results should be determined as described below, along with its 95% confidence interval. If the minimum acceptable systematic difference between the runs is 20% on average, then the 95% confidence interval of this ratio should be within 0.83 and 1.2. The interval for a 25% systematic difference is 0.80 to 1.25. In both instances the confidence interval calculated needs to include 1, in order to conclude that there are no statistically significant systematic differences between initial and reanalyzed results, but this is not required (ie, an interval of 1.05 to 1.15 would be acceptable as it is within 0.8 to 1.25; although the systematic difference is statistically significant, it is well within 25%).
2. Precision Assessment. The 67% limits of agreement of the ratio of sample results should be determined as described below. These limits are interpreted as the range within which the ratio of sample results is expected to fall two thirds of the time. If the difference between any 2 repeat samples should be within 20% of each other, then these limits of agreement should be within 0.83 to 1.2. The interval for a 25% difference is 0.80 to 1.25.

It should be noted that the mean ratio suggested in (1) above relates to the bias or trueness of the analytical results. The limits of agreement suggested in (2) above relates specifically to the imprecision of the analytical results. Since the bias and precision of the analytical results are 2 primary elements of analytical validation,[1-6] these determinations for the repeat testing of incurred samples are consistent with the performance characteristics derived during prestudy analytical validation of spiked samples. The criteria used for the mean ratio in (1) and limits of agreement in (2) are consistent with those used during the prestudy and in-study validation of the assay.[6] In addition, the motivation for using the 67% (rather than 95%) limits of agreement comes from the "4-6-20 rule"[1] that is often used to accept analytical runs for small molecule liquid chromatography/tandem mass spectroscopy (LC/MS/MS) assays. Given this rule, imposing a 95% limits of agreement criterion during the repeat analysis of incurred samples would be far too stringent. The 67% limits of agreement rule was developed to permit more variation than is acceptable in a single analytical run. Furthermore, it is easy to calculate since 67% represents ±1 standard deviation (SD) and appears to work well when applied to sample data for small molecules and macromolecules (see Statistical Computations, examples 1-3).

Statistical Computations

The statistical analysis for the above determinations is outlined below. This is similar to the steps outlined by Eastwood et al,[13] for evaluating the reproducibility of compound potency results between 2 results.[14]

1. Compute the difference ratio for individual samples as the difference in log (base-10) transformed values between the first and second run for each sample. Let $\bar{d}$, and s_d be the sample mean and standard deviation, respectively, of the difference in log transformed values. Since ratios of sample results are usually more meaningful than differences, log transformation allows the ratios to be analyzed as differences. The following calculations are needed for the 2 determinations outlined above. They can be easily carried out using many commercially available software packages.
2. Compute the geometric mean for each sample as (first result × second result)$^{1/2}$.
3. Compute the mean-ratio: $MR = 10^{\bar{d}}$. This is the average fold change in the sample results between 2 runs. The 95% confidence interval of this mean ratio, the ratio limits (RLs) is calculated using the formula: $RLs = 10^{\bar{d} \pm 2s_d/\sqrt{n}}$, where n is the number of samples.

Table 1. Analytical Data Presented for a Small Molecule Analyzed by LC/MS/MS*

Subject	Original Result (ng/mL)	Repeat Result (ng/mL)	Percentage Difference†
1	478	406	–16.3
2	107	107	0.0
31	826	718	–14.0
31	108	109	0.9
3	248	250	0.8
4	696	674	–3.2
4	141	135	–4.3
5	194	179	–8.0
6	548	564	2.9
7	676	598	–12.2
9	636	676	6.1
2	635	624	–1.7
2	244	240	–1.7
9	527	579	9.4
9	139	117	–17.2
10	107	99.3	–7.5
33	664	583	–13.0
33	187	176	–6.1
33	690	610	–12.3
33	187	190	1.6

*LC/MS/MS indicates liquid chromatography/tandem mass spectrometry.
†(Repeat-original)/average expressed as a percentage.

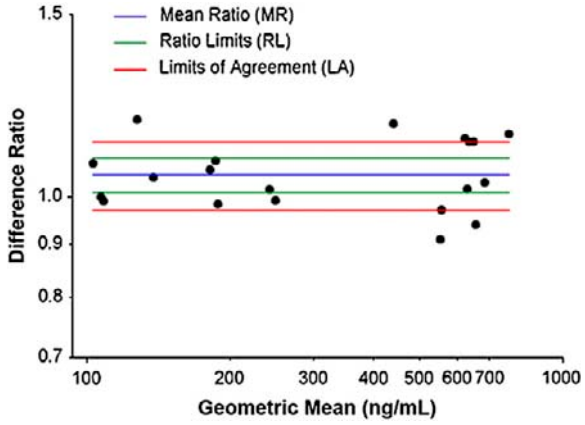

Figure 1. Bland-Altman plot of the data presented in Table 1.

4. Compute the limits of agreement: $LsA = 10^{\bar{d}+Sd}$. The ratio of repeat results for any sample is expected to fall within these limits two thirds of the time.

The above calculations should be accompanied by a plot of the difference ratio vs the geometric mean of the repeat results.[11] These calculations are illustrated in the examples which follow.

Example 1: Analysis of a Small Molecule by LC/MS/MS

Twenty incurred plasma samples, initially analyzed using an LC/MS/MS assay were reanalyzed using the same procedure in a different run. The results obtained, as well as their percentage difference, calculated using a traditional method, are presented in Table 1. These data were evaluated statistically using the procedure described above with the detection of a 20% difference considered desirable. The results of this analysis are presented in Figure 1.

Statistical analysis of this data revealed a mean ratio (MR) of 1.01; ratio limits (RL) of 1.01 to 1.09, not including 1 (acceptance range, 0.83-1.20); and limits of agreement (LA) of 0.97 to 1.13 (acceptance range, 0.83-1.20). The acceptance ranges for both the RL as well as the LA were met. The fact that the RL did not quite include 1 tends to suggest a very slight bias to the data. As can be seen upon inspection of the percentage differences in Table 1, 13 of the percentage differences calculated are negative compared with 6 that were positive. An evaluation as to the potential cause for this slight negative bias could prove useful. Aside from this minor bias, the data appear to be reproducible.

Example 2: Analysis of a Macromolecule by Enzyme-Linked Immunosorbent Assay

Twenty incurred plasma samples, initially analyzed using an enzyme-linked immunosorbent assay (ELISA), were reanalyzed using the same procedure in a different run. The results obtained, as well as their percentage difference, calculated

Table 2. Analytical Data Presented for a Macromolecule Analyzed by ELISA*

Subject	Original Result (ng/mL)	Repeat Result (ng/mL)	Percentage Difference†
1	1.96	1.59	20.9
2	0.779	1.00	–24.8
3	0.0577	0.0406	34.8
4	0.00462	0.00333	32.4
5	0.0403	0.0363	10.4
6	0.0973	0.0773	22.9
7	0.840	0.758	10.3
8	4.14	3.26	23.8
9	0.645	0.617	4.44
10	0.196	0.211	–7.37
11	1.16	1.16	0.00
12	3.35	4.46	–28.4
13	0.330	0.273	18.9
14	0.352	0.346	1.72
15	3.88	3.63	6.66
16	2.28	2.35	–3.02
17	0.0601	0.0614	–2.14
18	0.0747	0.0797	–6.48
19	0.221	0.236	–6.56
20	0.170	0.180	–5.71

*ELISA indicates enzyme-linked immunosorbent assay.
†(Repeat-original)/average expressed as a percentage.

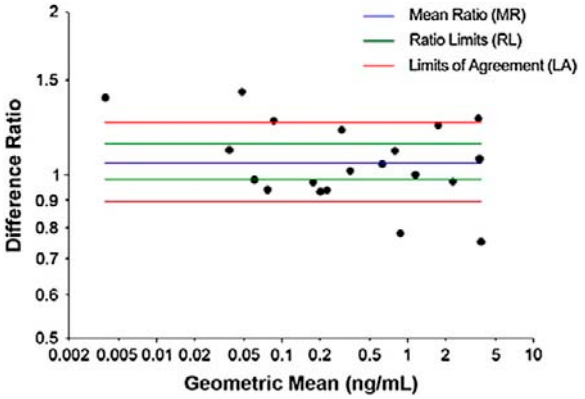

Figure 2. Bland-Altman plot of the data presented in Table 2.

using a traditional method, are presented in Table 2. These data were evaluated statistically using the procedure described above with the detection of a 25% difference considered desirable. The results of this analysis are presented in Figure 2.

Statistical analysis of this data revealed an MR of 1.05; RL of 0.98 to 1.14, including 1 (acceptance range, 0.80-1.25); and LA of 0.89 to 1.25 (acceptance range, 0.80-1.25). The acceptance ranges for both the RL as well as the LA were met. As a result, the method appears reproducible. The agreement between the replicate measurements for this data set appear to be at the outer limits of acceptability as evidenced by the closeness of the upper LA with the upper limit of the acceptance range. Please note that based on the percentage differences in Table 2, the greatest percentage differences in results occurred at the low and high concentration levels, further supporting the sample selection strategy discussed above.

Example 3: Analysis of a Small Molecule by LC/MS/MS

Twenty incurred plasma samples, initially analyzed using an LC/MS/MS assay, were reanalyzed using the same procedure in a different run. The results obtained, as well as their percentage difference, calculated using a traditional method, are presented in Table 3. These data were evaluated statistically using the procedure described above with the detection of a 20% difference considered desirable. The results of this analysis are presented in Figure 3.

Statistical analysis of these data revealed an MR of 1.01; RL of 0.97 to 1.05 including 1 (acceptance range, 0.83-1.20); and LA of 0.92 to 1.10 (acceptance range, 0.83-1.20). The acceptance ranges for both the RL as well as the LA were met, suggesting that the data are reproducible.

What Actions, if any, Should Be Taken, Once the Analysis Is Completed, and the Results Have Been Evaluated?

The primary purpose of reanalyzing incurred samples is to confirm the reproducibility of the bioanalytical method. If

Table 3. Analytical Data Presented for a Small Molecule Analyzed by LC/MS/MS*

Subject	Original Result (ng/mL)	Repeat Result (ng/mL)	Percentage Difference†
1	80.1	87.6	8.94
2	1891	1981	4.65
3	194	201	3.54
4	288	297	3.08
5	416	431	3.54
6	108	103	–4.74
7	172	165	–4.15
8	482	525	8.54
9	267	243	–9.41
10	611	697	13.15
11	482	499	3.47
12	345	365	5.63
13	459	511	10.72
14	189	183	–3.23
15	783	646	–19.17
16	191	171	–11.05
17	508	413	–20.63
18	235	224	–4.79
19	583	556	–4.74
20	209	217	3.76

*LC/MS/MS indicates liquid chromatography/tandem mass spectrometry.
†(Repeat-original)/average expressed as a percentage.

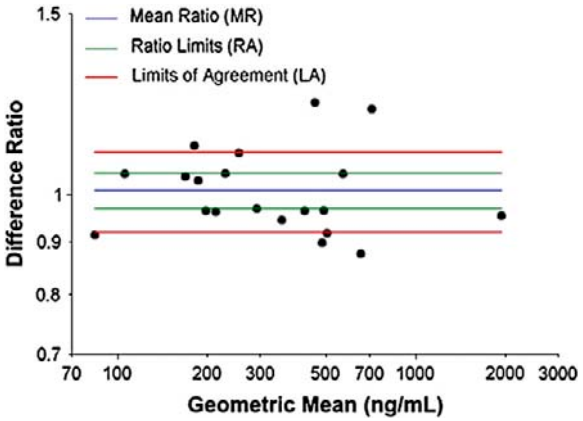

Figure 3. Bland-Altman plot of the data presented in Table 3.

the results suggest that the method is reproducible, we suggest reporting the initial result as the "final" reported value. We also suggest presenting the original and repeat results for the samples that were reanalyzed in tabular format along with the results of the statistical analysis in a separate section of the bioanalytical report.

In the event that the method proves reproducible, but disparate results are observed in a few of the samples, it may be worthwhile to investigate and document whether the samples in question have anything in common (all from the same patient, all are peak concentrations, etc). Depending on the results of this investigation, further laboratory work may be warranted to resolve any issues that are uncovered.

In the event that the method is not reproducible, further work will be necessary to investigate and resolve this problem.

Conclusion

The evaluation of bioanalytical methods through the reanalysis of incurred samples can be taken as one additional measure of assay reproducibility. This reanalysis expands upon the data collected during the prestudy validation of the method, since incurred samples may contain elements that are not present in the standards and quality control samples used during prestudy validation experiments. We have attempted to outline an approach to this reanalysis that is reasonable, practical, and statistically based. Furthermore, the data treatment required for this approach is straightforward and can be accomplished with off-the-shelf spreadsheet programs. The acceptance criteria discussed in this manuscript appear reasonable given the criteria typically employed in the bioanalysis of macromolecules and small molecules but should be further tested and adjusted as necessary.

References

1. Shah DD, Midha KK, Dighe SV, et al. Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies. *J Pharm Sci*. 1992;81:309-312.

2. Shah VP. The history of bioanalytical method validation and regulation: evolution of a guidance document on bioanalytical methods validation. *AAPS J*. 2007;9:E43-E47.

3. Viswanathan CT, Bansal S, Booth B, et al. Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *AAPS J*. 2007;9:E30-E42.

4. Findlay JWA, Smith WC, Lee JW, et al. Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J Pharm Biomed Anal*. 2000;21:1249-1273.

5. FDA/CDER. *Guidance for the Industry. Bioanalytical Method Validation.* Rockville, MD: US Department of Health and Human Services FDA (CDER) and (CVM); May 2001.

6. DeSilva B, Smith W, Weiner R, et al. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res*. 2003;20:1885-1900.

7. Bansal SB. AAPS Bioanalytical Survey. Paper presented at the AAPS 3rd Bioanalytical Workshop: Quantitative Bioanalytical Methods Validation and Implementation—Best Practices for Chromatographic and Ligand Binding Assays; May 1-3, 2006; Washington, DC.

8. Canadian Health Ministry. *Guidance for Industry. Conduct and Analysis of Bioavailability and Bioequivalence Studies—Part A: Oral Dosage Formulations Used for Systemic Effects.* 1992; Available at: http://hc-sc.gc.ca/dhp-mps/prodpharma/applic-demande/guide-ld/bio/bio-a_e.html. Accessed September 27, 2007.

9. Canadian Health Ministry. *Guidance for Industry. Conduct and Analysis of Bioavailability and Bioequivalence Studies—Part B: Oral Modified Release Formulations Used for Systemic Effects.* 1996; Available at: http://hc-sc.gc.ca/dhp-mps/prodpharma/applic-demande/guide-ld/bio/bio-b_e.html. Accessed September 27, 2007.

10. Canadian Health Ministry. *Notice to Industry—Removal of Requirement for 15% Random Replicate Samples*. 2003; Available at: http://hc-sc.gc.ca/dhp-mps/prodpharma/applic-demande/guide-ld/bio/15rep_e.html. Accessed September 27, 2007.

11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;i:307-310.

12. Bland JM, Altman DG. A note on the use of intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med*. 1990;20:337-340.

13. Eastwood BJ, Farmen MW, Iversen PW, et al. The minimum significant ratio: a statistical parameter to characterize the reproducibility of potency estimates from concentration-response assays and estimation by replicate-experiment studies. *J Biomol Screen*. 2006;11:253-261.

14. Eli Lilly and Company and the National Institutes of Health (NIH) Chemical Genomics Center (NCGC). Transfer of validated assays, 2005. Available at: http://www.ncgc.nih.gov/guidance/section2.html#analysis-potency. Accessed September 27, 2007.