Journal of Pharmaceutical and Biomedical Analysis 53 (2010) 729–734

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Short communication

A strategy for improving comparability across sites for ligand binding assays measuring therapeutic proteins

Chad A. Ray[c,*], Lei Zhou[b], Jennifer Tsoi[a], Lennie Uy[a], Jessie Gu[b], Jennifer Malella[a], Danielle DeSimone[a], Han Gunn[a], Mark Ma[a], Jean Lee[a], Binodh DeSilva[a]

[a] Pharmacokinetics and Drug Metabolism, Amgen Inc., Thousand Oaks, CA, United States
[b] GBE Biostats Med Sciences, Amgen Inc., Thousand Oaks, CA, United States
[c] Pharmaceutical Services, Radix BioSolutions, Georgetown, TX, United States

ARTICLE INFO

Article history:
Received 7 January 2010
Received in revised form 13 March 2010
Accepted 19 April 2010
Available online 24 April 2010

Keywords:
Bioanalytical comparability
Ligand binding assay
Pharmacokinetics

ABSTRACT

Outsourcing and multi-site testing has increased for ligand binding assays supporting protein therapeutic measurement. It is common to combine and compare data across studies with data from multiple bioanalytical sites. We designed a prospective study to determine the benefits of increasing control over the transfer process to improve ruggedness. The experiment involved the testing of 30 incurred samples at 3 stages with incremental laboratory harmonization in standard/quality controls and assay components: Stage I represented a transfer of a detailed protocol and critical reagents. Stage II, a single source of standards and quality controls were provided to each site. Stage III, standards and quality controls plus a ready-to-use kit were provided. The results indicated that all testing facilities failed agreement testing using the stage I procedure. The introduction of standards from a single source improved the agreement. The modification reduced variation by 33% compared to the stage I approach. There was no additional benefit when a packaged kit was provided. In conclusion, introduction of a single source of standards and quality controls reduced the inter-site component of variation and should allow for combinability of data.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Thought leaders in the field of ligand binding assay development have been successful at increasing the awareness of the limitations associated with ligand binding assays. They have written recommendations and guidance documents describing the requirements for validation of ligand binding assays used to support the pharmacokinetic measurement [1–3]. The guidance documents have placed an emphasis on analytical validation best practices, proper statistical assessment of the data, and recommendations on what constitutes a valid assay. As part of the best practices, it was recommended that robustness testing be conducted. The definition of robustness is being able to withstand stresses, pressures, or changes in procedure or circumstance within the anticipated boundaries. The translation of robustness within the laboratory has been to incorporate multiple operators, to perform the assay over multiple days, to prepare different lots of reagents, and to conduct the experiments using different incubators, washers, and plate readers. The sum total of all of these stresses should provide an estimate of variation that reflects the worst case scenario. However in modern drug development, a paradigm shift has occurred to more outsourcing and multi-site testing which creates challenges in method transfer [4]. Not only is cross site transfer challenging, but there seems to be a lack of consensus on whether it constitutes a new validation parameter. The International Conference on Harmonization (ICH) guidelines combined the terms robustness and ruggedness testing into a single definition [5]; however, the United States Pharmacopeia separates the two terms and defines method robustness as reproducibility with deliberate changes such as different laboratories [6]. Viswanathan et al. described a need for multi-site testing in the guidance document of 2007 under the section of cross-validation [7]. Regardless of the name, some evaluation of cross site testing is needed to reliably combine data across studies and laboratories.

As a result of the need, plans to test method comparability across sites must be carefully thought out and carried out. It is now imperative that sponsors understand the variation associated with changing bioanalytical sites. Sponsors should also attempt to reduce the variation to a level where data from different studies and sites can be combined. Otherwise, the variability will confound the interpretation of data and prevent comparisons across studies.

In general, the common practice for method transfer from the sponsor laboratory involves the transfer of a detailed procedure and

* Corresponding author Current address: 111 Cooperative Way, Suite 120, Georgetown, TX 78626, United States. Tel.: +1 512 869 8000; fax: +1 512 868 9040.
E-mail address: caray@radixbiosolutions.com (C.A. Ray).

0731-7085/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2010.04.022

all critical reagents required to perform the assay at the other site. We designed a prospective study to determine the level of agreement between sites using the common practice of assay transfer, a modified transfer approach that includes ready-to-use standards and quality controls in 100% human serum prepared by the sponsor, and a stringently controlled transfer where all assay components packaged into ready-to-use kits including standards and quality controls were provided. We tested the laboratories' agreement of incurred sample analysis to the sponsor laboratories results and also defined the variation associated with each approach.

2. Materials and methods

2.1. ELISA procedure

Two antibodies specific for the antibody binding region were used to develop a sandwich immunoassay to quantify a fully human monoclonal in serum. The assay was conducted in a sequential manner with wash steps using KPL Washing Solution (Kirkegaard and Perry Laboratories, Gaithersburg, MD.) following all incubation steps. The therapeutic protein was captured using a monoclonal antibody (Amgen Inc., Thousand Oaks, CA) that was passively adsorbed at 4 °C overnight at a concentration of 25 ng/mL. The sample was incubated for 1 h at room temperature. The detection reagent, a biotinylated rabbit anti-therapeutic polyclonal antibody, was prepared at a concentration of 40 ng/mL and incubated for 30 min at room temperature. Horseradish peroxidase (HRP) labeled NeutrAvidin (R&D Systems, Minneapolis, MN) prepared at 66 ng/mL was added to the wells for 30 min. Peroxide containing substrate solution tetramethylbenzidine (Applied Biosystems, Foster City, CA) was then added to the wells for 30 min at room temperature, which in the presence of HRP created a colorimetric signal that was proportional to the amount of analyte bound by the capture reagent in the initial step. A non-linear four parameter logistic model with weighting was used to fit the calibration model. The optimal non-linear regression model and weighting algorithm based on goodness of fit was selected prior to validation experiments.

The method was validated at Amgen by 3 analysts over multiple days using 3 individually prepared standard curves and a single preparation of QCs. The Std and QCs were prepared by a Tecan EVO using validated scripts. The validation accuracy and imprecision for LLOQ LQC, MQC, HQC and ULOQ validation samples were −5, −7, −11, −10, −14% bias; 6, 5, 5, 7, 7 imprecision and total errors of 11, 12, 16, 17 and 21, respectively.

The method was transferred to Laboratories (A, B, and D) for outsourcing and to Laboratory C within Amgen for additional bioanalytical support.

2.2. Study design

Three graded stages were designed to determine the effect of incremental control on inter laboratory assay agreement. Four laboratories (referred to as A, B, C, and D) participated in an exercise of analyzing 30 incurred samples with a different method transfer approach at each stage. Table 1 lists the components provided in each stage. In stage I, the sponsor (Amgen Inc.) provided a detailed procedure including the correct dilution factors and plate maps and all critical reagents required to perform the assay. In the second stage, the sponsor prepared a set of standards and quality control samples in 100% human serum, aliquoted, labeled, and shipped to the sites. In the third stage, all reagents were provided in a ready-to-use kit. For stages II and III, the same set of samples was diluted in human serum, pre-treated in assay buffer and aliquoted by the individual sites under the prescribed conditions. After all the samples were analyzed, the raw data were sent to Amgen for analysis. The 30 incurred samples were analyzed prior to shipping by the sponsor and considered the original results.

Table 1
Components provided at each stage of the experiment.

Reagents	Stage I	Stage II	Stage III
Reference material	X		
Stock capture reagent	X	X	
Stock detection reagent	X	X	
Human serum	X	X	X
Sponsor Std/QC	–**	X	X
Unknowns (30)	X	X	X
Sponsor kit*			X

* Kit includes: (1) Dry-coated/blocked plates, (2) Detection antibody solution (at working concentration), (3) NeutrAvidin-HRP solution (at working concentration) (4) 1 Component HRP Microwell Substrate (Bio FX Laboratories).
** Std/QC prepared by CRO.

2.3. Statistical analysis

Data analysis was done using SAS V9.3 on a Windows Professional operating system. To assess variance, a mixed effect model analysis was performed on the log of the responses (concentrations) with sample and group included as fixed effect and laboratory included as random effect. A 90% confidence interval for the ratio of the observed concentrations between the two groups was computed by first calculating the difference and its 95% confidence limit in the log of the observed concentration values. The difference and corresponding confidence limit was then exponentiated to obtain the ratio and confidence limit.

3. Results

3.1. Agreement testing

Each laboratory assayed 30 samples according to the protocol. The acceptance criteria were that a 90% confidence interval of the mean ratio must be contained within (0.80, 1.25) in order for two groups to be considered in agreement. The mean ratio was computed from the 30 individual ratios (original result/experimental result) for each laboratory and each stage.

For each laboratory, the sample results for each group were compared against the original result (stage 0) and divided into the three stages as shown in Figs. 1–3. The results were further parsed to separate out Laboratory C. This laboratory was composed of two analysts at Amgen previously not associated with the project. They prepared standards and quality control samples using the same liquid handler that was used for the original analyzes. As shown in Figs. 1–3, higher concentrations were observed for almost all samples and laboratories when compared to the original or control. This is represented by most points being above the diagonal line. The results for stage II (Fig. 2) and stage III (Fig. 3) were also above the 45° line; however, the results showed less departure from the 45° line and less spread compared to the results in Fig. 1. These results indicate that the increased control over the transfer process reduced the bias between results and improved the precision. These results obtained within the same laboratory demonstrated better agreement than outside laboratories. Agreement testing for stage I indicated that all external laboratories failed to agree with the original results. The only results that were in agreement were group C results (Table 2). All other laboratories had confidence limits that were outside the a priori acceptance of (0.8, 1.25). The estimated mean ratios of the original results vs. stage I results are 0.86 and 0.92 for the two analysts (C-1 and C-2), respectively. Corresponding 90% confidence intervals (CI) for the mean ratio are (0.84, 0.88) and

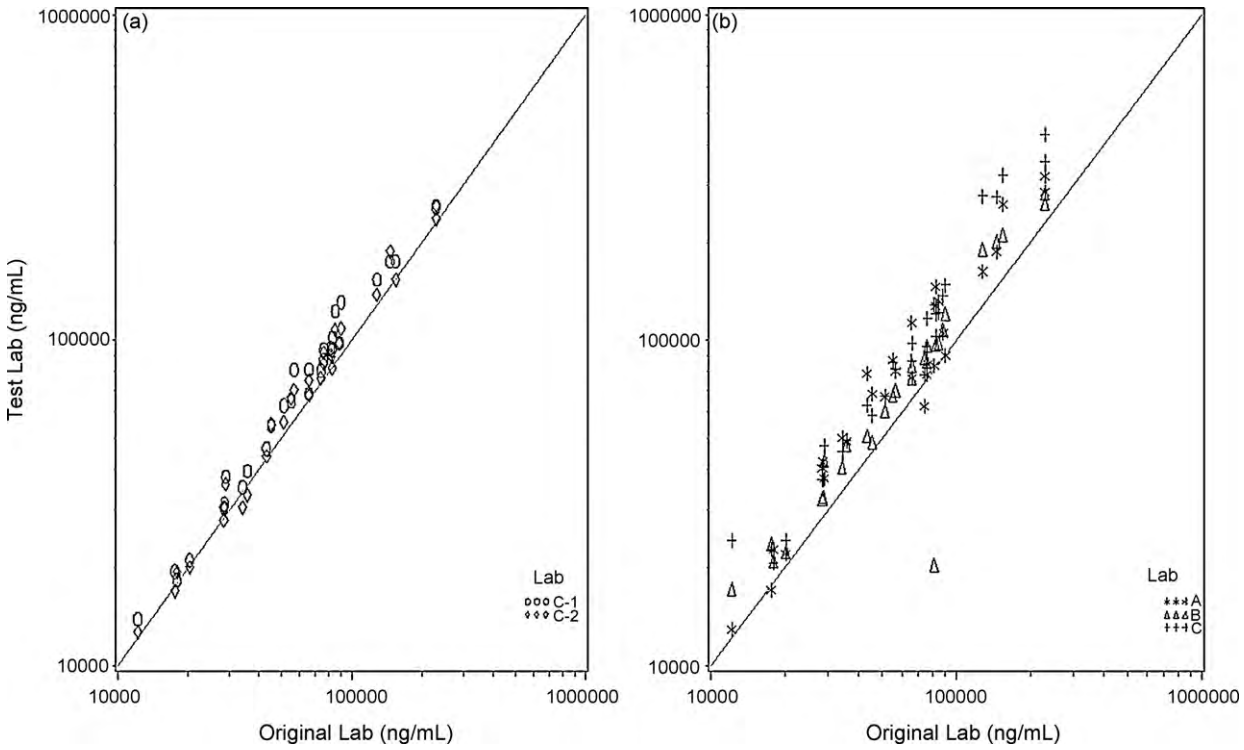

Fig. 1. Between and within laboratory comparison for stage I. (a) Two analysts from Amgen (Laboratory C) followed the stage I protocol and measured 30 incurred samples. All test results were compared to the original (control) results. The open circles represent analyst 1 and the open diamonds represent analyst 2. (b) Three outside laboratories also followed the stage I protocol and compared their results to the original (control). The following symbols were used to represent the respective laboratories: A (*), B (Δ) and D (+). The 45° line in both figures represents the line of absolute agreement.

(0.90, 0.94), both fall within the equivalence range of (0.80, 1.25). Stage II results showed that all laboratories met the acceptance criteria in agreement with the original results, with the exception of Laboratory A and B where the mean ratios of the original vs. stage II results are 0.82 (90% CI is (0.78, 0.86)) and 0.83 (90% CI is (0.76, 0.89)). However, if we remove two outliers (samples measured at 74,100 and 81,300) from Laboratory B results (B*), the mean ratio increased slightly to 0.85 and the 90% CI for the mean

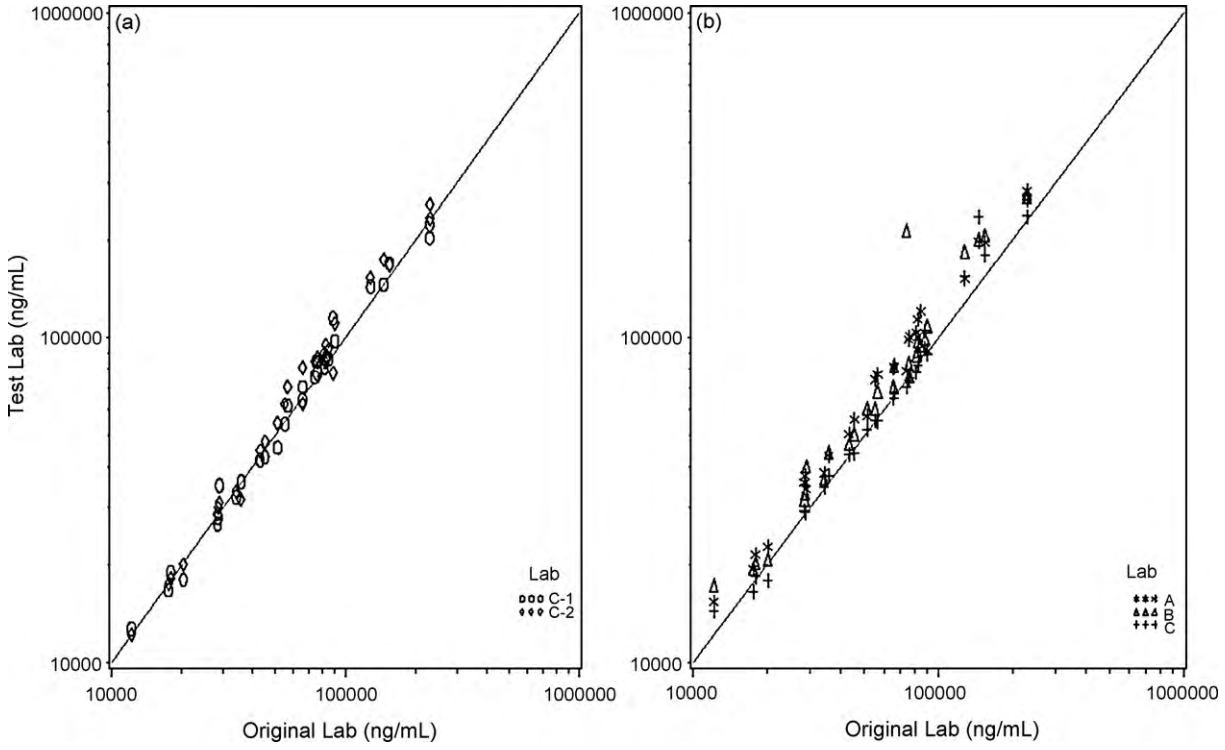

Fig. 2. Between and within laboratory comparison for stage II. (a) In stage II, standards and QCs were prepared by the sponsor laboratory. The open circles represent analyst 1 and the open diamonds represent analyst 2 results compared to the original results. (b) Three outside laboratories: A (*), B (Δ) and D (+) followed the stage II protocol and measured 30 incurred samples and compared their results to the original. The 45° line in both figures represents the line of absolute agreement.

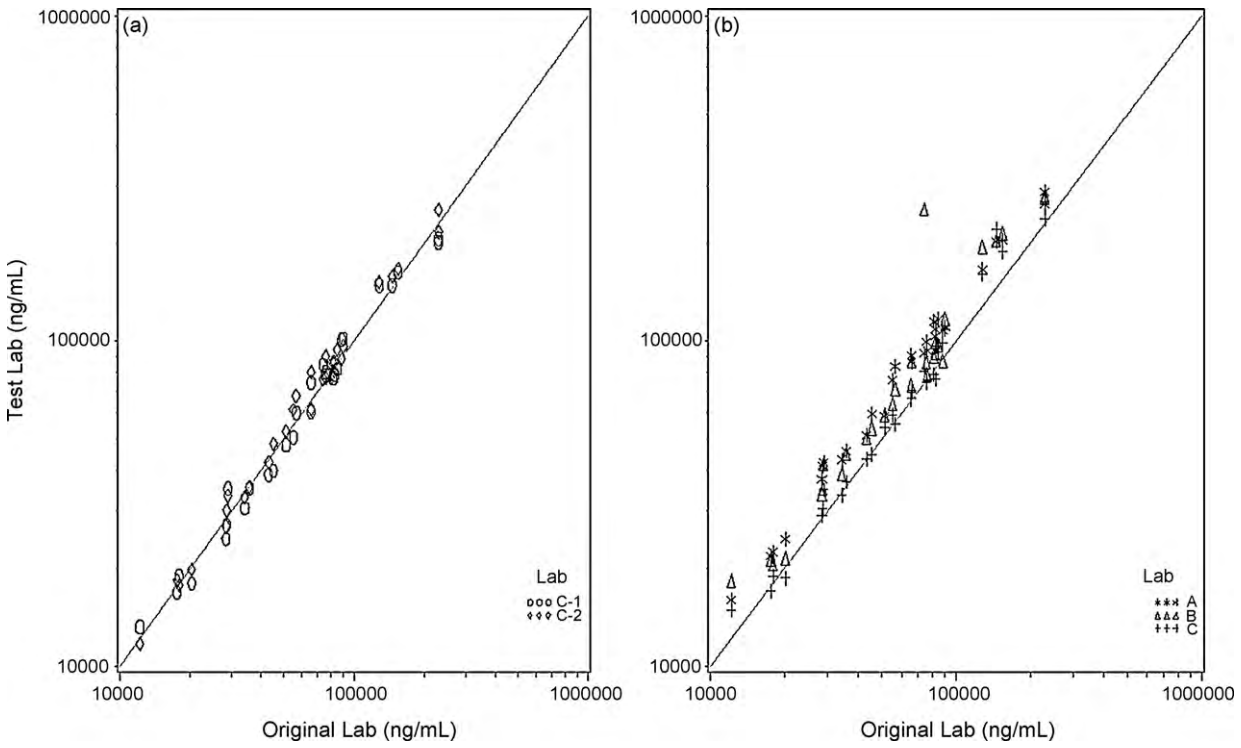

Fig. 3. Between and within laboratory comparison for stage III. (a) The stage III protocol was the same as stage II with the addition of a ready-to-use kit prepared by the sponsor laboratory. The open circles represent analyst 1 results and the open diamonds represent analyst 2 results compared against the control (original) results. (b) Laboratories: A (*), B (Δ) and D (+) followed the stage III protocol and measured 30 incurred samples and compared their results to the original. The 45° line in both figures represents the line of absolute agreement.

ratio becomes (0.83, 0.87) which does fall within the acceptance range. Stage III demonstrated the same trend as stage II. Laboratory A and B failed to meet the acceptance criteria. There were cases where the *p*-value is <0.0001 but the 90% confidence interval met the equivalence criterion. For example, Laboratory B with outliers removed, the mean ratio is 0.82 and the corresponding 90% CI is (0.80, 0.84), which again demonstrates that the results agree if outliers are removed. In each case the confidence interval does not include 1.0, the potential bias may be statistically significant. Furthermore, within a laboratory all results pass the a priori agreement criteria with the exception of Laboratory D where the mean ratios comparing stage I to stage II and stage I to stage III are substantially higher than the other comparisons within Laboratory D, similarly for the 90% confidence intervals.

To further test the impact of additional control on data harmony, the sample results for each group were compared against the original result and to each other across groups and the results were included in Table 3. The only stage that failed to meet the acceptance criteria of agreement was stage I when compared to the original data. That same stage had a mean ratio of 0.81 indicating that average observed concentrations for stage I were 19% higher than the original result on average. The 90% confidence interval for the mean ratio is (0.79, 0.82) which falls outside the equivalence range.

3.2. Variance assessment

The total variability observed in each test stage was separated into the contribution from the differences in laboratories and random noise (residual). The majority of the total variability was associated with random variation. The standard deviation of log transformed concentrations was 0.21 for stage I and 0.14 for stage II which translate to 21% and 14% coefficient of variation (CV) on the concentration scale, Table 4. The change in standard deviation from 0.21 to 0.14 represents a 33% improvement in percent CV of observed concentrations by introducing a single source of STDs/QCs.

4. Discussion

The main objective of the study was to better understand the differences in absolute value of patient samples when measured by multiple laboratories. In our study the original result was used as a comparator to simulate the actual condition that occurs in oncology clinical trials. For drug–drug interaction studies, the first in human results become the gold standard pharmacokinetic response, so all subsequent studies must be consistent with those results, otherwise the interpretation becomes confounded. For this reason, we chose to compare all laboratories to a single control result. The alternative and more statistically sound approach would have been to compare to the mean results from all observations, but it would not have reflected the real world scenario. A secondary objective was to determine if control mechanisms could be introduced that would reduce the differences across laboratories. The hypothesis was that each level of control (such as those introduced in stage II and stage III) would reduce the total variability in measurement. The results indicated that incorporating a single source of standards and QCs reduced the overall variation by 33% compared to the common practice of method transfer. Surprisingly, the ready-to-use kits did not provide additional benefits in this test. For this method, a single source of working standards provided the harmonized scale for quantification. Another important conclusion was that preparing standards and QCs within a site and computing the total error was not reflective of the total variation that existed across sites (data not shown). In addition, some of the sites had liquid handling systems for preparing the standards and quality controls and were unable to obtain agreement. These results indicate that liquid class optimization and harmonization are essential to achieving agreement in results. The final conclusion was that introducing a

Table 2
Agreement testing between laboratories.

Laboratory	Stage	_Stage	*p*-value	Ratio	Lower bound of 90% confidence interval of ratio	Upper bound of 90% confidence interval of ratio
A	0	1	<0.0001	0.77	0.73	0.80
	0	2	<0.0001	0.82	0.78	0.86
	0	3	<0.0001	0.77	0.74	0.81
	1	2	0.0120	1.07	1.02	1.12
	1	3	0.8124	1.01	0.96	1.05
	2	3	0.0222	0.94	0.90	0.98
B	0	1	0.0012	0.85	0.79	0.92
	0	2	0.0001	0.83	0.76	0.89
	0	3	<0.0001	0.80	0.74	0.86
	1	2	0.4822	0.97	0.89	1.05
	1	3	0.1444	0.93	0.86	1.01
	2	3	0.4451	0.96	0.89	1.04
B*	0	1	<0.0001	0.81	0.79	0.83
	0	2	<0.0001	0.85	0.83	0.87
	0	3	<0.0001	0.82	0.80	0.84
	1	2	0.0019	1.05	1.02	1.08
	1	3	0.2725	1.02	0.99	1.04
	2	3	0.0377	0.97	0.94	0.99
C-1	0	1	<0.0001	0.86	0.84	0.88
	0	2	0.2833	0.98	0.95	1.01
	0	3	0.6949	1.01	0.98	1.04
	1	2	<0.0001	1.14	1.11	1.18
	1	3	<0.0001	1.17	1.14	1.21
	2	3	0.1443	1.03	1.00	1.05
C-2	0	1	<0.0001	0.92	0.90	0.94
	0	2	<0.0001	0.94	0.91	0.96
	0	3	0.0024	0.96	0.93	0.98
	1	2	0.2237	1.02	0.99	1.04
	1	3	0.0102	1.04	1.01	1.06
	2	3	0.1651	1.02	1.00	1.05
D	0	1	<0.0001	0.68	0.66	0.71
	0	2	0.0305	0.95	0.92	0.99
	0	3	0.0019	0.93	0.90	0.97
	1	2	<0.0001	1.40	1.35	1.45
	1	3	<0.0001	1.37	1.32	1.42
	2	3	0.3143	0.98	0.94	1.01

Ratio is the average of one stage over that of the other stage as listed in the previous columns. Stage 0 represents original results. C-1 and C-2 are two analysts from Laboratory C. (*) The analysis removed samples measured at 74,100 and 81,300 as they appeared to be statistical outliers

Table 3
Agreement testing among groups.

Stage	_Stage	*p*-value	Ratio	Lower bound of 90% confidence interval of ratio	Upper bound of 90% confidence interval of ratio
0	1	<0.0001	0.81	0.79	0.82
0	2	<0.0001	0.91	0.89	0.92
0	3	<0.0001	0.89	0.88	0.91
1	2	<0.0001	1.12	1.10	1.14
1	3	<0.0001	1.11	1.09	1.13
2	3	0.2169	0.99	0.97	1.00

After outlier removal using 3 SD rule on the residual from the mixed effect model

single source of standards can improve the agreement between laboratories.

The challenge of comparability is not unique to PK bioanalysis for proteins. Clinical diagnostic assay manufacturers routinely deal with the challenge of comparability and have found that a universal calibrator was part of the solution to improve comparability [8]. In the clinical diagnostic setting, they have additional challenges such as different vendors making assays which contain different anti-

Table 4
Components of variance analysis.

CovParm	Stage I		Stage II		Stage III	
	Variance	Standard deviation	Variance	Standard deviation	Variance	Standard deviation
Laboratory	0.01	0.11	0.01	0.08	0.01	0.12
Residual	0.03	0.18	0.01	0.11	0.01	0.11
Total	0.04	0.21	0.02	0.14	0.03	0.16

bodies and process settings, so theoretically introduction of a single source calibrator in the PK setting could have greater value. While there are similarities between PK and clinical diagnostics, there are also differences between the two situations. PK bioanalysis does not have a third party organization such as National Institute of Standards Technology (NIST) to provide a gold standard or a governing body like College of American Pathologists to provide proficiency testing. Therefore, the first in human study calibrators could be considered the gold standard for subsequent lots. The method for qualifying subsequent lots will create additional debate, because to date the recommendations have been to use the nominal concentration and no correction factors or adjustments have been applied.

Obviously this is a limited dataset of only one test system, and similar experiments will need to be performed on other programs to understand the dimension and magnitude of laboratory discordance. Here we showed a potential problem with ligand binding assays, we designed experiments, analyzed the data to find the source of the problem and provided a recommendation for reducing the discordance between analytical sites. The adoption of stage II approach will have better probability for acceptable results that will allow for data combinability across different laboratories and studies.

Acknowledgement

The authors are grateful to Kendra Atkinson, Chris Wilson, Jim Gorman, Loesje Veyt, Sandra Beyl, Judith Leurs, Kate Watson, Clarinda De Guzman, and Rafiqul Islam for their participation and contributions.

References

[1] J.W. Findlay, W.C. Smith, J.W. Lee, G.D. Nordblom, I. Das, B.S. DeSilva, M.N. Khan, R.R. Bowsher, Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective, J. Pharm. Biomed. Anal. 21 (2000) 1249–1273.
[2] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, R. Tacey, H. Hill, A. Celniker, Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules, Pharm. Res. 20 (2003) 1885–1900.
[3] J. Smolec, B. DeSilva, W. Smith, R. Weiner, M. Kelly, B. Lee, M. Khan, R. Tacey, H. Hill, A. Celniker, V. Shah, R. Bowsher, A. Mire-Sluis, J.W. Findlay, M. Saltarelli, V. Quarmby, D. Lansky, R. Dillard, M. Ullmann, S. Keller, H.T. Karnes, Bioanalytical method validation for macromolecules in support of pharmacokinetic studies, Pharm. Res. 22 (2005) 1425–1431.
[4] C. Beaver, Challenges in Transferring Bioanalytical Methods, Drug Discov. World (2008) 2–6.
[5] International Conference on Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use, Validation of analytical procedures: definitions and terminology, Geneva (1996).
[6] General Chapter <1225>, Validation of compendial methods,United States Pharmacopeia XXIII, National Formulary, XVIII, Rockville, MD, The United States Pharmacopeial Convention, Inc, 1995, pp. 1710–1612.
[7] C.T. Viswanathan, S. Bansal, B. Booth, A.J. DeStefano, M.J. Rose, J. Sailstad, V.P. Shah, J.P. Skelly, P.G. Swann, R. Weiner, Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays, Pharm. Res. 24 (2007) 1962–1973.
[8] C.W. Weykamp, T.J. Penders, K. Miedema, F.A. Muskiet, W. van der Slik, Standardization of glycohemoglobin results and reference values in whole blood studied in 103 laboratories using 20 methods, Clin. Chem. 41 (1995) 82–86.