Journal of Pharmaceutical and Biomedical Analysis 53 (2010) 623–630

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

# Strategies to minimize variability and bias associated with manual pipetting in ligand binding assays to assure data quality of protein therapeutic quantification

Kinnari Pandya [a], Chad A. Ray [b], Laura Brunner [a], Jin Wang [a], Jean W. Lee [a,*], Binodh DeSilva [a]

[a] Pharmacokinetics and Drug Metabolism, Amgen Inc., Thousand Oaks, CA, United States
[b] Radix BioSolutions, Georgetown, TX, United States

ARTICLE INFO

*Article history:*
Received 22 December 2009
Received in revised form 20 April 2010
Accepted 21 April 2010
Available online 28 April 2010

*Keywords:*
Pipetting
Sample dilution
Ligand binding assay
Pharmacokinetics

ABSTRACT

Bioanalytical laboratories require accurate and precise pipetting to assure reproducible and accurate results for reliable data. Two areas where pipetting differences among analysts lead to poor reproducibility are long term stability testing and sample dilution. The purpose of this paper is to illustrate the problems with manual pipetting, describe an automation strategy to mitigate risks associated with manual pipetting, and provide recommendations on a control strategy that properly monitors samples requiring dilutions.

We determined differences among various manual pipetting techniques by analysts within a laboratory. To reduce variability in pipetting, a flexible modular liquid handling script was created on the Hamilton Microlab Star (HMS) to perform sample dilution, pre-treatment and plate loading. The script is capable of handling variable dilution factors. Additionally, two dilution controls were prepared and tested at concentrations of high and mid quality controls (QC). These same dilution controls were incorporated into both pre-study validation and in-study QCs to monitor dilution processing and assay performance.

Variability of manual pipetting among 11 analysts was more negatively biased with increasing dilution. Forward and reverse pipetting delivering different volumes contributed to the discordance. The dilutional bias with manual pipetting was eliminated using the liquid handler. Total error of dilution controls was less than 20%. The in-study pass rate was 100%.

Application of liquid handlers minimizes the variability and bias due to manual pipetting differences among analysts. The incorporation of dilution QCs serves a dual purpose to monitor the dilution process of the samples as well as the binding assay performance.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Ligand binding assays have been used extensively to measure protein therapeutics. These methods are more variable than chromatographic methods that are used to quantify chemical entities [1]. The major contributing source of variation is pipetting technique, which is often performed manually. This problem is further amplified by the additional pipetting required for dilutions of most protein therapeutic study samples into the standard curve range of the binding assay. Variability in manual pipetting has not been investigated for its impact on ligand binding assay used to support protein therapeutic pharmacokinetic (PK) or toxicokinetic (TK) studies.

Publications from a pipette vendor have described optimal pipetting procedures, factors and techniques that influence the accuracy of pipetting [2–4]. The delivered volume can be significantly different due to environmental factors such as temperature [3] of the fluid being pipetted, the delivery method (forward or reverse pipetting) [4], and relative humidity in the ambient environment [2]. Therefore, operators using different methods of pipetting at different conditions could lead to different results.

Long term stability (LTS) testing has been challenging because multiple analysts and standard (STD) preparations are involved over time. To set up the LTS tests, often one analyst prepares stability test samples and at the designated time point another analyst prepares a fresh STD set to be used for the test of the previously prepared stability sample. Higher variability has been observed for the stability samples than that of QCs during pre-study validation. It is not uncommon that results at later test time points would be outside of the *a priori* established acceptance criteria. In some instances the observed values are higher than that of the baseline value for a protein therapeutic that is known to be stable. The method variance prohibits a clear interpretation on the analyte stability since the differences could be related to operator differences and not to changes in analyte stability.

* Corresponding author. Current address: 1 Amgen Center Drive, Thousand Oaks, CA 91320-1799, United States. Tel.: +1 805 447 9463; fax: +1 805 480 3057.
*E-mail address:* jwlee@amgen.com (J.W. Lee).

0731-7085/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2010.04.025

Protein therapeutic quantification often requires very large dilutions. During an inter-laboratory comparison, we observed that higher concentration samples had greater variance. These results indicated that the dilution process could be a source of the increased variation. To address variability issues of the stability testing and sample dilution, we have developed an approach to reduce the variation associated with STD/QC preparations and sample dilution. We use automated liquid handlers for both processes to improve accuracy, precision and reproducibility. In addition, we have incorporated QC at concentrations above the STD curve range to reflect the study samples and determined their pre-study validation accuracy/precision total error, and use these QCs to monitor and accept/reject in-study runs.

Liquid handlers have been widely used for nearly three decades in the pharmaceutical industry in various aspects of drug development. However, it has been challenging to integrate them effectively for bioanalysis of protein therapeutics that support PK/TK studies. One of the challenges is the requirement of large and variable dilution factors depending on the dose and time point collected. In 2006 a paper described an in-house developed software program that fully automated the process for sample preparation using a Tecan Genesis integrated with Watson LIMS [5]. We asked the vendor to write a script for a Hamilton liquid handler and adopted a similar integration strategy by creating a Watson export file to be read by the liquid handler to prepare STDS, QCs and perform dilution at various dilution factors.

The purpose of this manuscript is to describe our findings on variability from manual pipetting, to illustrate the use of automated liquid handlers to successfully reduce this source of variation. We also present a revised monitoring process that accounts for both the sample dilution and binding portion of the immunoassay.

## 2. Experimental

### 2.1. Materials and equipment

The following equipments were used: Spectra Max 340PC plate readers (Molecular Devices, Sunnyvale, CA), ELX-405 plate washers (Biotek, Winooski, VT), Titermix 100 plate shakers (Brinkmann, Westbury, NY), model 2005 incubators (VWR, West Chester, PA), Hamilton Microlab Star (Hamilton Robotics, Reno, NV), and Rainin pipettes (Oakland, CA).

All protein therapeutic standards and immunoassay reagents were produced and prepared by Amgen Inc. (Thousand Oaks, CA). Sera from non-human primates and humans were obtained from Bioreclamation (Hicksville, NY). All solutions were stored at 2–8 °C except for the 20× wash buffer and Dulbecco's Phosphate buffer saline which were stored at ambient room temperature (ART). The 20× wash buffer (catalog number 50-63-00) was purchased from KPL, Inc. (Gaithersburg, MD). Dulbecco's Phosphate buffer saline (without $CaCl_2$ and $MgCl_2$) was purchased from Invitrogen (Carlsbad, CA). Neutravidin was from R&D Systems (Minneapolis, MN) and the hydrogen peroxide and tetramethyl benzene (TMB) substrate solutions were from BioFX (Owings Mills, MD).

### 2.2. ELISA procedure

The data presented in this manuscript come from several enzyme linked immunoassays (ELISA) that were used to quantify multiple protein therapeutics. These assays were based on a sandwich immunoassay approach. STD, test samples and QC in 100% serum, were loaded into the 96-well plate after being diluted with the assay buffer at the minimum required dilution of the specified method. The therapeutic protein was captured by an antibody that had been passively adsorbed on the plate. After unbound materials were removed by washing the wells, horseradish peroxidase tagged anti-therapeutic antibody was added for detection of the captured analyte. After another wash step, an enzyme substrate solution was added to develop the colorimetric signal, which was proportional to the amount of analyte bound. The color development was stopped and the optical density signal was measured at 450 nm with reference to 650 nm. The Watson version 7.0.0.01 LIMS using 4- or 5-parameter logistic regression model converted the signals to concentrations for the test samples and QCs comparing to a concurrently analyzed STD curve.

### 2.3. Long term stability test

LTS experiments were conducted to support the validation of a method intended for the determination of a protein therapeutic over the time span required for sample storage. All STD, QCs and LTS samples were prepared manually. Bulk LTS samples were prepared in 100% serum at concentrations of 2250 ng/mL for high QC (HQC) and 150 ng/mL for low QC (LQC). These bulk LTS samples were assayed by analyst 1 within 3 days of preparation and run against a set of STD and QC prepared and stored overnight at −70 °C. The results were used as the baseline to be compared with subsequent LTS time point determinations to assess analyte stability. A fresh set of STDs and QCs was prepared and stored overnight at −70 °C prior each LTS time point by the analyst performing the stability test. In addition to establishing the baseline, analyst 1 assayed the LTS samples for the 2 and 4 weeks time points. The 3-month LTS assessment was performed by analyst 2. Because of the discordant results, the 3-month LTS time point was repeated by both analysts.

### 2.4. Manual dilution variability among analysts

Eleven analysts participated in the test to determine relative dilution accuracy within our laboratory. Test samples were diluted by each analyst and a Hamilton liquid handler $10–10^6$-fold in 100% serum to obtain a final concentration within the range of the STD curve: those at 10 and 100 μg/mL were diluted to 10- and 100-fold, those at 1000 μg/mL to $10^3$- and $10^4$-fold, and those at 20,200 μg/mL to $10^5$- and $10^6$-fold. To reduce the effects of plate-to-plate variation and operator-to-operator variation, all dilutions from an analyst or the liquid handler were assayed on one plate, and all procedures after dilution were performed by a single analyst. Data assessment was performed using JMP 7.0 (JMP, Cary, NC).

### 2.5. Variability due to pipetting technique

Following the manual dilution test, a comparison of pipetting techniques was conducted. The experiment followed a full factorial design to test all combinations of STD preparer, STD curve delivery, QC preparer, and QC delivery. Two analysts prepared STD curves and QCs using two different pipetting techniques. For one set of STD and QC, a forward pipetting technique was used and for the other set a reverse pipetting was used. Fig. 1 illustrates the procedural differences between forward and reverse pipetting. In order to control the inter-operator variability all assay steps except for STD and QC preparation were carried out only by analyst 1 (Table 1). A total of 16 combinations were tested. For each condition, QCs were assayed in duplicate. The interpolated value was compared against the nominal concentration and the mean % bias for each condition was computed. In addition, pipetting accuracy of each technique was determined by the Artel pipetting calibration system (Westbrook, ME) using two dye solutions. Three representative pipettes (P-10, P-200, and P-1000 from Rainin, Oakland, CA) were tested at 5 different volumes. The volumes were selected because they represent the most common volumes used for sample dilution (10, 100, 200, 500, and 900 μL). A total of 5 replicates at each volume were

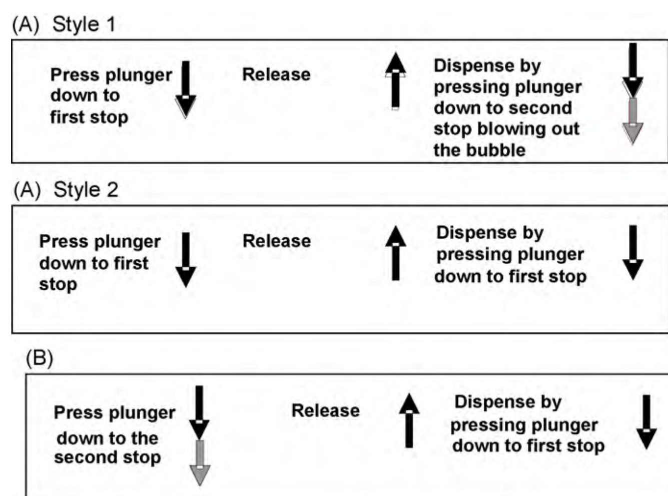


**Fig. 1.** Procedural differences between forward and reverse pipetting. (A) Two styles of forward pipetting: (1) The operator presses the pipette plunger down to the first stop, releases to aspirate and dispenses the liquid by pressing the plunger pass the second stop to blow out any air. (2) The operator presses pipette plunger down to the first stop, releases to aspirate and dispenses the liquid by pressing the plunger down to the first stop. (B) Reverse pipetting style: the operator presses pipette plunger down to the second stop, releases to aspirate and then dispenses the liquid by pressing the plunger to the first stop.

tested using both the forward and the reverse pipetting technique. The target volume was compared against the observed volume and the % bias was computed.

### 2.6. Robotic liquid handling script

A custom liquid handling script was developed by Hamilton Robotics Company (Reno, NV) to dilute samples up to $10^6$-fold, add the diluted samples into wells containing assay buffer (pre-treatment), and load the pretreated sample onto the assay plate. The script was written to fit our laboratory's modular approach of automation, so that the operator could choose any combination of steps for flexible execution of any portion of the process [6]. Fig. 2 depicts the process flow of an analyst using the script: a sample dilution work list created in Watson was saved as a comma delimited (.csv) file. Upon script execution, the user was prompted to import the .csv file that contained sample information of the dilution factors and the final destination on the assay plate. Based on the desired parameters of sample dilution, pre-treatment and/or plate loading, the user specified sample tube type, number of pre-treatment steps, sample and diluent volumes for pre-treatment and final volume to assay plate. Following user inputs, the script performed the steps and used the information found in the .csv file to properly dilute the samples and load in the correct location. All of the subsequent immunoassay steps were then performed manually by the analyst.

**Table 1**
Factorial design to test variability from manual pipetting.

| Assay plate | Region[a] | Operator | Technique used for STDs Prep | Technique used for QCs[b] Prep |
|---|---|---|---|---|
| 1 | 1 | 1 | Reverse | Reverse |
| | 2 | 1 | Forward | Forward |
| 2 | 1 | 2 | Reverse | Reverse |
| | 2 | 2 | Forward | Forward |
| 3 | 1 | 1 | Forward | Forward |
| | 2 | 2 | Forward | Forward |
| 4 | 1 | 1 | Reverse | Reverse |
| | 2 | 2 | Reverse | Reverse |
| 5 | 1 | 1 | Reverse | Reverse |
| | 2 | 2 | Forward | Forward |
| 6 | 1 | 1 | Forward | Forward |
| | 2 | 2 | Reverse | Reverse |

STD and QC were prepared by analyst 1 and 2 using reverse and forward pipetting technique. A total of 16 combinations were tested across 6 immunoassay plates.

[a] Region 1 and 2 reflects columns 1–6 and columns 7–12, respectively in a 96-well microtiter assay plate.

[b] QCs from each assay plate were compared against both the STD curves run on the same assay plate.

### 2.7. Liquid handler performance assessment

The robotics performance using the liquid handling script was verified with spiked and study samples. Five different dilution factors ranging from 1- to 10,000-fold were tested with spiked samples. Concentrations of 0.197, 1.97, 19.7, 197, and 1970 μg/mL were diluted 1 (no dilution), 10-, $10^2$-, $10^3$- and $10^4$-fold in 100% serum. These experiments generated 10 values per dilution factor and the analysis was conducted over 4 assays.

### 2.8. Revised QC strategy

As study samples of protein therapeutic require high dilution into the standard curve range, we revised the QC levels to reflect the expected concentrations of study samples to include the control of the dilution process. Two QCs at concentrations above of the STD range were added to the traditional accuracy and precision assessment to reflect the process of sample dilution. For the specified assay, the high dilutional QC (DHQC) and mid dilutional QC (DMQC) were diluted into the high and mid range of the standard curve before assay, respectively. A third QC at low concentration (LQC) was assayed without dilution. This approach incorporated the error components associated with both sample dilution and immunoassay. During method validation, the intra- and inter-assay accuracy and precision were evaluated with seven levels of Validation Samples/QC over 9 independent assays. The concentrations of neat QC samples were 100 at the low limit of quantification (LLOQ), 300 (LQC), 2000 (MQC), 3200 (HQC) and 4500 ng/mL at the upper limit of quantification (ULOQ). The dilutional QC concentrations were 32,000 (DHQC, diluted 10-fold) and 2,000,000 ng/mL (DMQC, diluted 1000-fold). The dilutions were prepared by the liquid handling script. Each validation run consists of STD and 4 replicates of validation sample at each concentration.

Bias, impression and total error of the accuracy and precision experiments were calculated. The run acceptance criteria were established using the 4-6-*X* rule with at least one QC must be acceptable at each level. "*X*" was determined to be 20% based on the total errors of LQC, DMQC and DHQC, instead of the neat QCs. These acceptance criteria were applied to a clinical study using the specified assay. The longitudinal QC results were evaluated for the passing rate and overall error.

## 3. Results

### 3.1. Long term stability test

LTS was tested over 3 months by two analysts using manual pipetting. Analyst 1 prepared the stability samples, assayed the baseline stability samples and two subsequent stability time point samples. Analyst 2 assayed the 3-month time point samples. Fig. 3 shows that the 3-month results biased high and were unacceptable, indicating stability problem. These results were repeated by the same analyst confirming the high bias. In order to rule out operator differences, analyst 1 also assayed the 3-month test samples; however, the results were within the acceptance criteria. Upon further

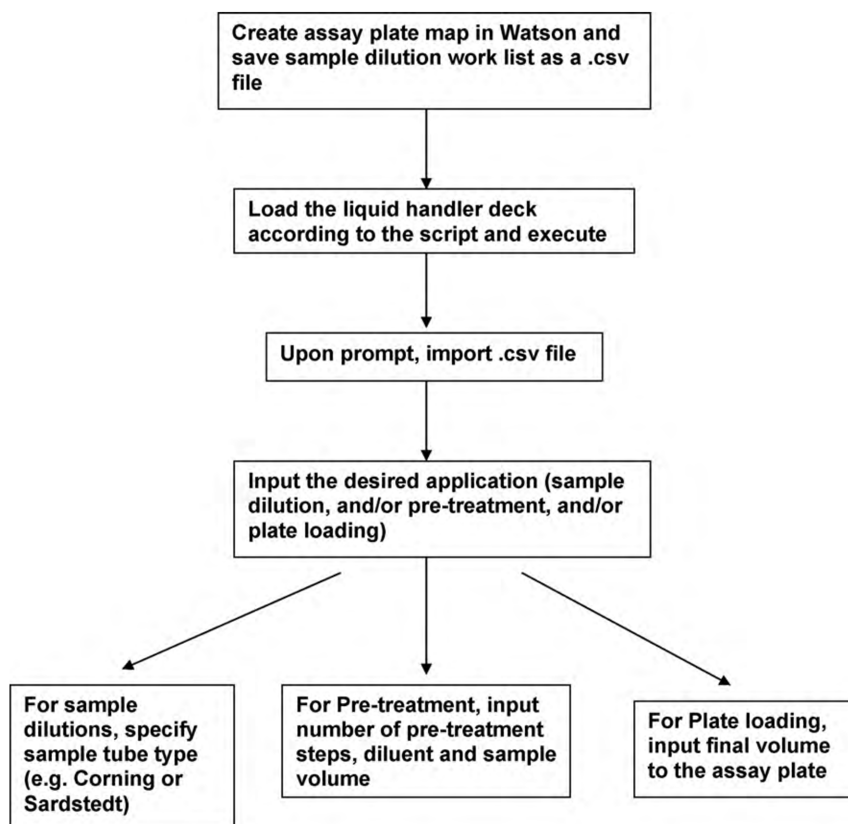


**Fig. 2.** Work flow of modular script using a liquid handler. A sample dilution work list created in Watson is saved as a comma separated variable (.csv) file. Upon script execution, the user is prompted to import the .csv file containing sample information (the dilution factors and position on the assay plate). The user specifies sample tube type, pre-treatment steps, sample and diluent volumes for pre-treatment, and volume to be transferred to assay plate. The script then performs the steps defined by the user. The subsequent ELISA assay steps are usually performed manually by the analyst.

discussion, it was found that analysts 1 and 2 differed in their practice of forward or reverse pipetting techniques during the manual preparation of the fresh set of STDs that was used for the stability sample calculation (Fig. 1), which may have contributed to the disagreement of LTS results.

### 3.2. Manual test of dilution factors

Within a laboratory, analysts may have different pipetting techniques as a result of variable training and previous experience. Eleven analysts in our laboratory were randomly selected to assess the inter-operator variation on manual sample dilution. Dilutions of 10–$10^6$-fold were chosen to reflect dilutions used in both preclinical and clinical studies. Fig. 4 shows an overall increase in negative bias with increasing dilution factor for most of the analysts with the exception of analysts 1, 9, and 11; and the liquid handler. More important, the inconsistencies among the analysts were high to cause concern. Three distinct patterns were observed: (1) a consistent negative bias, (2) a consistent positive bias, and (3) no observable trend. In addition to the different pipetting techniques, inappropriate mixing, loss of transferring volume, and carry-over by not changing pipette tips during transfer may also contribute to the discordance. An obvious risk from the disagreement among the analysts could be an inability to reconcile discrepant data from experiments such as incurred sample reanalysis (ISR) [7], stability assessments and cross-study comparisons.

### 3.3. Effect of pipetting technique on accuracy

We investigated differences in pipetting techniques among analysts that could contribute to different volumes being delivered, resulting in variable values. We performed a full factorial design experiment to assess the impact of two prevalent pipetting techniques (reverse and forward) on accuracy using STD and QC samples prepared by two operators. Fig. 5 shows three key observations: (1) QC prepared by forward pipetting interpolated against the reverse pipetted STDs resulted in an overall negative bias with one exception (marked by an asterisk). This could be attributed to an increased volume delivered by reverse pipetted STD vs. the lower

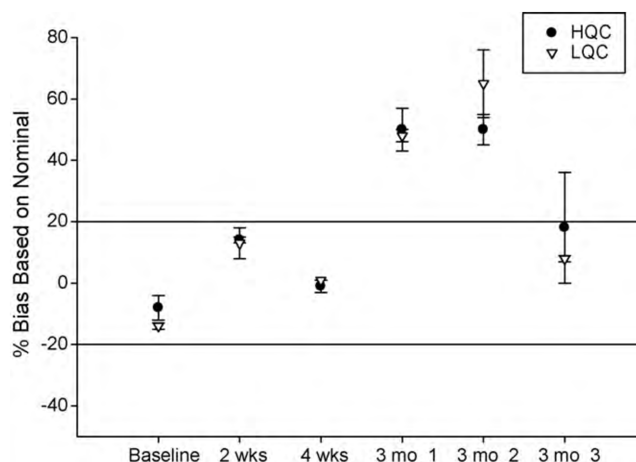


**Fig. 3.** Long term stability test. LTS tests were performed at 2 and 4 weeks (wks), and 3 months (mo) time points. Baseline: stability samples assayed within 3 days of preparation. Stability tests at 3mo_1 and 3mo_2 were performed by analyst 2. All other tests were performed by analyst 1. Solid lines: acceptance boundaries of ± 20% bias from the nominal. Solid circles: HQC; triangles: LQC.

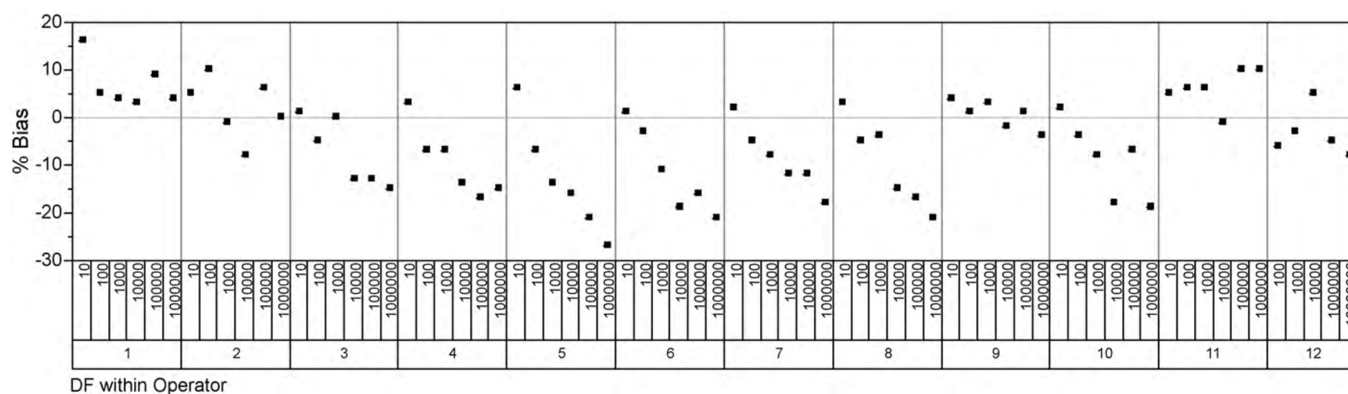


**Fig. 4.** Dilution variation with manual pipetting. Eleven analysts (designated as 1–11) manually diluted QCs using dilution factors ranging from 10 to $10^6$. Panel 12: dilution performed by automation. The assay was carried out by the same operator on the diluted samples from the various sources.

volume delivered by the forward pipetted QC. No explanation was identified for the exception. (2) QC prepared by reverse pipetting interpolated against forward pipetted STD showed an overall positive bias. This could be attributed to a decreased volume delivered by the forward pipetted STD vs. the increased volume delivered by reverse pipetted QC. (3) The mean % bias for each analyst using the same pipetting technique for the STD and QC preparation was well within the assay acceptance criteria. Forward pipetting reflected more variation when compared to reverse pipetting. This could be attributed to the two different styles of forward pipetting compared to only one style for reverse pipetting (Fig. 1). In addition to the differences in the pipetting styles, another variable of pre-wetting the tips in both reverse and forward pipetting and viscosity of the liquid can also contribute to the variations.

In addition to immunoassay, the Artel pipette calibration system was used to pipette a dye into another dye based sample solution to test pipetting accuracy on five pipetting volumes from 10 to 900 μL. Table 2 shows that the mean bias for forward pipetting range was 0.4–1.9% while that of reverse pipetting was 1.6–3.6%.

### 3.4. Liquid handler performance assessment

To minimize pipetting variability, a strategy using robotic liquid handler to replace the manual pipetting has been adopted in our laboratory. Scripts were written to execute various pipetting processes for STD and QC preparations and sample dilution. The assay performance was evaluated on spiked samples and study samples. The first experiment evaluated five dilution factors from 1- to 10,000-fold. Ten samples per dilution factor were processed and assayed on 4 different immunoassay plates. The mean recovery across all dilution factors ranged from 88% to 92%. The slope across dilution factors was 0.0003 using linear regression (Fig. 6).

**Table 2**
Test of forward and reverse pipetting by Artel pipette calibration system.

| Target volume (μL) | Forward pipetting Mean (μL) $n=5$ | % bias | % CV | Reverse pipetting Mean (μL) | % bias | % CV |
|---|---|---|---|---|---|---|
| 10 | 10.19 | 1.9 | 0.1 | 10.36 | 3.6 | 0.6 |
| 100 | 100.36 | 0.4 | 0.2 | 102.73 | 2.7 | 0.2 |
| 200 | 202.50 | 1.3 | 0.6 | 204.92 | 2.5 | 0.1 |
| 500 | 503.58 | 0.7 | 0.2 | 512.48 | 2.5 | 0.3 |
| 900 | 903.44 | 0.4 | 0.1 | 914.60 | 1.6 | 0.5 |

### 3.5. Revised QC strategy

To mimic study samples that required dilution, two QCs at concentrations above the highest STD calibrator were included to reflect the diluting process of the sample. Data from a specified study are presented here as illustration: The first QC concentration was designed to mimic the predicted $C_{max}$ concentration of 2,000,000 ng/mL (DMQC) and was diluted 1000× to the middle of

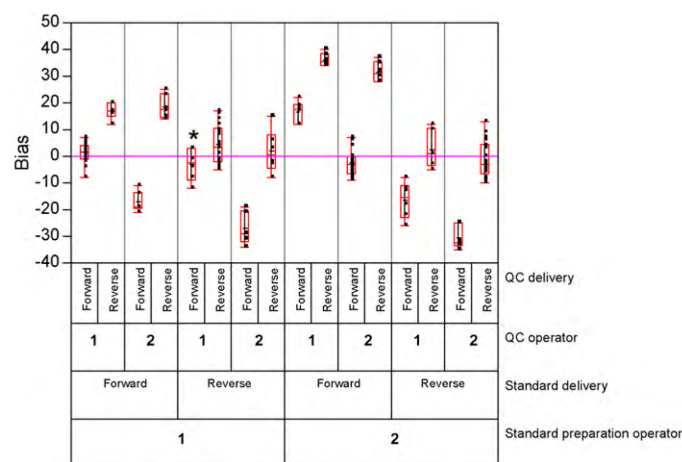


**Fig. 5.** Assay bias from different manual pipetting techniques. Two analysts (designated as 1 and 2) manually prepared two sets each of STD and QC using the forward and reverse pipetting techniques.

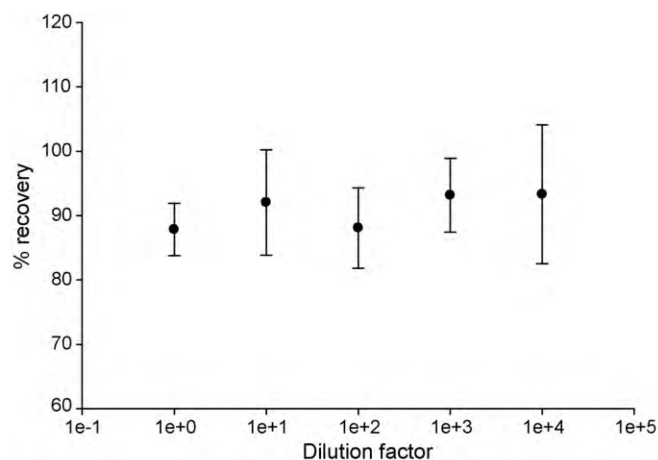


**Fig. 6.** Liquid handler dilutional linearity assessment. Five dilution factors ranged from 1 to 10,000 were tested using the liquid handler script. The error bars are standard error of the mean. Ten dilution preparations were made on multiple days and assayed on 4 different plates. The mean recovery was 91% the slope across the dilution factors was 0.0003 using linear regression.

**Table 3**
Intra- and inter-assay accuracy and precision of neat and diluted validation samples/QC.

| QC level | LLOQ | LQC | MQC | DMQC | HQC | DHQC | ULOQ |
|---|---|---|---|---|---|---|---|
| Nominal conc. (ng/ml) | 100 | 300 | 2000 | 2,000,000 | 3200 | 3200 | 4500 |
| *N* | 36 | 36 | 36 | 36 | 36 | 36 | 36 |
| Accuracy (mean bias) | −0.8 | 1.2 | −1.5 | −0.4 | −1.3 | −2.8 | −7.8 |
| Precision | 10.9 | 7.8 | 7.2 | 10.9 | 12.7 | 14.8 | 16.9 |
| Total error | 11.7 | 9.0 | 8.7 | 11.2 | 14.0 | 17.5 | 24.6 |

Validation samples at each level were assayed in 4 replicates in each of 9 assays. All dilutions were performed with a liquid handler. The dilutional QCs, DHQC and DMQC were diluted 10- and 1000-fold, respectively, prior to the analysis. Neat samples were LLOQ, LQC, MQC, HQC, and ULOQ.

the STD curve. The second QC was designed at 32,000 ng/mL (DHQC) to be diluted 10× to represent the other high concentration samples. After dilution, their concentrations coincided with the neat MQC and HQC to allow performance assessment at those regions of the dynamic range. The method validation data showed that total errors from DMQC and DHQC were 3.5% and 2.6% higher than those of the neat QCs, respectively (Table 3). The third QC was assayed neat at the LQC level. The total error data were 20% at all validation sample levels except 25% at the ULOQ, meeting the intended purpose of the method.

The assay acceptance criteria applied the 4-6-*X* rule, with *X* (20%) based on neat and dilution QC data from the accuracy and precision experiments during pre-study validation. The QC strategy of incorporating both neat and dilutions QC was applied to a clinical study. Six QCs (2 replicates at each of DHQC, DMQC and LQC) were included in every run. At least 4 QCs and one at each level had to be within the *a priori* acceptance criteria [8]. The in-study total error was 15% and 14% for the DMQC and DHQC, respectively. Two out of 17 assays (run ID's 13 and 14) showed an unusually high bias for one of the two LQC replicates, which were shown to be outliers by the Dixon-Q test [9]. The total error with and without the outliers for the LQC was 116% and 5%, respectively. A 100% pass rate was achieved for 17 runs. The total error observed during pre-study validation was predictive of the in-study performance (Table 3). The control charts in Fig. 7A and B show that 91% of the diluted QC and 94% of the LQC were within 20% of the nominal concentration.

## 4. Discussion

This paper illustrates that variation in manual pipetting can impact the final results for ligand binding assays. The processes where manual pipetting can have a great impact are STD and QC preparations and sample dilution. Implementation of a standardized manual pipetting program is not an easy task that would involve consensus-building, thorough training, change of habits, and continual monitoring. An obvious option to mitigate the risk of manual pipetting was to remove the human component by using automated liquid handlers for these two crucial processes. In addition, we revised the control strategy to include QC samples that require dilution in the run acceptance criteria. These dilution controls were evaluated in pre-study validation and in-study performance monitoring. We believe that the combination of these two approaches has reduced the risk associated with manual pipetting and provided realistic control data that better reflect the study samples.

Our awareness of the variation associated with manual pipetting came from LTS experiments. In those experiments, it was common for multiple analysts to prepare STD curves to measure the stability samples. The experimental design with multiple analysts created a situation where the variables (analyst and time) were confounded. As a result, it was not possible to distinguish stability problems from differences in the performance of the operator. Fig. 3 demonstrated that analyst 1 and analyst 2 generated different results for the 3-month time point when evaluating the same stability samples. The only difference was the fresh STD curve prepared by each analyst conducting the stability test at the specified storage time. The data generated by analyst 1 were within the *a priori* acceptance criteria indicating that the test material was stable. In order to further investigate these results, a series of experiments were conducted to estimate the inter-operator variation and the pipetting factors that contribute to the differences.

A sample dilution experiment was performed by 11 analysts and the results showed an overall trend of increasing negative bias with increasing dilution factor. The data also demonstrated disparity among some of the analysts. For example if analyst 5 and 11 performed analysis on the same sample, the results could be differ-

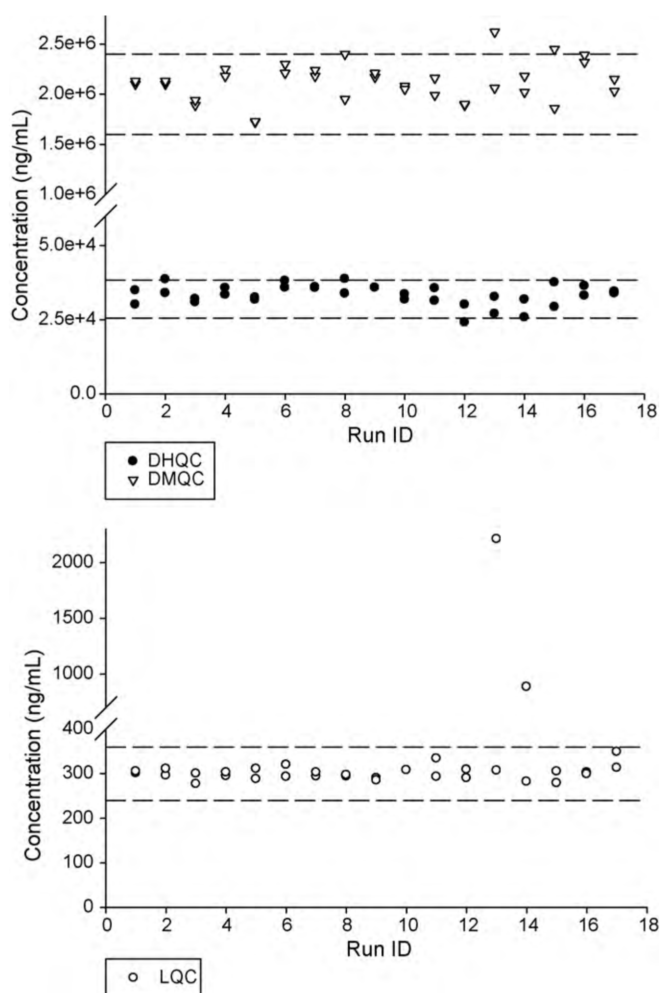


**Fig. 7.** In-study quality control performance. Dashed lines: acceptance boundaries of ±20% Bias. All 17 assays were accepted according to the 4-6-20 criteria for this method. (A) Diluted QCs, solid circles: DHQC at 32 μg/mL, diluted 10-fold; triangles: DMQC at 2000 μg/mL, diluted 1000-fold. (B) Neat QC, circles: LQC at 300 ng/mL.

ent by more than 35% (Fig. 4). Training may be a contributing factor in the pattern associated with dilutions. Specifically in this case analysts 6, 7, and 8 were trained by one person, whereas 1 and 9 by another. The patterns across training groups were very different. One hypothesis was a certain pipetting style may be used in one group compared to the other. For example some analysts primarily used reverse pipetting and others used a combination of forward and reverse pipetting. The Artel experiment showed that reverse pipetting delivered more volume than that of forward pipetting at all test volumes. As expected, low volume pipetting showed higher bias than those of larger volumes, however, the delivered volumes were consistently lower for forward than reverse pipetting by about 2% at all volumes. The pipetting accuracy of a dye in buffer would be different from that of serum. In order to test the effect of pipetting techniques on STD/QC preparations and sample dilutions, a specific immunoassay in serum was performed in a factorial designed experiment.

The immunoassay data also showed higher values from reverse pipetting compared to forward pipetting which could impact the immunoassay results. Reverse delivered QCs usually had a positive bias against the forward delivered STDs, while the forward delivered QCs had negative bias against the reverse delivered STDs (Fig. 5). Overall, these results indicate that consistent pipetting technique is necessary to reduce bias. The aggregate data indicated that analysts prepared sample dilutions differently, the pipetting styles were different across groups of analysts, and volume differences occurred depending on the style.

It should be noted that the immunoassays selected for these experiments were robust assays. The accuracy and precision experiments were conducted over multiple days by multiple operators using multiple STD and QC preparations and multiple instruments. Each of the assays described had maximum total error less than 20% (data not shown). By selecting these assays that were validated for robustness, we increased the probability that the observed effect was due to the test variable of pipetting and not immunoassay variation. However, other factors besides forward and reverse pipetting were not further deciphered for their effect on dilution variability. These differences could include: pre-wetting, tip touching to the side, wiping the tips, vortexing speed, vortexing duration, sample density, viscosity, and speed of aspiration and dispensing. As a result of all of these potential factors, we made the decision to minimize the human variability by incorporating automated liquid handlers.

For years, the potential productivity benefits of liquid handlers have been recognized, but the quality gains have not been widely discussed, especially in the bioanalysis of protein therapeutics. One expectation was that liquid handlers would provide more consistency in pipetting and mixing than the manual processes. We worked with a liquid handler manufacturer to develop a modular script that was capable of reading a LIMS output file for diluting, pre-treating, and plating samples. The resulting performance from the script demonstrated elimination of bias and variability due to dilution, with recoveries of 88–92% and almost a zero slope across the dilution factors (Fig. 6). The experiment was designed to dilute different spiked samples to the same region of the STD curve, so that the variability would be mainly attributed to dilution and not to differences in variation associated with STD regions. Additional dilutional linearity experiments showed that a single sample can be diluted across the entire range of the STD curve accurately (data not shown).

The liquid handler results across dilution factors were more consistent than the majority of the manually prepared dilutions. The reliability of the script and the quality advantage of using liquid handlers over manual processes were demonstrated. This level of consistency is essential for comparability across studies and analytical sites. It is important to note that a considerable amount of time was spent adjusting and optimizing parameters such as the aspiration and dispensing speed, the number of mixing steps, and the volume of mixing. Proper optimization was needed to avoid upward or downward trends across a range of dilutions. The time spent was worthwhile for a robust script to cover the dynamic applications across various methods, dilution factors, studies, and dose groups. The end product is user-friendly, which requires minimal input from the operator.

In order to assure proper dilution and assay performance, a strategy was applied to incorporate dilution QCs into pre-study method validation and run acceptance. We selected two QC concentrations above the upper range of the STD curve. The DHQC level was 10× the HQC and the DMQC was 1000× the MQC. The QC concentrations reflected those of the study samples, and facilitated to monitor both the dilution process and ELISA performance. The procedural change was minimal in preparing QCs at 10× and 1000× the usual HQC and MQC, respectively. The same 4-6-*X* rule was applied for run acceptance or rejection.

The method validation accuracy and precision data provided a direct comparison of the dilution QC with neat samples to determine the error associated with the dilution processing. To our knowledge, this is the first attempt in the pre-study validation to dissect out the total error caused by the dilution processing in addition to the errors due to other immunoassay steps. Our results indicated that dilution QCs had an increased total error (imprecision plus bias) of ∼3% mainly attributed to the imprecision. The similar amount of errors from the 10× and 1000× dilution confirms that the dilution process using the robotic scripts is independent of dilution factor and concentrations. The total error of the dilution QC and LQC was used to set the in-study QC acceptance criteria. The in-study data performance has been comparable to the pre-study predictions.

This approach was simpler than the option of using additional dilution QC as a sample processing control. If the dilution QC is outside the acceptable range, then the diluted samples with that specific dilution factor would be rejected and those samples repeated. This sample processing approach requires cherry picking to identify failed samples to be reanalyzed, which can be logistically challenging. However, using the 4-6-*X* rule, the entire assay plate will be repeated upon QC failure, thereby eliminating the need for cherry picking.

The risk mitigation strategy described in this manuscript is congruent with an initiative led by pharmaceutical scientists, technology providers, and contract research organization scientists. The initiative referred to as the 21st Century Bioanalytical Laboratory plans to build on successful examples like this one to improve efficiency and increase quality through innovation. One of the goals of this initiative is to remove restrictions of custom liquid handling scripts like the one described here to be broadly available. This will bring more harmonization across labs.

## 5. Conclusion

Manual pipetting differences among individuals can lead to differences in final immunoassay results. Operator training within a laboratory to assure consistent pipetting technique would be one way to decrease data variability and bias. However, a better approach is to identify the critical processing points and remove the human error component by using automation. We have optimized the use of liquid handlers to reduce the inter-operator variation in STD/QC preparations and sample dilutions and increased data reliability. Considerable amount of time was spent on the script writing, feasibility test and iterative optimization. The effort was rewarded by the flexibility, user-friendliness and successful implementation to various analytes. In addition, automation can be used

as a common standard for conformance tests if manual pipetting has to be used, such as in the case of outsourcing to a laboratory without automation or pipetting samples of limited volume.

To monitor performance of both the sample processing and binding assay, we included dilution QCs in pre- and in-study method validation. This approach has been effective for performance control of ligand binding assays supporting protein therapeutic PK and TK studies that require multiple dilutions. We have implemented the use of automated liquid handlers to minimize risk of data variability due to manual pipetting and a revised QC approach to monitor both sample dilution and assay performance to assure reliable data. The combined effort will improve data agreement in experiments such as long term stability sample testing, ISR and cross-study comparisons.

## Acknowledgements

We would like to acknowledge contributions from Dom Calamba, Beth Johnson, Chris Macaraeg, Jessica Manlongat, Robert Ortiz, Vimal Patel, Ramak Pourvasei, Amir Sharifi, and Jennifer Tsoi.

## References

[1] M. Kelley, B. DeSilva, Key elements of bioanalytical method validation for macromolecules, AAPS J. 9 (2007) E156–E163.
[2] D. Rumery, B. Carle, Extreme pipetting IV: it's not the heat, it's the. . .pipettes perform exceptionally well in high humidity, Pharm. Formulation Qual. February/March (2008) 40–43.
[3] G. Rodrigue, D. Rumery, Extreme pipetting II: the effect of fluid temperature on data integrity, Pharm. Formulation Qual. September (2007) 46–48.
[4] R. Curtis, Minimizing liquid delivery risk pipettes as sources of error, Am. Lab. News 39 (2007) 8.
[5] H. Gu, S. Unger, Y. Deng, Automated Tecan programming for bioanalytical sample preparation with EZTecan, Assay Drug Dev. Technol. 4 (2006) 721–733.
[6] K. Pandya, C. Ray, C. Macaraeg, J.W. Lee, B. DeSilva, Automating the dilution and pre-treatment of pharmacokinetic samples for ligand binding assays: a key to reproducibility and efficiency, in: AAPS National Biotech Conference, Toronto, Canada, 2008.
[7] M.L. Rocci Jr., V. Devanarayan, D.B. Haughey, P. Jardieu, Confirmatory reanalysis of incurred bioanalytical samples, AAPS J. 9 (2007) E336–E343.
[8] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, R. Tracey, 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.
[9] R.B. Dean, W.J. Dixon, Simplified statistics for small numbers of observations, Anal. Chem. 23 (1951) 636–638.