Design of experiment in assessing robustness and for qualification of a cell-based potency assay

Robert Wilson^{*,1}

¹ Piramal Pharma Solutions, Analytical Development, Earls Road, Grangemouth, Stirlingshire FK3 8XG, United Kingdom

*Author for correspondence: Tel.: +44 1324 494206; Robert.Wilson@piramal.com

Aim: This paper describes the salient points in establishing robustness of the cell-based potency assay, with reference to phase-specific qualification. **Materials & methods:** The methodology was qualified in a thaw for use format utilizing standard parameters. At a later stage, the assay was further developed, and concomitant with requalification the robustness was assessed by fractional factorial design of experiment, the center point data being used for qualification. **Results:** The methodology was qualified with a precision of 8.5% and mean bias of 3.6%. The robustness data generated a similar precision and accuracy/bias to the qualification data. Assay factors were optimized to be within the design space model of the design of experiment. **Conclusion:** The methodology was qualified and was robust with regard to assay parameters tested.

First draft submitted: 23 October 2017; Accepted for publication: 1 March 2018; Published online: 17 May 2018

Keywords: antibody–drug conjugate • bioassay • parallel curve analysis (parallelism) • potency • relative potency • similarity

A key milestone in bioassay development is the execution of robustness studies, which are usually conducted prior to validation and the testing of commercial product [1]. It is important that the bioassay is sufficiently robust to determine manufacturing consistency and to achieve this a bioassay may require further development and refinement throughout a product's life cycle. Design of Experiment (DoE) is a tool that can be used to systemically evaluate multiple variables and can also provide an assessment of the impact from the interaction of related experimental conditions [2]. DoE has been applied to cell-based bioassay robustness and qualification studies [3–5], the ultimate aim of such studies is to define the 'design space' to ensure that the bioassay methodology is fit to support manufacture and ongoing stability [1].

DoE or a Quality by Design (QbD) approach to analytical validation is increasingly being advocated for bioassays [1,2]. Indeed, United States Pharmacopeia (USP) Chapter <1033> Biological Assay Validation provides an example of a three-factor DoE approach to validation [6]. The factors, in this nested design, were analyst, cell preparation and reagent lot, which represent key variables that would change over the product's life cycle. In the USP approach, individual results are used to determine the format that has acceptable precision and method capability (Cpm) that will meet the release specification and to mitigate the risk of generating an Out Of Specification (OOS) [6].

We have previously used a two factor (operator, cell bank) design for validation of a bioassay [7]. In this paper, we report the qualification of a cytotoxicity bioassay (operator, cell bank) using fixed assay conditions and the use of DoE to establish the robustness of the bioassay 'design space'. These studies utilized five mock potency samples to establish the precision, accuracy/bias, linearity and range of the methodology and were used to assess the requirements for validation of the bioassay methodology using DoE. The robustness of the methodology was assessed and confirmed using Minitab and Design expert software. This approach, in which the assessment of robustness is coupled to qualification, provided a larger control dataset giving greater assurance to where trends occur in the DoE and when these are significantly different. The normalizing of the mock potency data to 100%

provided additional power to the analysis allowing for optimal assay condition to be established that would minimize inaccuracy in the methodology at the limits of the optimal assay conditions.

Methodology & experimental design

The cell-based assay, in a thaw-for-use format, was developed to measure the specific cytotoxic activity of an antibody–drug conjugate toward a cell line expressing the target antigen. Cells were seeded into three 96-well plates at a nominal density of 3000 cells per well, and within 4 h an 11-point dose dilution series of drug was added to the wells. The activity of the drug was determined after 4 days exposure using Alamar blue as metabolic reagent.

The methodology utilized optimized assay conditions, not only the number of cells per well but also recovery period prior to adding drug, the duration of drug exposure and Alamar blue development time. The objective of the current study was to assess the robustness of the methodology and to qualify the methodology for determining the relative potency of the antibody–drug conjugate.

There were eight factors (cell bank, cell number, dose dilution series, cell recovery time, drug exposure time, Alamar blue concentration, Alamar blue development time and Fetal Bovine Serum (FBS) concentration) considered to be important in assessing the robustness of the methodology. The goal of the experimental design was to test all of the variables, testing for interactions and if required allowing for parameter optimization. With eight factors, the fractional factorial design becomes very large and difficult to manage and a different approach was taken to incorporate all eight factors into the study.

Alamar blue is a nondestructive metabolic reagent and at each event assay plates were read at both the upper (5 h) and lower limit (3 h) time points generating two sets of data. Five mock potency samples (50, 71, 100, 141 and 200%) were independently prepared using either a 2.7-fold or threefold dilution series and analyzed at each event. Analyzing these two dilution series provided sample replication at events and was used to assess the robustness of the dose dilution series. The remaining six factors were studied in a fractional factorial design where cell bank was a categorical variable (see Supplementary Table 1). Four center points were incorporated into the design and a further four center points were generated in the assessment of environmental factors (temperature and % CO₂). Two analysts conducted a similar number of events.

Assay qualification

The primary assumption underlying potency assays is that test sample will act as a dilution of the reference standard preparation and that mock potency samples are able to generate an increase or decrease in response proportional to the expected concentration (see Supplementary Figure 1 for a typical dose response and for the purpose of this manuscript, the parallel line analysis [PLA] and relative potency calculations are presented as a single calculation). It is becoming increasingly important to incorporate factors which may be a source of assay variability (i.e., plate location, operator, day, etc.) into validation and the use of experimental design is now advocated for bioassay validation [5]. A multioperator design, conducted over several days, with replication at each event will provide estimates of precision (intermediate precision), accuracy/bias, linearity and the range of the methodology [5].

It is generally recognized that potency validation/qualification should use independent dilutions. A 1.41-fold geometric dilution series gives five mock potency samples at 200, 141, 100, 71 and 50% potency. The activity (relative potency) is determined relative to the standard preparation (100%), following PLA of the nonlinear dose response curves [4–7]. It should be recognized that relative potency data are not normally distributed and log transformation provides data that are more appropriate to statistical analysis [5].

In the present study, the center point DoE data were used for assay qualification against predetermined acceptance criteria, and these were compared with the robustness data generated from the experimental design to determine the number of events required to generate a similar level of performance. Combining qualification to DoE robustness generated a much larger dataset such that significant trends could be established and assay parameters could be optimized.

Results & discussion

Qualification for precision, linearity, accuracy/bias & range

A sample matrix was used such that each operator analyzed the five mock potency samples in different plate locations over the eight independent events using two working cell banks. A second set of independently prepared mock potency samples and standard was analyzed to provide the replication for determining intermediate precision [5].

Table 1. Center point data used for qualification of the cell-killing assay.

Event	Potency				
	50	71	100	141	200
1	45.0	62.5	97.1	159.4	194.1
	52.6	79.2	107.6	151.6	227.8
2	51.6	67.5	93.6	136.0	214.9
	51.3	68.8	106.0	151.7	217.0
3	57.5	75.6	110.0	148.6	201.7
	51.6	78.8	109.4	146.3	204.8
4	50.9	76.6	113.0	155.5	210.6
	52.2	73.6	103.0	140.9	214.9
5	47.6	70.7	96.5	131.7	209.4
	49.7	77.4	107.6	137.5	222.3
6	44.5	79.3	96.3	149.6	175.1
	51.8	73.3	116.2	144.7	213.4
7	56.5	75.0	103.5	128.8	224.8
	49.7	72.7	97.2	135.6	203.4
8	45.3	67.7	91.5	128.0	180.8
	48.5	74.1	104.8	131.7	209.5
Geometric mean	50.3	73.1	103.1	142.0	207.3
Mean lower 90% CI	48.7	71.0	99.9	137.7	200.8
Mean upper 90% CI	51.9	75.3	106.3	146.4	214.0
Mean bias	0.5	3.0	3.1	0.7	3.6
Bias lower 90% CI	-2.6	0.0	-0.1	-2.3	0.4
Bias upper 90% CI	3.8	6.1	6.3	3.9	7.0

Figure 1. Percentage relative bias plot of a cell-based potency assay.

Intermediate precision, dilutional linearity and accuracy/bias were determined using the $\log_{10}$ derivative of the relative potency data [5]. The center point data are presented in Table 1.

The $\log_{10}$ derivative was used to calculate the geometric mean (GM), 90% CI of the GM, mean relative bias (RB) and the 90% CI of the mean RB according to <USP 1033> (Table 1). The %RB is shown as a plot of mean bias (■) with the 90% CI of the mean (represented by the bars), and this demonstrates equally variability across the mock potency samples centered at zero with no obvious pattern across the potency range (Figure 1) indicative of accuracy in the relative potency measurement over the range of samples tested. For the 200% mock potency sample, there was a small positive bias that was within the acceptance criteria of +15 to -13% (dotted line). The

Table 2. Variance component estimates and intermediate precision.

Variance component	50%	71%	100%	141%	200%	Average
Within run (Var(error))	0.0009759	0.0009817	0.001228	0.000364	0.0011881	0.000947
Between run (Var(run))	0.0000389	-0.000139	-0.00031	0.000594	-0.000207	0.000127
Intermediate precision (GCV%)	7.6	7.5	8.4	7.4	8.3	7.8

Note: Value of 0 was used for the negative Var(run) values in the intermediate precision calculations [8].

DoE data ($n = 32$) generated comparable GMs and not surprising tighter 90% CIs. The mean bias was comparable and all CIs spanned zero bias. As expected for the larger dataset, the 90% bias CIs were tighter than shown above.

For the assessment of linearity, target potency values (50, 71, 100, 141 and 200%) and the individual-observed relative potency values are plotted as $\log_{10}$ -transformed data and regression analysis performed on the data. Log transformation gives points that are equally spaced such that high concentrations cannot dominate the analysis. For the qualification data, the correlation coefficient was 0.99, with a slope of 1.0112 (95% CI = 0.979–1.0430) and the intercept (-0.0129) was within the bounds of the 95% CI of zero intercept (-0.077, 0.051; see Supplementary Figure 2). These data met the acceptance criteria for correlation coefficient of $r > 0.95$ and slope between 0.9 and 1.1.

Intermediate precision was calculated by analysis of variance (ANOVA) [5]. The variance component analysis of the center point data is shown in Table 2. For analysis performed on the same day (repeatability), each operator conducted two assays using independent potency preparations (two reference standard and ten mock potency samples), using the same cells and culture media. The five mock potency samples were loaded on different plate locations over two assays. The within run component of variability (Var(err)) includes variability due to the cells, sample preparation and plate location. For analysis performed on different days (intermediate precision), an operator conducted two assays using independent preparations (as above), using a different cell bank and cell preparation, different culture media and different plate locations. The between run component of variability (Var(run)) includes variability due to cell bank, cell preparation, operator, day, sample preparation and plate location. The intermediate precision for the qualification data was less than 7.8% over the five mock potency samples (Table 2) and met the acceptance criteria Geometric Coefficient of Variation GCV% <15%. The precision, relative bias/accuracy and linearity data demonstrate the capability of the method to determine potency values over the range from 50 to 200%.

These calculations were also performed with the DoE data (six factor $\frac{1}{4}$ -factorial DoE [$n = 16$]) to assess the number of events required to obtain a similar level of precision to the qualification data. The selection of subsets of the DoE data was based on a six factor $\frac{1}{8}$ -factorial ($n = 8$) or a three factor $\frac{1}{2}$ -factorial DoE ($n = 4$). For $\frac{1}{4}$ - and $\frac{1}{8}$ -factorial subset of data, the within run component (Var(error)) and between run component (Var(run)) were similar in magnitude to that detailed in Table 2 and the estimates of intermediate precision were slightly higher to the qualification data, the average precision was 9.2 and 9.6%. For the three factor $\frac{1}{2}$ -factorial subset of data, the within run component (Var(error)) was higher and the between run component (Var(run)) comparable to that detailed in Table 2, the estimates of intermediate precision higher compared with the qualification data, the average precision was 11.4% (see Supplementary Table 2).

In the above studies, mock potency samples were analyzed as independent preparations in two assay runs providing replication at each event. The within run (repeatability) element (Var(err)) was the largest component that increased as the sample size decreased. DoE is increasing being advocated for validation and a four factor $\frac{1}{2}$ -factorial DoE with replication would provide the number of assay runs to achieve good estimates of precision and bias.

The variance component estimates (Var(error) and Var(run)) can also be used to calculate the precision of different assay formats. For the methodology present in this report, the sample preparation is loaded on different locations over three plates, the reportable value as detailed in Table 1 was the result from the single assay. There is, however, a potential improvement in precision from 7.8 to 6.0% by increasing replication from one to two independent assay runs, and this is an effective method of reducing variability and bias in the reportable value. There is a further improvement in precision to 4.8% from three assay runs (see Supplementary Table 3).

Figure 2. Normal plot showing factors having either a positive or negative effect on upper asymptote.

It is important to understand the specification that the methodology will support and the potential for an OOS as a result of assay variability. An OOS is where the test result is outside the specification limits or acceptance criteria established for a particular drug. There are a number of capability indices (Cp, Cpk, Cpm) that can be used to determine process capability [9,10]. It is generally accepted that in manufacturing the process capability for critical quality attributes should be 1.33 or better. Cpm is a capability index that is a function of the specification limits that incorporates assay variability and assay bias and if known product variance into the calculation [1,6]. An analytical method is deemed capable if the index is $1.00 \leq \text{Cpm} < 1.33$, satisfactory if the index is $1.33 \leq \text{Cpm} < 1.50$ and excellent if the index is $1.50 \leq \text{Cpm} < 2.00$ [9,10]. The capability (Cpm) of the methodology can be determined against the proposed specification and the probability of OOS can be calculated using the formulae provided in USP <1033> and the potential for nonconformance to specification calculated [5].

In the early stages of development, a small number of batches may have been manufactured and the lot-to-lot variability is often not known. A single-assay run methodology used to generate a reportable value is able to support a specification of 71–141% with a 0.06% probability of OOS (see Supplementary Table 4). As the drug progresses toward Phase III and commercialization several batches will be manufacture and it is not unreasonable for a process to achieve 5% variance in the critical quality attributes. Incorporating process variability not surprisingly increases the probability of OOS. To mitigate the probability of an OOS, the precision and bias of the methodology must improve and this can be achieved by independent replication. Increasing from 1 to 2 assay runs to generate the reportable value improves precision with a concomitant decrease in bias, the Cpm increases and the probability of OOS decreases (see Supplementary Table 4). These calculations provide information as to the level of risk of failing a batch due to both product and assay variability.

DoE in assessing assay robustness: system suitability

The system suitability was R2, asymptote ratio and CV% between replicate dilutions and for parallelism, delta lower asymptote, delta upper asymptote, delta slope, R2 (PLA) and χ^2 . Three plates failed R2 system suitability for reference standard and three individual samples failed R2 system suitability due to poor replication. Design expert software was used to investigate if there were factors that had a significant impact on system suitability parameters. There were a number of factors that had either a positive or a negative effect on Alamar blue conversion (Figure 2), generating a range of values for the reference standard lower asymptote (437–1556), upper asymptote (1968–8116) and slope (0.77–1.14). Cell concentration, not surprisingly, had a highly significant effect on the upper asymptote (Bonferroni correction) and to a lesser extent on the lower asymptote. Drug time and cell bank exerted a significant negative effect on both the lower asymptote and upper asymptotes. In fact, most factors and their interactions

Figure 3. Boxplot of relative potency versus cell bank.

had either a positive or negative effect on the upper and lower asymptote, the interaction predictable based on the combination of factors. The asymptotes and slope define the shape of the four parameter dose–response curve, the asymptote ratio acting as system suitability to ensure sufficient depth of curve for subsequent PLA. There were no significant factors associated with slope. It is essential in relative potency methodology that one can demonstrate sufficient similarity between the reference standard dose–response curve and that of the test material prior to performing PLA. It is important to note that the change in Alamar blue conversion did not impact on the similarity/parallelism criteria (delta lower asymptote, delta upper asymptote or delta slope) between reference standard and each of the five mock potency samples.

In developing relative potency methodology, it is essential to apply appropriate system suitability criteria that can be used to fail inappropriate dose–response curve rather than generate an erroneous relative potency. The DoE demonstrated that the system suitability is appropriate without compromising quality of relative potency data. The DoE also demonstrated that the methodology would meet all system suitability criteria at the limits of the test factors. However, under certain incubations, the lowest target cell concentration generated the smallest asymptote ratio and this could lead to system suitability failures. An improvement in the methodology would, therefore, be to increase the cell concentration toward the center point.

DoE in assessing assay robustness: relative potency

It is important to note that the critical fold difference of the single assay reportable value methodology was 1.18 for the repeated measurement of a single batch of material under the same conditions and 1.26 over different assay runs (not shown). For the center point, all values were within 118% of the nominal target. Similarly for the 3- and 5- hour high- and low-point data of all values bar one was within 118% of the nominal target. Similar results were obtained for the 3- and 5-hour data and in a paired t-test, there was no significant difference between the relative potency data ($p = 0.37\text{--}0.96$) for each of the five mock potency samples indicating that the assay is robust with regard to Alamar blue development time, which was linear over time (not shown). The remaining discussion of the DoE data is with the 3-hour data, the 5-hour data generating comparable results.

The six factor $\frac{1}{4}$ -factorial DoE with replication, center point and additional assays, generated a total 30 assays analyzed as pairs. The robustness of the methodology was assessed in Design Expert and Minitab software packages for changes in the mock potency reportable value, curve fit parameters and system suitability applied to the methodology. As there are considerable amount of data for brevity, the following is provided as examples of the analysis.

The DoE examined six parameters (dilutional series, cell number, cell recovery, drug incubation time, cell bank, %FBS, Alamar blue concentration) and also Alamar blue incubation time at 3 and 5 h with 4-h incubation at the center point. Cell bank was a categorical variable and data were evaluated to determine if there was any significant difference between the two cell banks. An example of the relative potency data is provided in Figure 3, the boxplot of the respective group data showing the median, (line), the interquartile range (box) and the whiskers, the limits of

Figure 4. Interaction between cell time and alamar blue concentration.

the data overlap at each mock potency level between the two cell banks, and none of the pairings were statistically significant (ANOVA, $p = 0.070\text{--}0.682$). The boxplot shows outliers [*] to that particular dataset, which were not excluded in this or in subsequent analysis as these values represent expected variability in the methodology. There was also no significant difference between operators ($p = 0.106\text{--}0.740$) and no significant difference between the threefold and 2.7-fold dilution series ($p = 0.273\text{--}0.916$) for each of the five mock potency samples.

When considered as an individual factor, in other words, cell number (see Supplementary Figure 3), there was no significant difference between the low, center and high points of the DoE (ANOVA, $p = 0.264\text{--}0.972$) for each of the five relative potency calculations, or for cell recovery time ($p = 0.105\text{--}0.734$), drug incubation time ($p = 0.181\text{--}0.779$), Alamar blue concentration ($p = 0.242\text{--}0.765$) or FBS concentration ($p = 0.180\text{--}0.940$). As there was no significant difference on the relative potency over the five mock potency samples, the relative methodology can be considered robust.

Design expert software was used to investigate possible interactions between the factors and their effect on relative potency. Interpretation was difficult as there was no significant trend over the five mock potency levels, the data confounded by the resolution of the design ($\frac{1}{4}$ -factorial) and the number of factors selected in the models. The coupling qualification to the DoE robustness provided substantial number of center point data and although some interactions were evident in the both the 3- and 5-h data for the 100% mock potency sample, it is important to note that these interactions were often within the variability of the center point data. To improve the resolution of the DoE, the mock potency sample data were normalized to 100% giving ten values for each arm of the DoE. Prior to model selection, there was one highly significant negative effect (Bonferroni correction) between cell concentration (A) and FBS concentration (E) and one highly significant positive effect (Bonferroni correction) between drug time (C) and FBS concentration (E) the largest being the CE interaction (Figure 4). These data indicated that those assays initiated using 8% FBS that relative potency decreased as drug time increased and those assay initiated in 12% FBS that relative potency increased as drug time increased. As noted above, there is considerable overlap between the center point data and test data (Figure 4) and although the data have been normalized, the data are not completely independent as the mock potency data have been generated against two reference preparations (100%), five mock potency in each assay. The analysis, however, demonstrates that cell number, cell time and drug time have a significant effect and these should be adjusted toward the center point as an improvement in the methodology. Alamar blue concentration will be fixed at 20 μl , which should counteract the observed negative effect between

Figure 5. Robustness of the method design space.

cell time and Alamar concentration. FBS will be fixed at 10% which may counteract the observed AE and CE interactions. It also important to note that the CE and AE interactions are confounded with other two factor, three factor and four factor interactions and can be elucidated by additional testing.

The design space for the 100% normalized mock potency for three factors (cell number, cell time and drug time) with Alamar blue, %FBS and cell bank fixed is shown in Figure 5. This model indicates that with cell bank A (or cell bank B), the extremes of the DoE are within the expected variability of the qualified methodology. The model indicates that at low cell concentration potency increases as cell recovery time increases. Potency also increases as cell number increases when cell recovery time is 1 h, whereas potency decreases as cell number increases when cell recovery time is 4 h. There is a smaller increase in potency for drug time with cell concentration or with drug time and cell recovery time.

Design expert software allows for optimization of factors to provide the best recovery for the normalized mock potency samples. For the optimization, cell bank was set to A, although a similar number of solutions were obtained for cell bank B or if combined. Alamar blue and FBS were fixed at nominal concentration. The methodology generated a normalized potency value between 74.8 and 118.4% and for optimization the normalized potency goal was set to between 90 and 111%. The optimization provided one-hundred solutions of cell concentration, cell time and drug time combinations, all with a desirability index of 1, one of these solutions was that used for the center points. The large number of solutions indicates that methodology is robust with regard to the parameters tested.

Factor optimization was a repetitive process fixing one factor at a time and analyzing the generated solutions. As noted above, low cell density could give rise to asymptote (signal-to-noise) system suitability failure and therefore cell concentration was fixed to between 3000 and 4000 cells and optimization conducted for cell time and drug exposure time. During initial optimization for cell recovery, some 75% of the solutions were less than 3.0 h and for drug exposure time, some 65% of the solutions were between 90 and 94 h. Optimization analysis indicates that methodology could potentially be improved and to achieve this the assay factors were moved to within the design space and therefore the cell recovery was set to between 1 and 2.5 and drug exposure time was set to between 90 and 94 h. Under these conditions, the majority of the solutions gave optimal conditions of 3000 cells, 2.5 h cell time and 94-h drug time for a normalized potency goal of 90–111%. The design space of the optimized method conditions and desirability index is shown in Figure 6.

The environmental factors (incubator temperature, CO₂ concentration), where one factor is fixed and the other varied over the accepted operation parameter were assessed at the center points of the DoE. The incubator environmental factors were incubator temperature 36 or 38°C against a control of 37°C all at 5% CO₂. The CO₂ was assessed at 4% and 6% against 5% CO₂ as control arm all at 37°C. The control arm of these studies (37°C and 5% CO₂) was used as qualification data and as center points in the DoE. The high and low incubator set points

Figure 6. The design space of optimized method conditions.

generated comparable data to the qualification data and therefore the environmental factors were not considered to affect the relative potency methodology during the incubators' normal operation conditions.

Conclusion

The requirements for validation of bioassays are defined in USP monograph <1033> [6], ICH Q2R1 [11] and the US FDA and EMA bioanalytical guidelines [12,13]. The use of experimental design (DoE) is increasingly being utilized for bioassay validation [3–5] and a multioperator DoE is typically used to provide estimates of precision, accuracy/bias and to define the range of the methodology. This paper provides further evidence that qualification and validation can be achieved using DoE with similar precision and bias to that achieved using fixed parameters. The DoE data for the five mock potency samples were all within 118% of nominal potency with the center point data generating similar precision (bias, range, etc.) to six factor $\frac{1}{8}$ -factorial DoE (8.5 vs 9.6%). The DoE analysis demonstrated that the methodology was robust with regard to the assay parameters utilized. Coupling qualification to DoE robustness and normalizing the data to 100% provided a large center point dataset that allowed for interactions to be established and for the assay factors to be optimized to be within the 'design space,' and closer to that typically used during routine operation. Under optimal assay conditions, the predicted potency values are within 4% of target.

The data also demonstrated that ruggedness factors that may change over the life cycle of the product such as cell bank, and operator did not have a major impact on assay performance, although it is recognized that new operators will have to be trained and new cell banks qualified before use. The data presented herein demonstrate that a single test result is able to support a specification 71–141%, with a small risk of generating an OOS due to assay variability ($C_{pm} = 1.16$). However, in progressing toward commercialization, the risk of OOS due to combined product and assay variability is increased which can be mitigated through replication ($C_{pm} = 1.57$).

Future perspective

Experimental design will be increasing used to understand the robustness of cell-based potency assays. The use of experimental design, incorporating factors which may be a source of assay variability, is advocated by regulatory authorities and will become increasingly used for qualification and validation of cell-based potency assays. Five mock potency values are typically used in bioassay studies and normalizing data to 100% will provide additional power for subsequent analysis. The use of design expert software to understand interactions and their effect on the relative potency will become increasing used as will factor optimization to establish the best assay conditions that will generate the most accurate relative potency data.

Executive summary

- The aim of this paper is to describe the salient points in establishing robustness of the cell-based potency assay, with reference to activities required for phase-specific qualification.
- The methodology was assessed by fractional factorial design of experiment, the center point data being used for qualification.
- The methodology was qualified with a precision of 8.5% and mean bias of 3.6%.
- The robustness data generated during execution of the design of experiment generated a precision and accuracy/bias similar to the qualification data.
- Statistical analysis demonstrated the methodology to be robust with regard to assay factors tested.
- Assay factors were optimized using design expert software.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Financial & competing interests disclosure

The data presented are from an original work conducted at Piramal paid by an external client. The studies were conducted using concepts detailed in USP monographs and the client gave permission to create a dataset, so that examples could be presented in this article. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The studies used human cell lines, did not involve human subjects and did not require ethical approval.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.4155/bio-2017-0234

References

1. Schofield T. Lifecycle approach to bioassay. In: *Nonclinical Statistics for Pharmaceutical and Biotechnology Industries*. Zhang L (Ed.). Springer Publishing, NY, USA, 433–460 (2016).
2. Little TA. Design of experiments for analytical method development and validation. *Bio. Pharm. Int.* 27(3) (2014). www.biopharminternational.com/design-experiments-analytical-method-development-and-validation
3. Chen XC, Zhou L, Gupta S *et al.* Implementation of design of experiments (DOE) in the development and validation of a cell-based bioassay for the detection of anti-drug neutralizing antibodies in human serum. *J. Immunol. Meth.* 376, 32–45 (2012).
4. Coffy T, Grevenkamp M, Wilson A *et al.* Bioassay qualification using design of experiment. *Bioprocess Int.* 11(6), 42–49 (2013).
5. Miller KR, Mullan A, Novick S *et al.* Use of design of experiment (DOE) for optimization of ADC potency assays. *Am. Pharmaceut. Rev.* (2017). <https://www.americanpharmaceuticalreview.com/Featured-Articles/335152-Use-of-Design-of-Experiment-DOE-for-Optimization-of-ADC-Potency-Assays/>
6. USP <1033> *Validation of Biological Assays*.
7. Wilson R. A personal perspective in the development and validation of a phase specific antibody–drug conjugate cytotoxicity potency assay. *Bioanalysis* 5(9), 1083–1097 (2013).
8. USP <1010> *Analytical Data – Interpretation and Treatment*.
9. Pearn WL, Shu MH. Lower confidence bounds with sample size information for Cpm applied to production yield assurance. *Int. J. Prod. Res.* 24(4), 216–231 (2003).
10. Pearn WL, Shu MH, Hsu BM. Testing process capability based on Cpm in the presence of random measurement errors. *J. Appl. Stat.* 32(10), 1003–1024 (2014).
11. ICH. Q2R1 *Validation of Analytical Procedures: Text and Methodology*.
12. FDA. *Guidance for Industry, Bioanalytical Method Validation*. <https://www.fda.gov/downloads/drugs/guidances/ucm368107.pdf>
13. EMA. *Guideline on Validation of Bioanalytical Methods*. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/WC500109686.pdf