

Variance Components Analysis to Determine Sources of Variation for Vaccine Drug Product Assays

Brian K. Nunnally

ABSTRACT

After determining the analytical variability using the stability data for a vaccine drug product, a screening variance components study was performed on the vaccine content assay and related substance assays. For the content assay, the design included analyst, standard lot, mixing style, heater type, and container as factors. One significant factor (i.e., container) was identified; all other parameters were not deemed significant. For the related substance assay, the following factors were chosen because they could have an impact on the variability: analyst, gel activation, and standard curve. One factor (i.e., standard curve) was identified as having a significant impact on the assay while all the other factors were not. Utilizing the data generated, several recommendations for assay improvement were made. The study design and methodology illustrated in this paper are generally applicable to problem solving, variability reduction efforts, and other applications in addition to the specific analytical application described. Further, use of statistical analysis in support of process design and continuous improvement is supported in the 2008 US Food and Drug Administration draft process validation guidance (1) and Global Harmonization Task Force (GHTF) process validation guidance for medical devices (2).

INTRODUCTION

Variability comes in three forms—process, sampling, and analytical—as illustrated in the following equation:

$$\sigma^2 = \sigma^2P + \sigma^2S + \sigma^2T$$

where σ^2 is the total variance, σ^2P is the process (or lot to lot) variance, σ^2S is the sampling variance (e.g., content uniformity), and σ^2T is the test (analytical) variance. A control chart of the lot release values includes all three forms of variability. It is rare that all three types can be accurately quantitated without a carefully planned and executed hierarchical or similar study. The discussion and design of a hierarchical study has been previously discussed (3). For the analytical laboratory, it is important to know and understand the variance components for an assay.

Information on variation is used to assist in production and analytical investigations, as well as determining sample sizes and strategies for continuous improvement. The required data can be obtained from a number of sources, including method validation. A simple, and better, method for determining the analytical variability of an assay is to use stability data. See reference (3) for a more detailed review of variance component estimation methodologies, including the use of stability data. All stability samples are prepared

For more Author
information,
go to
gxpandjvt.com/bios

ABOUT THE AUTHOR

Brian K. Nunnally, Ph.D., is in charge of process validation at Wyeth in Sanford, NC. He may be reached at nunnalb@wyeth.com.

and filled at the same time, stored together, and are representative of the production material. This means the sampling variability is usually minimal (although it is not always negligible) and there is no manufacturing variability. For parameters that do not exhibit degradation or statistically significant change over time, the variability observed in the stability study can provide an estimate of the variability of the analytical assay. This method is an excellent way to estimate variability when laboratory control and blind control data do not exist. A blind control is a sample submitted to and tested in the laboratory that the laboratory does not realize is a control. It can also be a useful population to compare to control data in the absence of blind controls. This estimation is superior to validation data in most circumstances, because the spotlight effect has been minimized and is, therefore, more representative of production testing. However, this estimation can be inferior to both control and blind control data because the testing is spread over long time periods and may not allow a separation of any random or periodic day-to-day components present from the within-day variance component. Further, if the product is degrading, the estimation of the analytical variance component is complicated and is only possible if the fit of the model is excellent and the slopes are consistent. See "Analysis and Control of Variation" in the Summer 2009 issue of the *Journal of Validation Technology* for further discussion on determination of laboratory variation.

This discussion addresses sources of variation for the analytical determination of two quality attributes in a vaccine product. A number of different analytical methods are used to measure the quality attributes for the vaccine drug product. Two important assays are the content assay and related substance(s) assay. The content assay determines the amount of active ingredient in the vaccine product. The related substance assay determines the amount of impurities in the product. The analytical methods used to determine the attributes of the vaccine in this paper are colorimetric and can be technique dependent. The variability (% relative standard deviation [RSD]) of the content method was 2.3%, based on an older method validation study. This estimate was not believed to represent the current method variability for the assay for three reasons. First, the method validation was done many years ago in a different (development) laboratory. Second, the study was a short-term study (only a few days of experimentation). Finally, the full robustness of the assay was not studied. Using stability data (not shown) to estimate the variability

gave a better representation of the true variability. The stability data give a variability (%RSD) of 5.3%. As described in reference 3, the %RSD is calculated by plotting the data for each lot to determine if there was any consistent change over time (there was not). Then the mean, standard deviation, and relative standard deviation were calculated for each lot. The average relative standard deviation is the approximate variability for the method, because the data are normalized for lot-to-lot differences.

Content Assay Variation History

A Shewhart control chart of the content control sample is shown in Figure 1. These data indicate the variability (%RSD) of the assay to be approximately 2.8%. The %RSD is calculated by dividing the difference between the control chart limit and the mean by three (the standard deviation) and then dividing it by the mean. The estimate comes from the latest set of control chart limits, although this has changed over time (the inherent capability or the underlying variability of the method is approximately 1.1%).

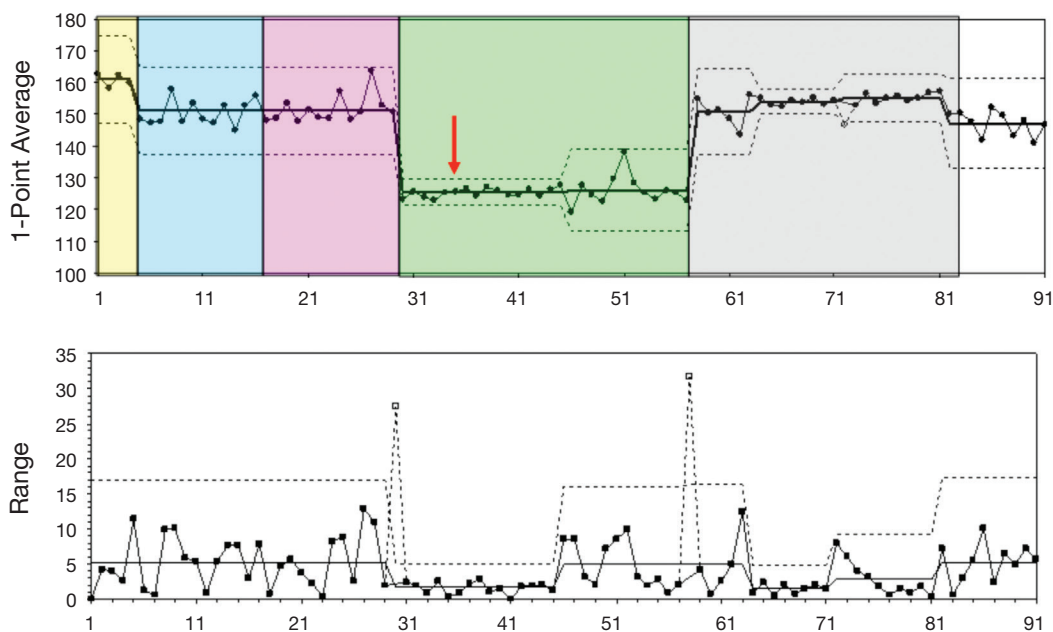
Related Substances Assay Variation History

The variability (%RSD) of the related substances method from an older method validation study is 5.5%. No control sample data were available for this assay. Using the stability data (not shown) to estimate the variability indicated variability (%RSD) of 16.4%. This estimation is approximately three-fold greater than the validation estimation.

"Spotlight" or Hawthorne Effect

It is common to see variability in practice more than twice the variability from the validation (data not shown, personal observation). Variability seen during validation tends to be smaller than the variability seen in routine testing due to a variety of reasons. Validations are characterized by the "spotlight effect" (sometimes called the Hawthorne effect). The spotlight effect is based on the tendency for an experiment to be performed with greater care and control when its outcome or performance is being assessed, closely scrutinized, or has particular impact. The spotlight effect in high impact studies such as validations can result in the under-estimation of variability due to intense concentration and the use of highly skilled analysts. Analysts selected for validations are usually high performers, and even if they are not, they will have an improved focus relative to "normal" samples.

Figure 1: Control chart of the content control sample. Changes in lots of control are shown in shading. The red arrow shows the inherent capability of the method (i.e., 1.1% RSD).



Another reason for under-estimation of variability from validation data is that long-term variability or periodicity is not included in validation data due to the short length of time required to collect the data.

Variance Component Analysis

In order to better understand the sources of variability in these assays and to ultimately improve and reduce the variability, a series of variance component studies were designed and developed. This methodology offers significant value. Its outcome is a better and quantitative understanding of the sources of the variability in the assays. This allows for data driven improvements in performance as well as better data quality, without significant work or time investment. The results of these variance component analyses are reported in this discussion.

CONTENT ASSAY STUDY

Statistical Design

A screening variance components analysis study was designed to identify the major sources of variance in a content assay (using JMP [SAS, v. 6.0.0] software). The design included analyst, standard lot, mixing style, heater

type, and container as factors. Specifics of these factors included the following:

- **Analyst.** Two different analysts performed each assay, designated as Analyst 1 or Analyst 2.
- **Standard lot.** Two different standard lots were used, designated at Lot A or Lot B.
- **Mixing style.** Samples were mixed by either vortexing individual tubes (designated "Individual") or shaking the entire collection of tubes (designated "All-at-Once").
- **Heater type.** Samples were heated by means of a heat block (designated "Dry Block") or a water bath (designated "Water Bath").
- **Container.** Samples were contained in either Eppendorf sealed tubes (designated "Eppendorf") or glass tubes (designated "Glass").

The run order for the trials was randomized. Each result represents an independent run of the assay. Two analysts ran the content assay, mixing the samples two different ways (either by vortexing individual tubes or shaking the entire collection of tubes) with two different standard lots, two different heater types (using either a heat block or a water bath), and two different containers (either glass tubes or Eppendorf sealed tubes) for a total of eight results. Statistical design for the study is shown in Table I.

Results and Data Analysis

The content method results from the variance components screening study are given in Table I. The variance components study results were analyzed using JMP (SAS, v. 6.0.0). A fixed effects model including all factors was used, and the JMP output is shown in Figure 2. The model exhibits a R^2 of 0.86 with a R^2 adjusted of 0.51. The model, including all five factors, is thus able to explain 86% (51% when adjusting for limited degrees of freedom) of the total variance in the results. The R^2 is simply a ratio of the model and total sum of squares, as given in the "Analysis of Variance" section of Figure 2. However, the model is not statistically significant, as evidenced by the $\text{Prob} > F$ of 0.312 (p-values less than 0.050 are considered significant). In addition, no individual factor is significant (Figure 2, "Parameter Estimates" section). The container factor has the largest influence on the model and exhibits borderline significance with a p-value of 0.08 ("Scaled Estimates" section). The model was revised by removing the mixing factor because the $\text{Prob} > |t|$ value was closest to 1.0 (indicating the parameter was least contributing to the variability observed in the model). In addition, the analyst factor was removed because it was not significant in previous validation work and was not significant in this variance components analysis study.

The revised model is shown in Figure 3. The model exhibits a similar R^2 (0.83) as the full factor model. However, it exhibits an improved R^2 adjusted (0.71), indicating this model is superior to the previous model. The model is able to explain 83% of variability using these factors (71% when adjusting for limited degrees of freedom). The model is significant, as evidenced by the $\text{Prob} > F$ of 0.049 (p-values less than 0.050 are considered significant). In addition, the container factor is now significant. The remaining two factors are clearly not statistically significant. The results for the Eppendorf tubes are statistically significantly higher than the results for the glass tubes; this is not surprising, because the closed Eppendorf tubes are considered technically superior to the open glass tubes. Based on this study, the use of Eppendorf tubes was standardized for future testing to ensure this variability does not bias the data. The overall variability for the assay from this variance components screening study was 10.9%; using only the data from the Eppendorf tubes gives a % RSD of 7.2%. Each of these estimates is higher than the variability assessment from the stability data, validation, or control sample data. This higher variability may be due to

Table I: Variance components analysis study results for the content assay.

Mixing	Standard	Analyst	Heater type	Container	Content result (mg/mL)
Individual	A	2	dry block	Eppendorf	18.701
All at once	B	1	water bath	Eppendorf	21.197
Individual	B	2	water bath	Eppendorf	22.232
All at once	B	2	dry block	Glass	17.526
All at once	A	1	dry block	Eppendorf	21.133
Individual	B	1	dry block	Glass	17.001
Individual	A	1	water bath	Glass	18.576
All at once	A	2	water bath	Glass	16.878

the presence of other factors (i.e., analyst and mixing style) whose minor effects were not accounted for in the final model and may over estimate the actual variability of the method.

RELATED SUBSTANCES ASSAY STUDY

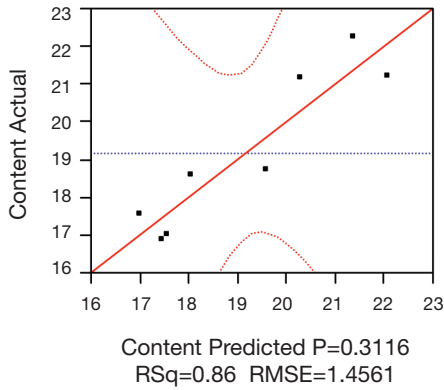
Statistical Design

A screening variance components study was developed and run. The study design is given in Table II including the assessment of analyst, gel activation, and standard curve as factors. Specifics of these variables included the following:

- **Analyst.** Two different analysts performed each assay, designated as Analyst 1 or Analyst 2.
- **Gel activation.** There were three different gel activations, designated as 1, 2, or 3. The gel is used to isolate the related substance from the active ingredient.
- **Standard curve.** There were four different standard curve preparations, designated as 1, 2, 3, or 4.

The factors were chosen because each could have an impact on the variability. Previous validation studies had shown analyst and day were not significant factors. These factors were not randomized in order to facilitate execution. Each result was performed as an independent run of the assay. Two analysts ran the related substances assay with three different gel activations on four different standard curves for a total of 12 results. The study design is shown in Table II.

Figure 2: Full factor model for the content assay.



Summary of Fit

RSquare	0.861285
RSquare Adj	0.514499
Root Mean Square Error	1.4561
Mean of Response	19.1555
Observations (or Sum Wgts)	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	26.329163	5.26583	2.4836
Error	2	4.240455	2.12023	Prob > F
C. Total	7	30.569618		0.3116

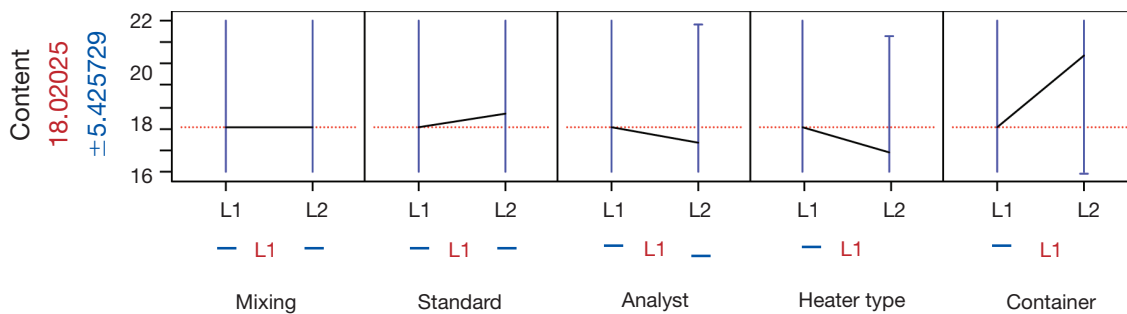
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	19.1555	0.514809	37.21	0.0007
Mixing[L1]	-0.028	0.514809	-0.05	0.9616
Standard[L1]	-0.3335	0.514809	-0.65	0.5835
Analyst[L1]	0.32125	0.514809	0.62	0.5963
Heater type[L1]	0.56525	0.514809	1.10	0.3867
Container[L1]	-1.66025	0.514809	-3.22	0.0842

Scaled Estimates

Term	Scaled Estimate	Plot Estimate
Intercept	19.1555	
Mixing[L1]	-0.028	
Mixing[L2]	0.028	
Standard[L1]	-0.3335	
Standard[L2]	0.3335	
Analyst[L1]	0.32125	
Analyst[L2]	-0.32125	
Heater type[L1]	0.56525	
Heater type[L2]	-0.56525	
Container[L1]	-1.66025	
Container[L2]	1.66025	

Prediction Profiler



Results and Data Analysis

The results from the variance components analysis study are shown in Table II. A suitable model including analyst, gel activation, and standard curve could not be constructed because there were not sufficient data (e.g., loss of degrees of freedom). A revised model is shown in Figure 4. The model exhibits a R^2 (coefficient of determination) of 0.31 with a R^2 adjusted of 0.05. The model is not able to explain much of the variability; the analysis of variance (ANOVA) F-test is not statistically significant (p -value = 0.37).

A revised model using standard curve and gel activation as factors is shown in Figure 5. Analyst was dropped because the Prob > F was higher than gel activation for the first model. The model exhibits an improved R^2 of 0.78 and a R^2 adjusted of 0.59. The model is able to explain 78% of variability using these factors. The model is still only borderline significant, as evidenced by the Prob > F of 0.056 (p -values less than 0.050 are considered significant). This means that there is not strong evidence that the factors in the model have any effect on the related substance level above random noise. The standard curve factor is significant in this model. The total variability was 13.4%. However, if one outlier were removed from the data, the overall variability would be 9.7%. Removal does not affect the previous parameter analysis. A data acceptance criterion related to the difference timepoint-to-timepoint was added to the assay to help average out and reduce this variability for future testing.

Because the absorbance of the sample is in the lower part of the standard curve, any variability in the standard curve and especially the lowest and highest standards, which are influential points, will be transferred to the variability in the sample results. In order to address this, a post-variance components analysis study was executed to try to improve the variability of the standard curve.

Follow-up Study

The purpose of the follow-up study was to determine the effect of increasing the number of replicates of the standards to improve the precision of the related substance assay. Increasing the number of replicates of each standard will reduce the error in the standard and will reduce the standard error of the slope and intercept for each standard curve (e.g., lower standard curve to standard curve %RSD). This in turn will lower the between run %RSD of the assay. Precision (%RSD) was the only response considered because this is the parameter that has generated concern.

Table II: Variance components analysis study results for the related substance assay.

Analyst	Gel activation	Standard curve	Related substance result (%)
1	1	1	10.06
1	2	1	10.55
1	3	1	10.70
2	1	2	10.65
2	2	2	8.96
2	3	2	8.92
1	1	3	14.36
1	2	3	10.67
1	3	3	11.42
2	1	4	12.49
2	2	4	11.16
2	3	4	11.14

Eight replicates of each standard were performed on a single 96-well plate by a single analyst. Data were analyzed in several different ways to determine the effect of increasing the replicates of all standards ($n = 5$) and increasing the replicates of the lowest and highest standard.

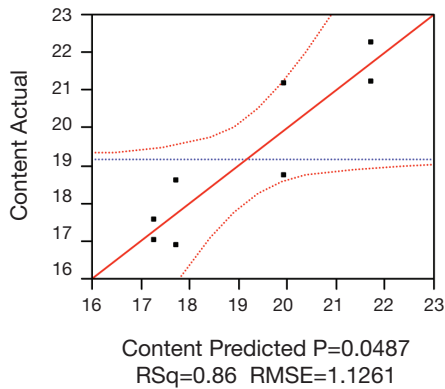
Table III shows the line characteristics (e.g., slope, intercept, and R^2) for the three options, four replicates ($n=4$) used for each standard (hereafter referred to as the Baseline because this is the current normal procedure), eight replicates ($n=8$) for each standard, and eight replicates for the influential points only. Based on the results in Table III, there is no clear advantage to using the extra replication to decrease the slope variability. This is due to the design of the study rather than an indictment of the increased replication concept. This is discussed in the following sections.

Table IV shows the % RSD when only four replicates are used for each standard for the standard curve (this was done by only selecting the first four replicates of each time point. The three different conditions are not independent—all data came from a single calibration run. This represents the baseline case.

Table IV shows the % RSD when eight replicates are used for each standard for the standard curve. The variabilities of the standards were not consistently reduced, as would be expected by theory (assuming the %RSD of the method is constant). This is likely due to the limited amount of data.

In the final instance, the replicates of the lowest and highest standards (only) were increased; only

Figure 3: Revised model for the content assay, removing the most non-significant parameters.



Summary of Fit

RSquare	0.834073
RSquare Adj	0.709627
Root Mean Square Error	1.126093
Mean of Response	19.1555
Observations (or Sum Wgts)	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	25.497279	8.49909	6.7023
Error	4	5.072339	1.26808	Prob > F
C. Total	7	30.569618		0.0487

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	19.1555	0.398134	48.11	<.0001
Standard[L1]	-0.3335	0.398134	-0.84	0.4494
Heater type[L1]	0.56525	0.398134	1.42	0.2287
Container[L1]	-1.66025	0.398134	-4.17	0.0140

Scaled Estimates

Term	Scaled Estimate	Plot Estimate
Intercept	19.1555	
Standard[L1]	-0.3335	
Standard[L2]	0.3335	
Heater type[L1]	0.56525	
Heater type[L2]	-0.56525	
Container[L1]	-1.66025	
Container[L2]	1.66025	

Prediction Profiler

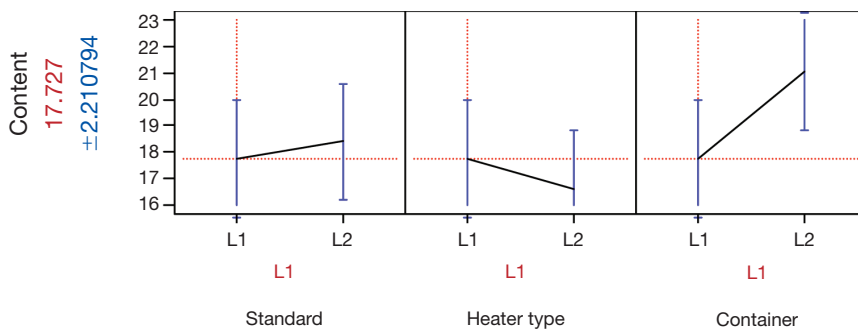
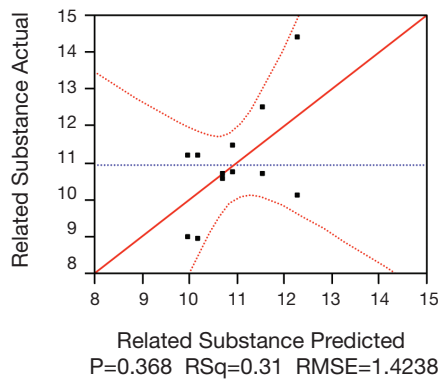


Figure 4: Model including gel activation and analyst for the related substances assay.



Summary of Fit

RSquare	0.311495
RSquare Adj	0.053306
Root Mean Square Error	1.423842
Mean of Response	10.92333
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	7.337667	2.44589	1.2065
Error	8	16.218600	2.02733	Prob > F
C. Total	11	23.556267		0.3680

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	10.923333	0.411028	26.58	<.0001
Analyst[L1]	0.37	0.411028	0.90	0.3943
Gel Activation[L1]	0.9666667	0.581281	1.66	0.1349
Gel Activation[L2]	-0.588333	0.581281	-1.01	0.3411

Scaled Estimates

Term	Scaled Estimate	Plot Estimate
Intercept	10.923333	
Analyst[L1]	0.37	
Analyst[L2]	-0.37	
Gel Activation[L1]	0.9666667	
Gel Activation[L2]	-0.588333	
Gel Activation[L3]	-0.378333	

Prediction Profiler

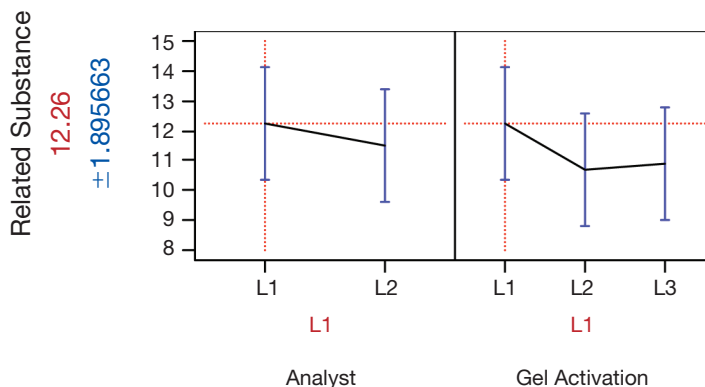


Table III: Related substances assay. Slope, intercept, and R² for the different standard curve options.

Condition	Slope	Intercept	R2
n=4 for each (baseline)	0.00643	0.00105	0.999
n=8 for all standards	0.00643	0.00169	0.999
n=8 for influential points only	0.00642	0.00061	0.999

Table IV: Related substances assay. %RSD when using different numbers of replicates for each standard.

Condition	% RSD (Std 1)	%RSD (Std 2)	%RSD (Std 3)	%RSD (Std 4)	%RSD (Std 5)
N=4 for each (baseline)	5.41%	2.48%	1.12%	1.59%	1.31%
n=8 for all standards	3.88%	2.57%	1.13%	1.66%	1.26%
n=8 for influential points only	3.88%	2.48%	1.12%	1.59%	1.26%

four replicates were used for the middle points on the standard curve. Table IV shows the % RSD when eight replicates are used for the influential (first and last) data points for the standard curve.

Based on the data, there is no clear advantage to adding extra replicates to the standard curve; this is due to the design of the study rather than an indictment of the increased replication concept. Due to time constraints, a larger study could not be conducted. However, there are no negative consequences to the implementation of the extra replication either. Statistical and analytical theory states that increasing the replication of the influential points (the lowest and highest point of the standard curve) should increase the confidence of the result (between run) and therefore help stabilize the slope (thus reducing the slope variability across runs). Because most of the related substance values correspond to absorbances in the lower portion of the standard curve, reducing the slope variability will result in decreased between-standard-curve to standard-curve variability. Therefore, it is logical to increase the replicates of the lowest and highest standard to eight and re-validate the related substance assay with the increased standard replication. The most effective way to run a study like this is to evaluate the data over a long period of time (e.g., up to 30 independent runs). Because the data

requirement is for a long-term study, it makes sense to implement based on the theoretical benefits and to monitor the data to determine if the changes have the intended impact.

IMPLICATIONS FOR VALIDATION AND COMPLIANCE

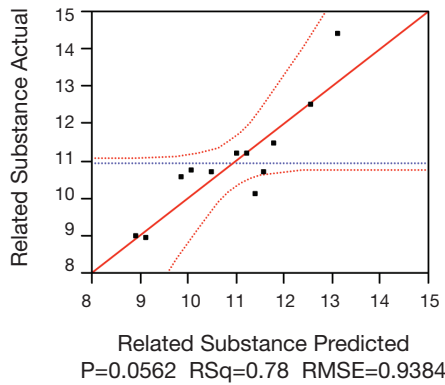
This case study demonstrates the following key points:

- Analytical method validation data may not accurately represent method performance on a day-to-day basis. Often this work is conducted under optimum conditions (i.e., analysts with best technique, recently qualified instruments, most reliable instruments, etc.). The use of stability data, control data, and blind control data is superior to validation data.
- When excessive analytical data variation is noted, running a study to determine the sources of variation is recommended to determine the respective components of the total variation. The first step in this process is to identify the potential sources of variation.
- Once the sources of variation are identified, steps should be taken to replicate across the parameters to ultimately reduce the overall analytical variability.
- The extra replication for the standard curve has been implemented as a way to reduce standard curve to standard curve variability.
- The approach and methods described in this study for analytical test methods are generally applicable to evaluating process variation and other situations where multiple sources of variation may be individually quantitated or optimized process parameters are being investigated. The use of designed experiments with statistical analysis of data is recommended.
- Consultation with personnel trained in experimental design and statistical methods is recommended.

CONCLUSIONS

Two screening variance studies were performed to better understand the variability of the content and related substance assays. For each assay, one parameter was identified as significantly affecting the overall assay variability. In the content assay, the container was identified as a major source of variation. In the related substances assay, standard curve was identified as the major source of variation. A follow-up study was inconclusive, but provided some data to indicate improved variation. While controlling these factors

Figure 5: Revised model for the related substances assay using gel activation and standard curve as factors.



Summary of Fit

RSquare	0.775683
RSquare Adj	0.588753
Root Mean Square Error	0.938444
Mean of Response	10.92333
Observations (or Sum Wgts)	12

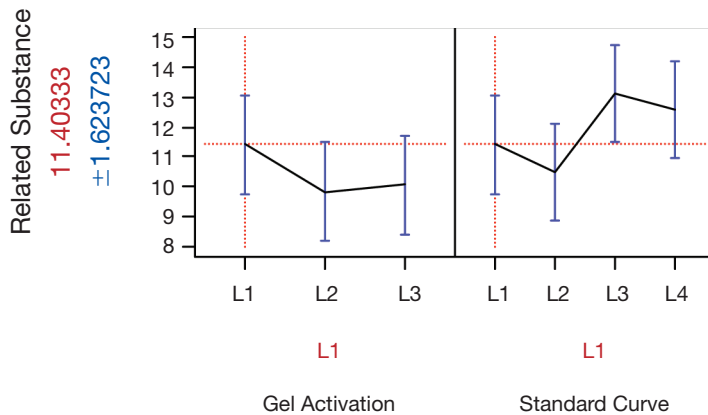
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	18.272200	3.65444	4.1496
Error	6	5.284067	0.88068	Prob > F
C. Total	11	23.556267		0.0562

Scaled Estimates

Term	Scaled Estimate	Plot Estimate
Intercept	10.923333	
Gel Activation[L1]	0.966667	
Gel Activation[L2]	-0.588333	
Gel Activation[L3]	-0.378333	
Standard Curve[L1]	-0.486667	
Standard Curve[L2]	-1.413333	
Standard Curve[L3]	1.226667	
Standard Curve[L4]	0.6733333	

Prediction Profiler



will reduce the variability, this may not be sufficient to ultimately control the assay to the levels desired. Additional variation improvements could be realized by utilizing harmonization efforts discussed in the Winter 2009 issue of the *Journal of Validation Technology* (4). These harmonization efforts are simple, easy to implement, and reduce the controllable variability. This includes a comprehensive review of the procedure, determination of common practices, and identification of targets. All parameters within this study have been harmonized, even though most of them were insignificant. The discussion in this manuscript demonstrates the use of designed experiments and statistical analysis to improve two analytical procedures. However, this same strategy and approach is generally applicable to evaluating process variation and other situations where multiple sources of variation may be individually identified and quantitated. The approach and methods of this discussion are consistent with respective process validation guidelines issued by FDA and GHTF for pharmaceutical and medical device applications.

ACKNOWLEDGMENTS

The author gratefully acknowledges the help and assistance of Stefan Jordi, Dominik Kuenzler, Maud Frerejean, Nathalie Dubois, and Lee Fox with the work detailed in this manuscript.

REFERENCES

1. FDA, *Guidance for Industry, Process Validation: General Principles and Practices, Draft Guidance*, November 2008.
2. GHTF, *Quality Management Systems—Process Validation Guidance*, January 2004.
3. Nunnally, B.K. and McConnell, J.S., *Six Sigma in the Pharmaceutical Industry*, Boca Raton: CRC Press, 2007.
4. Nunnally, B.K.; Johnson, A.; Kaiser, R.K., "Reduction of Analytical Method Variability in a R&D Laboratory: Case Study," *Journal of Validation Technology*, Winter 2009. **JVT**