Organic Process Research & Development

Concept Article

pubs.acs.org/OPRD

# Applications of Gage Reproducibility & Repeatability (GRR): Understanding and Quantifying the Effect of Variations from Different Sources on a Robust Process Development

Amol A. Deshpande,[†] Ayyalasomayajula Ramya,[†] Vavilala Vishweshwar,[†] Girish R. Deshpande,*[,†] and Amrendra Kumar Roy*[,‡]

[†]Custom Pharmaceutical Services, Dr. Reddy's Laboratories Ltd., Bollaram Road, Miyapur, Hyderabad, Telangana State 500049, India

[‡]Process Engineering Excellence, CTO-III, Dr. Reddy's Laboratories Ltd., Plot 116, 126C & Survey Number 157, S.V. Co-operative Industrial Estate, IDA Bollaram, Jinnaram Mandal, Medak District, Telangana State 502325, India



**ABSTRACT:** During process development, it is always a debatable issue whether the variation in analytical results is due to the measurement system (MS) or due to the process. The best approach is to quantify total variation coming from the MS prior to any process improvement activity. This quantification is done by "Gage Reproducibility & Repeatability" (GRR). This article describes the usage of GRR for quantifying variation from various sources and selecting a suitable MS for the analysis. In this study, two instruments, a potentiometer and ultra high pressure liquid chromatography (UPLC), were evaluated for the assay measurement of a key starting material (KSM) supplied by a vendor. As a result of the GRR study, it was found that the potentiometer was not a suitable instrument, because of the high variation contributed by it, whereas UPLC was found to be suitable, because of the insignificant variation contributed by it towards the assay. In addition to this, it was also observed that the variation contributed by the KSM samples was insignificant, indicating that those samples were coming from a robust process, and the vendor was found to be suitable for supplying the KSM.

## INTRODUCTION

Nowadays, quality by design (QbD) is an essential part of any process development activity. This is evident by the fact that there is a plethora of literature available on the QbD approach for the process development of drug substance,[1] drug product,[2] and analytical method development.[3] These research articles focus on the process development of drug substances and drug products assuming that the measurement systems (MS) are perfect for the intended test, and even if they are not, it has rarely been discussed in these articles with some exceptions.[4,5] The real outcome of any process consists of measurable attributes such as critical quality attributes (CQA) associated with the product, for example, assay, purity, impurity levels, pH, and so forth, and in order to measure these attributes one requires some kind of MS such as high-pressure liquid chromatography (HPLC), gas chromatography (GC), and so forth. There is always a measurement uncertainty (Figure 1) in all analytical procedures, and only a few articles on QbD have highlighted this issue in pharmaceutical analysis.[6]

## PROBLEM STATEMENT

During the initial phase of a process development, it is difficult to carry out a developmental activity in the absence of validated analytical methods. As method validation is a time-consuming and is a parallel activity, there is a need for a tool that could provide a temporary solution for selecting a proper analytical instrument for a given analysis. In other words, it is of interest to estimate the total variation in a given analytical result and the contribution of the process and the MS towards the total variation. This quantification helps to take an appropriate action. In case there is high variation from the process, one needs to work on the robustness of the process, and if the culprit is the MS, then one has to use an alternative MS, or one should try to reduce the variation coming from the existing MS. The first step in doing so is to calculate the total variation contributed by the MS itself.

As MS is the integral part of any process development, it is imperative to have an idea about a variation contributed by it towards the total variation in the measurement. MS variation is often found to be the major contributor. If we cannot trust the MS, we cannot trust the data it produces. If an unreliable MS is used during process validation or during a process improvement program using QbD/Six Sigma strategy,[7] it could mislead the analysts to look for the cause of variation in the process or in the raw material instead of looking for it in the MS. Hence it is important to have an idea about the variation contributed by the MS. The study of the variation coming from a MS can be estimated by a well-known "Six Sigma" tool called "Gage Reproducibility & Repeatability"[8,9] or simply GRR, which was initially developed by the automotive industry.[10] If this variation is found to be within the specified range, only then is a MS certified for the analysis.[9,11] This article illustrates how GRR can help in quantifying the variation coming from a MS and how it can help in selecting the most suitable analytical instruments for a given analysis. This concept was extended further for vendor qualification.

**Brief Introduction to GRR.** It is important to have an overview of GRR, as it is seldom used in the pharmaceutical industry. As stated earlier, this tool helps in estimating the

**Received:** September 14, 2014
**Published:** November 14, 2014

   dx.doi.org/10.1021/op5002935 | Org. Process Res. Dev. 2014, 18, 1614−1621

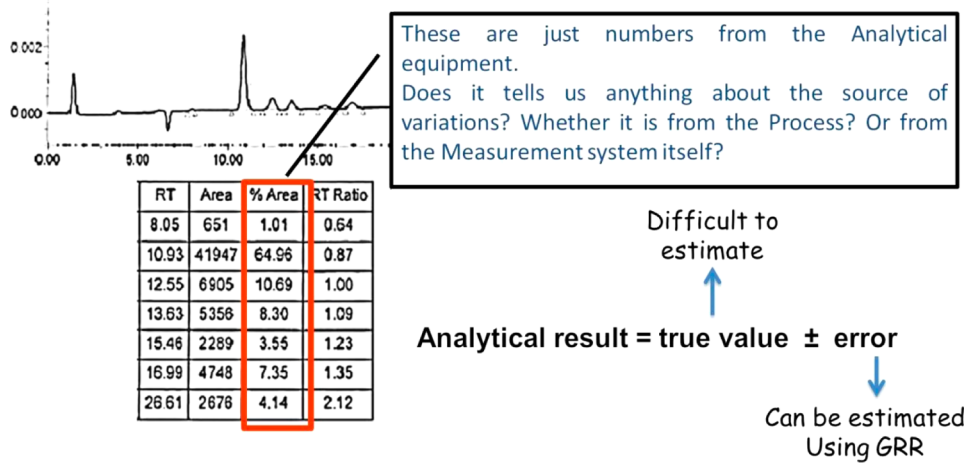


**Figure 1.** Typical HPLC chromatogram: what do analytical results represent?

amount of variation in an analysis introduced by a MS and compares it with the total variation which in turn evaluates the suitability of the MS prior to its use for any analysis.

The total variation in any measurement is a culmination of the variation coming from the MS and the variation contributed by the process as shown in Figure 2 and also by eqs 1 and 2.

$$\text{total variation (TV)} = \text{process variation (PV)} + \text{MS variation (GRR)} \quad (1)$$

$$\sigma^2_{TV} = \sigma^2_{PV} + \sigma^2_{GRR} \quad (2)$$

$$\sigma^2_{GRR} = \sigma^2_{AV} + \sigma^2_{EV} \quad (3)$$

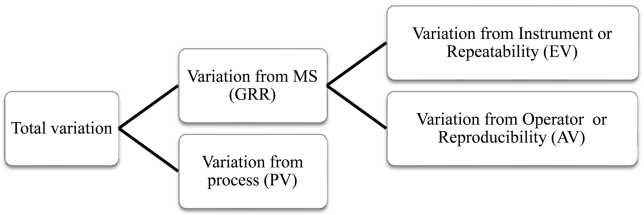


**Figure 2.** Components of total variation in a measurement.

MS variation in eq 1 is popularly known as GRR which consists of the variation due to operator/appraiser (AV) and the variation due to equipment (EV) as shown in eq 3.[12] If the variation from the MS is known, one can calculate the true variation from the process. This enables researchers to work on the process development effectively. Most importantly GRR aids in selecting a suitable MS for a given analysis based on the magnitude of variation contributed by a given MS. A GRR study quantifies three things:

Gage Repeatability (R) or variation from the instrument (EV): is the variation obtained from one gage and one operator when measuring the same sample several times.

Gage Reproducibility (R) or variation from the operators using the instrument (AV): is the difference in the average of the measurements made by different operators using the same gage and measuring the same sample.

Total Gage Repeatability & Reproducibility (GRR): is the vector sum of EV and AV as shown in eq 3 and Figure 3.

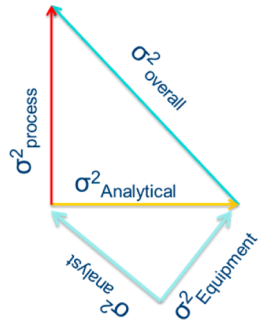


**Figure 3.** Analogy of variance addition with the Pythagoream theorem ($\sigma^2$ = variance).

The percent GRR (eq 4) estimates the percent of the total variation that is contributed by the MS.[13]

$$\%GRR = \frac{GRR^2}{TV^2} \quad (4)$$

Finally, true variation from the process (PV) could be obtained by subtracting the GRR from the total variance (TV) using eq 2.

It is evident from the above discussion that GRR studies are not only useful in estimating the MS variation but can also help in estimating the variation coming from the process as shown in eq 2 and eq 3. Hence GRR can be used in the early phase of the process or method development to avoid any surprises later during the commercialization[14] of the process as summarized in Figure 4.

**How Does GRR Work?** Usually GRR is performed by taking samples deliberately in such a way that it represents the entire spectrum of the specification. Preferably, some samples are taken closer to lower specification limit (LSL), some samples are taken closer to upper specification limit (USL), and some samples are taken closer to the mean value of the specification. The main aim of the GRR study is to test the ability of the MS to differentiate between such samples. In layman terms, if the variation from the MS is significantly small (with respect to the total variation), then it can differentiate all the three samples, but on the other hand if the variation from the MS is large or equal to the total variation, it cannot judge the difference between the samples, and hence the MS would not be suitable for the analysis.


dx.doi.org/10.1021/op5002935 | Org. Process Res. Dev. 2014, 18, 1614−1621

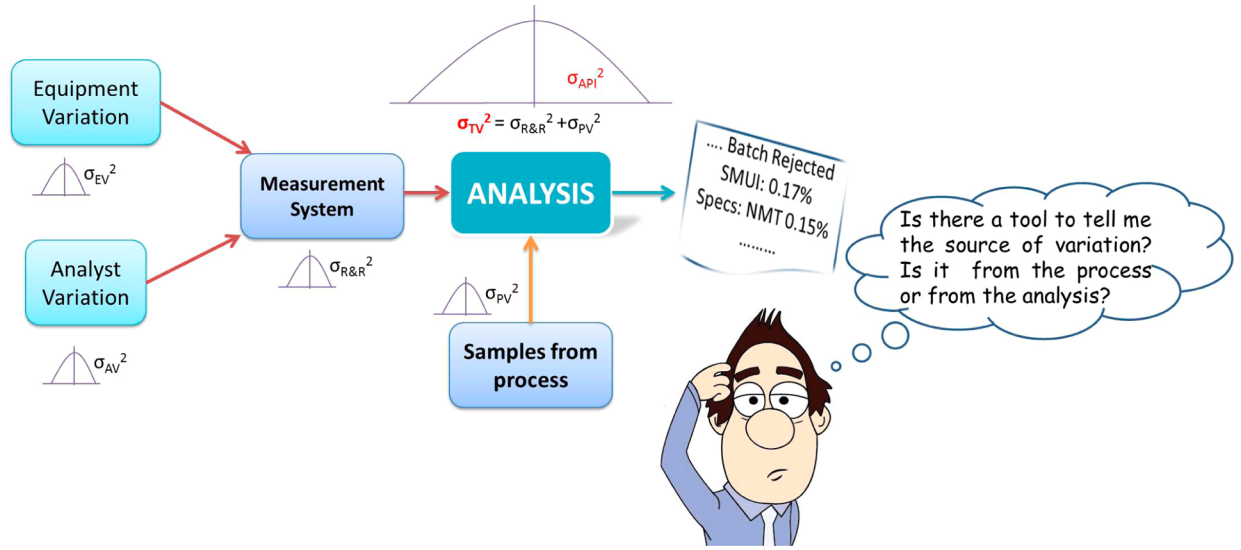


**Figure 4.** Situation faced during the commercialization of the process if sources of variation are not studied properly.

For example, for the assay specification of an Active Pharmaceutical Ingredient (API) between $\mu_1$ and $\mu_3$, three samples: Sample 1, Sample 2, and Sample 3 were taken representing the entire specification range. Each sample was analyzed multiple times for the assay by two different instruments: Instrument 1 and Instrument 2, which resulted in the mean assay value of $\mu_1$, $\mu_2$, and $\mu_3$ for Sample 1, Sample 2, and Sample 3, respectively, where $\mu_1 < \mu_2 < \mu_3$. Let the overall or total variance in the system be $\sigma^2$.

Let us assume that the inherent variation $(\sigma_1^2)$ from Instrument 1 is greater than that of Instrument 2 $(\sigma_2^2)$ as shown in Figure 5. Further $\sigma_1^2$ is greater than or equal to the

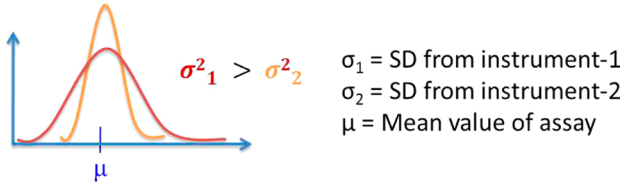


**Figure 5.** Inherent variance from two instruments. The curves represent the distribution of the data points around the mean value $\mu$. A wider curve means more variation in the experiment.

total variation of the system $(\sigma^2)$. Because of this higher inherent variance $(\sigma_1^2)$ from Instrument 1, the overall distribution of the assay results from the three samples would overlap due to which Instrument 1 cannot distinguish between the three samples. In GRR terms, the number of distinct categories is 1, which means Instrument 1 considers all the samples as the same. Hence, Instrument 1 could not be used for the analysis. This scenario is depicted in Case 1, Figure 6.

On the other hand, if the inherent variance $(\sigma_2^2)$ is less than the total variation $(\sigma^2)$, as with Instrument 2 (Figure 6), the distribution of the assay results for each sample would be much narrower, and there would be no overlap of the distribution. This would enable Instrument 2 to distinguish between the three samples, and in this case the number of categories would be 3, as it can differentiate between all the three types of samples. Hence, Instrument 2 would be preferred for the analysis. This scenario is depicted in Case 2, Figure 6.

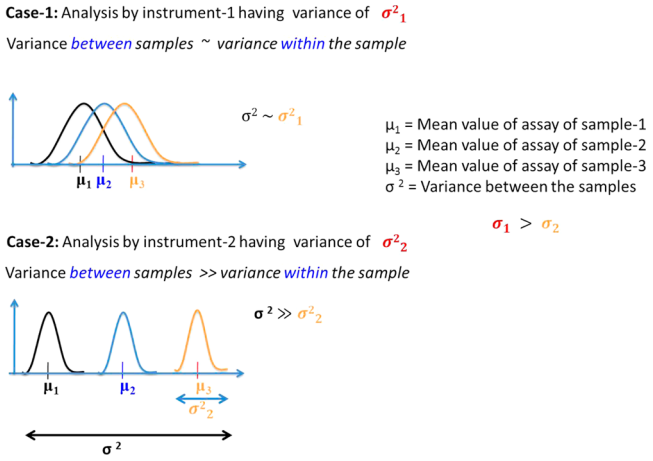


**Figure 6.** Capability of two instruments in differentiating between the samples.

To understand GRR better, we need to draw an analogy of the above discussion with analysis of variance (ANOVA). Let us assume that we are trying to study the effect of three samples and three analysts on the assay measurement using one HPLC. Each sample is analyzed thrice by each analyst. In this case, the ANOVA table would be similar to Table 1. Based on the *p*-value, the ANOVA table would indicate whether a variation from a given source is significant or not.

**Table 1. Understanding GRR using ANOVA**

| source of variation | DF | SS | MSS | *F*-statistics[a] | *p*-value |
|---|---|---|---|---|---|
| samples (part-to-part/ between group variation) | 2 | a | $x_1 = a/2$ | $x_1/x_4$ | |
| analysts (reproducibility/ precision) | 2 | b | $x_2 = b/2$ | $x_2/x_4$ | |
| interaction effect of sample and analyst | 4 | c | $x_3 = c/4$ | $x_3/x_4$ | |
| error (equipment/within/ repeatability/accuracy) | 18 | d | $x_4 = d/18$ | | |
| total | 26 | | | | |

[a]If this ratio is >4, then the given term has a significant effect, this would be reflected by *p*-value <0.05.


dx.doi.org/10.1021/op5002935 | Org. Process Res. Dev. 2014, 18, 1614−1621

Note that the samples represent the entire specification range, i.e., there is a variation *between* samples (as stated earlier, this is done deliberately in a GRR study). If the variation from the instrument (*within* variation) is really less (as depicted in Case 2, Figure 6), then the *F*-ratio ($x_1/x_4$, Table 1) of $MSS_{between/samples}$ and $MSS_{instrument/error}$ would be high (usually, the minimum criterion is >4) which means the samples are not the same. This would be indicated by *p*-value <0.05 at 95% confidence level. It indicates that the instrument is sensitive enough to differentiate between the samples and can group them in multiple categories.[15]

On the other hand, if $MSS_{instrument/error}$ is high (more variation from the instrument), then the *F*-ratio of $MSS_{between/samples}$ and $MSS_{instrument/error}$ is close to 1, it indicates that the instrument is not sensitive enough to differentiate between the samples even if there exists a variation among the samples and *p*-value would be >0.05. Similarly, the variation due to different analysts and the variation due to the interaction effect of analyst and sample can be explained.

Thus, the whole purpose of GRR is to check the instrument suitability for a given analysis and also to check whether the MS is sensitive enough to differentiate between the samples, provided the samples represent the entire spectrum of the specification. If the MS is not suitable, then one either needs to change the MS or reduce the GRR further.

In layman terms, $MSS_{instrument/error}$ could be taken as a system noise, and $MSS_{between/samples}$ and $MSS_{analyst}$ could be considered as a signal. If signal-to-noise ratio is >4, we can say that the measurement system is capable enough to detect the variation between two samples.

## EXPANDING THE SCOPE OF GRR FROM THE ABOVE DISCUSSION

The concept of GRR can also be used in a different way for resolving the following day-to-day objectives:

**Objective 1.** To have a quantitative idea about the total variation in the analytical results and to understand the contribution of process and MS towards that total variation.

This is required during process optimization, because usually a time-validated analytical method is not available during the initial stages of process development. This is also quite helpful prior to initiating any QbD study, in particular during the design of experiments (DoE) stage. Conducting DoE is ineffective, in the absence of stabilized/optimized processes and reliable analytical methods. Partitioning the total variation into its components (eq 2) puts an end to any acrimonious debate between Process Research and Development (PR&D) and Analytical Research and Development (AR&D) (Figure 4). If the GRR indicates that the variation is due to the process, then PR&D needs to improve its process, and if the variation is from the MS, then AR&D needs to work on its analytical method.

**Objective 2.** Selecting a suitable MS and developing a method of analysis quickly to support the process optimization in the absence of a validated method.

This is a typical case for GRR studies. As the process is usually not stable during the initial stages of process development, the samples from different lab batches would have significant variation in quality parameters. This should be followed by a GRR study with multiple samples using a suitable analytical instrument. The output from the GRR would be in the form of ANOVA as shown in Table 1. Using the ANOVA table, one can infer the major source of variation and derive a plan of action accordingly. This process is described as a flow diagram in Figure 7.

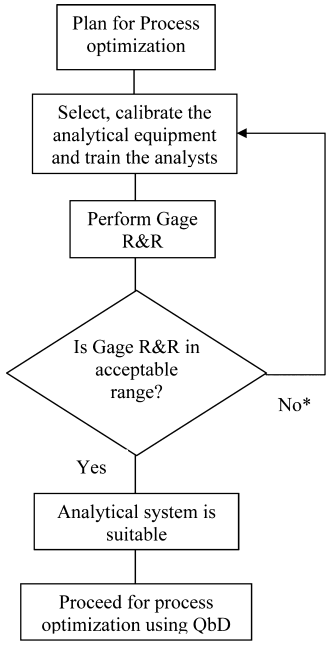


**Figure 7.** Flow diagram for GRR study. If GRR is high even after 2−3 iterations, go for an alternate analytical tool.

The significance of this GRR study for MS suitability is that one does not require a reference standard for the analytical method development.

**Objective 3.** To qualify a vendor for supplying a KSM, based on the robustness of the vendor's process. However, a GRR needs to be performed for all of the quality parameters. This is not so difficult, as a single chromatogram can provide data related to all of the impurities which can be used for GRR study.

Consider Case 2 (Figure 6), where the variation from MS is much less. On the basis of this, we will try to answer the question: "What happens to the GRR if the samples are from a robust process?" In this case, the means of all of the samples are too close, which results in complete overlap of the three distributions as shown in Figure 8. Hence, the MS would see

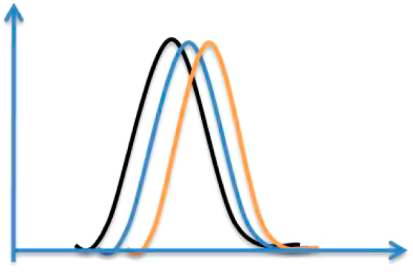

**Figure 8.** Samples from a robust process as seen by the MS.

them as a single sample as described earlier. This observation would also be confirmed from the ANOVA table where $MSS_{between/samples}$ and $MSS_{instrument/error}$ would be nearly equal, which is quite obvious as they all are coming from same robust process. In this case, the number of distinct categories would be equal to 1. Hence, one can conclude that the vendor's process


dx.doi.org/10.1021/op5002935 | Org. Process Res. Dev. 2014, 18, 1614−1621

is quite robust, and it can be qualified, provided variation from MS is insignificant. One needs to understand that, in the absence of sample variation, the total variation is loaded on the MS; i.e., nearly 80−100% variation would be contributed by the MS. In this case, one needs to focus on the absolute value of the standard deviation coming from the MS and not the percent variation.

Note: Even Case 1 (Figure 6) gives the number of distinct categories as 1. But, in that case, the variation from the MS is too high, because of which there is an overlap of the distribution resulting in the number of categories as 1.

Similarly, if a process is not robust at the vendor's end, then the ratio of $MSS_{between/samples}$ and $MSS_{instrument/error}$ would be high, which means all samples are not same, and in this case, the number of distinct categories would be higher. It needs more work for vendors to make their process robust and would lead to the rejection of the vendor's material without any ambiguity. In this case, traditional interpretation of GRR will work as shown in Case 2 (Figure 6), and one can inform the vendor that too much variation is coming from the vendor's process and the variation from the MS is very less and request the vendor to make the process robust. The vendor should be able to understand and accept the results without any ambiguity as the scenario has been explained with transparency.

To summarize, if the variation from the MS is less and the number of distinct categories is 1, it means that the process is robust.[16] Hence, the process is under control, and R&D can start the QbD study (Objective 1), or if samples are provided by a vendor, then the vendor is qualified, as the process is robust.

## RESULTS AND DISCUSSION

This article is about quantifying the variation from the MS and from the process and also about finding a suitable MS for the assay analysis of a KSM. It was essential to qualify the supplier of the KSM based on GRR data. This was necessary because the MS was used not only for vendor qualification but also for the assay analysis during the DoE studies where the KSM was used as an input material for the reaction.

As the KSM has acidic functionality, a potentiometer was considered to be the right choice because of lesser setup time, lesser analysis time, and ease of operation, and most importantly because it was affected by lesser number of variables (compared to a HPLC). The main aim of this article is to qualify potentiometer as a suitable instrument for assay measurement for future DoE studies. In addition, it was used to estimate the real process variation (PV) coming from the KSM by using eq 2 through eq 4 so that the vendor could be qualified based on potentiometric titration.

A typical GRR study involves multiple operators measuring multiple samples, and each sample has to be measured multiple times. The number of samples, operators, and trials can vary from case to case (in the present case there are three operators, three samples, and three trials each). Multiple trials are essential to estimate repeatability (EV) and multiple operators are required to estimate reproducibility (AV). Multiple samples were taken to get a better estimate of the process variation or part to part variation (PV). The commonly used methods for calculating GRR are AIAG and ANOVA.[17,18]

**AIAG Method.** This method uses the average and range[19] to estimate repeatability and reproducibility of MS. It cannot estimate the operator by part interaction. This methodology was not used in the present case.

**ANOVA Method.** This method uses average and variance ($\sigma^2$) to estimate repeatability and reproducibility of MS. Contrary to the AIAG method, it allows the estimation of the variance component due to operator by part interaction.

## PREPARING FOR GRR STUDY

The following preparations are made before starting a GRR study.

**Gage Calibration.** It ensures the accuracy of the readings through its operating range.

**Checking Gage Resolution.** The gage should be able to distinguish at least 10 readings within the tolerance range (USL-LSL). As a potentiometer can detect up to 100 parts, the resolution is good enough for our requirements.

**Sample Collection.** For GRR study, preferably, the samples from production should be chosen[20] in such a way that they represent the entire variation in the process. In the present case, three such samples were selected at random from the vendor, assuming a nonrobust process.

**Identification and Training of Operators.** A minimum of three operators were chosen and trained on the equipment for GRR study. All of the three operators were trained on potentiometer with the same sample of KSM.

**Randomization of the Experiments.** Analysis has to be done in a random fashion to eliminate the block effect of any uncontrolled factors (such as temperature and humidity in analytical lab on a given date). Hence, three samples were analyzed by three operators with three replications, summing up to a total of 27 experiments in a random fashion.

The results of the assay obtained from the potentiometer are captured in Table 2.

**Table 2. Assay results obtained from a potentiometer for GRR calculation**

| | | samples of KSM | | |
|---|---|---|---|---|
| | trials | 1 | 2 | 3 |
| Analyst A | 1 | 100.30 | 100.43 | 100.37 |
| | 2 | 99.83 | 100.31 | 100.03 |
| | 3 | 100.09 | 100.16 | 99.86 |
| Analyst B | 1 | 100.09 | 100.63 | 99.46 |
| | 2 | 99.95 | 100.04 | 99.36 |
| | 3 | 100.53 | 100.46 | 99.68 |
| Analyst C | 1 | 98.59 | 98.49 | 98.22 |
| | 2 | 99.66 | 99.35 | 99.2 |
| | 3 | 98.78 | 99.08 | 99.01 |

The GRR was calculated by ANOVA method using Minitab statistical software (Table 3).

As evident from Table 3, the operator to part interaction was insignificant; hence ANOVA was calculated again by ignoring this interaction term (Table 4). It is evident from the *p*-value that the variation from the operators was significant.

Finally, Minitab gave the GRR data for the potentiometer as shown in Table 5. The GRR contribution towards total variance was found to be 94.76%, whereas the %PV contribution was only 5.24%. Out of the total 94.76% GRR variation, 72.75% was contributed by the operator (AV), and 22% was contributed by the potentiometer (EV). Hence, it was concluded that %AV needs to be reduced in order to reduce % GRR, which could be done by training the operators. But, as the % EV was also too high (Table 5), the potentiometer was considered unsuitable


dx.doi.org/10.1021/op5002935 | Org. Process Res. Dev. 2014, 18, 1614−1621

**Table 3. Minitab output two-way ANOVA with interaction**

| source | DF | SS | MS | *F* | *P* | remarks |
|---|---|---|---|---|---|---|
| parts (sample) | 2 | 0.8276 | 0.41378 | 2.6818 | 0.182 | not significant |
| operators[a] | 2 | 8.0942 | 4.04712 | 26.2308 | 0.005 | significant |
| parts × operators | 4 | 0.6172 | 0.15429 | 1.2190 | 0.337 | not significant |
| repeatability | 18 | 2.2783 | 0.12657 | | | |
| total | 26 | 11.8172 | | | | |

[a]Operator = reproducibility.

**Table 4. Minitab output of two-way ANOVA table without interaction**

| source | DF | SS | MS | *F* | *P* | remarks |
|---|---|---|---|---|---|---|
| parts (sample) | 2 | 0.8276 | 0.41378 | 3.1439 | 0.063 | not significant |
| operators[a] | 2 | 8.0942 | 4.04712 | 30.7507 | 0.000 | significant |
| repeatability | 22 | 2.8954 | 0.13161 | | | |
| total | 26 | 11.8172 | | | | |

[a]Operator = reproducibility.

**Table 5. Minitab output of GRR for potentiometer**

| source | variance component | % contribution |
|---|---|---|
| total GRR | 0.566667 | 94.76 |
| repeatability (EV) | 0.131611 | 22.01 |
| reproducibility (AV) | 0.435057 | 72.75 |
| process | 0.031352 | 5.24 |
| total variation | 0.598019 | 100.00 |

for assay analysis. High variation from the potentiometer was also evident from the absolute value of standard deviation (SD) given in Table 6. Out of the total SD of 0.77, a potentiometer was contributing almost 0.36 SD, which was quite high.

**Table 6. Standard deviation of various components for potentiometer**

| source | SD |
|---|---|
| total GRR | 0.75 |
| repeatability (EV) | 0.36 |
| reproducibility (AV) | 0.66 |
| process | 0.17 |
| total variation | 0.77 |

Further data from Minitab showed that the distinct number of categories was 1 for the potentiometer. This meant that either the potentiometer was unable to distinguish between the three samples due to high EV or the samples were coming from a robust process. Later on, it was found that the KSM, which has an acidic functionality, contains some impurities that were also acidic in nature, and during titration both of them were getting titrated giving a large variation. Hence, it was decided not to pursue the potentiometer any further for method development, and a decision was taken to test UPLC for GRR study.

Later on, the UPLC method was developed for fast analysis, and it was subjected to GRR study, whose results are presented in Tables 7, 8, and 9.

The GRR results of both the analytical equipment were compared, and it was observed that the major difference was in the percent variance due to analyst (or operator). It was zero for UPLC and ∼72.75% for potentiometer.

Percent variation contributed by the process towards the total variance was 6.65% for potentiometer and 0% for UPLC.

**Table 7. Assay results obtained from UPLC for GRR calculation[21]**

| | trial | Analyst 1 | Analyst 2 | Analyst 3 |
|---|---|---|---|---|
| Sample 1 | 1 | 99.12 | 99.28 | 99.18 |
| | 2 | 99.39 | 99.38 | 99.45 |
| | 3 | 99.37 | 99.87 | 99.33 |
| Sample 2 | 1 | 99.31 | 99.20 | 99.60 |
| | 2 | 99.35 | 99.27 | 99.26 |
| | 3 | 99.55 | 99.76 | 99.55 |
| Sample 3 | 1 | 99.20 | 99.34 | 99.31 |
| | 2 | 99.58 | 99.13 | 99.31 |
| | 3 | 99.20 | 99.37 | 99.51 |

**Table 8. Assay results obtained from UPLC for GRR calculation**

| source | VarComp | % contribution |
|---|---|---|
| total GRR | 0.05 | 100.00 |
| repeatability (EV) | 0.05 | 100.00 |
| reproducibility (AV) | 0.00 | 0.00 |
| process or sample (PV) | 0.00 | 0.00 |
| total variation | 0.05 | |

**Table 9. Standard deviation of various components for UPLC**

| source | SD |
|---|---|
| total GRR | 0.22 |
| repeatability (EV) | 0.22 |
| reproducibility[22] (AV) | 0.00 |
| process | 0.00 |
| total variation | 0.22 |

This low variation from the samples indicated that they were coming from a robust process, because of which the number of distinct categories from both the equipment was 1. Hence, it was concluded that the vendor samples could be approved for further use, as negligible variation was contributed by the samples to the total variation. At this point it is important to mention that the decision on qualifying a vendor cannot be made only on the basis of one quality parameter but based on all the quality parameters.

Readers may wonder that the percent contribution by potentiometer towards total variance was lower (∼22%) than


dx.doi.org/10.1021/op5002935 | Org. Process Res. Dev. 2014, 18, 1614−1621

that of UPLC, which was 100% (Tables 8 and 9), rendering UPLC as unsuitable for the analysis. But, in case of UPLC, the % variation contributed by the analyst and process was zero, because of which entire variance was loaded onto the UPLC equipment (= 100%). Hence, in the present case, it would be prudent to look at the absolute values of variances rather than relative percent variances. One needs to understand, in present context, that the total variance from UPLC was 0.05, whereas it was 0.13 from the potentiometer, which was 2.6 times higher. The total variance of 0.05 from UPLC for the assay measurement was quite insignificant (compare Tables 5 and 8). The same conclusion could be drawn from the standard deviation data given in Tables 6 and 9.

Hence, UPLC was the best choice for assay measurement, and it could also be used for vendor qualification and for monitoring DoE experiments in the future.

## NOTE ON CURRENT METHOD VALIDATION

As per the guidelines, a % GRR < 20% is good, and <30% is acceptable for other industries. We were unable to find similar guidelines for the pharmaceutical analysis. Q2 (R1) guidelines of ICH[5] do insist on calculating the repeatability (EV) and reproducibility (AV) for method development, but these two cannot be added as per eq 2 to give GRR because they are calculated separately by two different set of experiments. Moreover, reference standards or pure samples were used for the method development which ensured that the % PV becomes zero (Tables 5 and 9), which in turn ensured that any variation whatsoever was due to the equipment and the analyst. During method validation, efforts were directed towards reducing the variance from the equipment (EV) and because of the analyst (AV). This served the purpose of the analytical team for providing a robust analytical method. But, it does no good to PR&D as they could not know the contribution of their process towards total variation and which can create confusion during scale up (Figures 4 and 7). Hence, GRR is the better way of developing a method of analysis, which is useful during scale-up, as the various sources of variation are known.

## CONCLUSION

This article describes the importance of estimating the variance coming from the MS before any improvement program is initiated. Using GRR, we could quantify the variation contributed by the analyst and by the equipment, separately. This study also enabled us in selecting the right analytical equipment for the analysis, which later on helped in qualifying the vendor as the variation from the vendor's samples was quite low. This concept could further be applied during the process development, where it is possible to partition the total variation into its components (GRR and PV). If the variation is more from the analysis (GRR), then analysts need to improve their method either by reducing EV or AV; otherwise they need to change the equipment itself. On the other hand, if the variation is more from the process (PV), then process chemists need to make the process more robust. Last but not the least, a GRR needs to be performed for all of the quality parameters associated with the molecule. This is not so difficult, as a single HPLC chromatogram can provide data related to all of the impurities which can be used for GRR study.

## AUTHOR INFORMATION

### Corresponding Authors

*Telephone: +919701346355. Fax: + 91 08458 279619 E-mail: amrendrakr@drreddys.com.

*E-mail: girirajd@gmail.com.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We appreciate the efforts of Shilpi Lathia from DRL corporate communication for drawing the animations used in this article. DRL Communication Number IPDO-IPM 00368.

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AIAG | Automotive Industry Action Group |
| ANOVA | analysis of variance |
| API | active pharmaceutical ingredient |
| AR&D | Analytical Research and Development |
| AV | appraiser or operator variance |
| CQA | critical quality attribute |
| DF | degree of freedom |
| DoE | design of experiments |
| EV | equipment variance |
| GRR | Gage R&R |
| HPLC | high-pressure liquid chromatography |
| KSM | key starting material |
| LSL | lower specification limit |
| MS | measurement system |
| MSS | mean sum of squares |
| PR&D | Process Research and Development |
| PV | process variance |
| QbD | quality by design |
| $s,\sigma$ | standard deviation/variation/spread of a set of data points |
| SD | standard deviation |
| SS | sum of squares |
| TV | total variance |
| UPLC | ultra high-pressure liquid chromatography |
| USL | upper specification limit |

## REFERENCES

(1) (a) Mohanty, S.; Roy, A. K.; Kumar, V. K. P.; Reddy, S. G.; Karmakar, A. C. *Tetrahedron Lett.* **2014**, *55*, 4585. (b) Mohanty, S.; Roy, A. K.; Phani Kiran, S.; Eduardo Rafael, G.; Kumar, K. P. V.; Karmakar, A. C. *Org. Process Res. Dev.* **2014**, *18*, 875.

(2) (a) Musters, J.; van den Bos, L.; Kellenbach, E. *Org. Process Res. Dev.* **2013**, *17*, 87. (b) Cimarosti, Z.; Bravo, F.; Castoldi, D.; Tinazzi, F.; Provera, S.; Perboni, A.; Papini, D.; Westerduin, P. *Org. Process Res. Dev.* **2010**, *14*, 805. (c) Bravo, F.; Cimarosti, Z.; Tinazzi, F.; Smith, G. E.; Castoldi, D.; Provera, S.; Westerduin, P. *Org. Process Res. Dev.* **2010**, *14*, 1162. (d) Ende, D.; Bronk, K. S.; Mustakis, J.; O'Connor, G.; Santa Maria, C. L.; Nosal, R.; Watson, T. J. N. *J. Pharm. Innov.* **2007**, *2*, 71.

(3) (a) Furlanetto, S.; Orlandini, S.; Mura, P.; Sergent, M.; Pinzauti, S. *Anal Bioanal Chem.* **2003**, *377*, 937. (b) Deshpande, G. R.; Roy, A. K.; Rao, N. S.; Rao, B. M.; Reddy, J. R. *Chromatographia* **2011**, *73*, 639.

(4) (a) Skrdla, P. J.; Wang, T.; Antonucci, V.; Dowling, T.; Ge, Z.; Ellison, D.; Curran, J.; Mohan, G.; Wyvratt, J. *J. Pharm. Biomed Anal.* **2009**, *50*, 794. (b) Hofer, J. D.; Olsen, B. A.; Rickard, E. C. *J. Pharm. Biomed. Anal.* **2007**, *44*, 906.

(5) Sun, M.; Liu, D. Q.; Kord, A. S. *Org. Process Res. Dev.* **2010**, *14*, 977. (c) Snodin, D. J. *Org. Process Res. Dev.* **2011**, *15*, 1243.

(6) Dejaegher, B.; Jimidar, M.; De Smet, M.; Cockaerts, P.; Smeyers-Verbeke, J.; Vander Heyden, Y. *J. Pharm. Biomed. Analysis* **2006**, *42*,

155. (c) Houson, I. *Process Understanding: For Scale-Up and Manufacture of Active Ingredients*, 1st ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Berlin, 2011; Chapter 1.
(7) Kubiack, T. M.; Benbow, D. W. *The Certified Six Sigma Black Belt Handbook*, 2nd ed.; Pearson Education: Upper Saddle River, NJ, 2009.
(8) (a) Houf, R. F.; Berman, D. B. *IEE Trans. Components, Hybrids Manufacturing Technol.* **1988**, *11* (4), 516. (b) Burdick, R. K.; Borror, C. M.; Montgomery, D. C. *J. Quality Technol.* **2003**, *35*, 342. (c) Montgomery, D. C.; Runger, G. C. *Quality Eng.* **1993a**, *6*, 115;(d) *ibid* **1993b**, *6*, 289. (e) Montgomery, D. *Introduction to Statistical Quality Control*; John Wiley & Sons, Inc.: Hoboken, NJ, 2005. (f) *NIST/SEMATECH e-Handbook of Statistical Methods*. http://www.itl.nist.gov/div898/handbook.
(9) Burdick, R. K.; Borror, C. M.; Montgomery, D. C. *Design and Analysis of Gauge R&R Studies*; Society for Industrial and Applied Mathematics: Philadelphia, PA, 2005.
(10) (a) Automotive Industry Action Group (AIAG), *Measurement System Analysis*, 2nd ed.; AIAG: Detroit, MI, 1985. (b) Automotive Industry Action Group (AIAG), 2002, *Measurement System Analysis*, 3rd ed., Detroid, MI. (c) Kubiack, T. M.; Benbow, D. W. *The Certified Six Sigma Black Belt Handbook*, 2nd ed.; Pearson Education: Upper Saddle River, NJ, 2009.
(11) Fanelli, T. *Advances in Process Measurements for the Ceramic Industry*; Jillavenkatesa, A., Onoda, G. Y., Eds.; Wiley: New York, 1999; pp 367−376.
(12) The relationship shown in eqs 2 and 3 is analogous to the Pythagorean theorem.
(13) This calculation is based on the ANOVA method where the ratio of variances ($\sigma^2$) is taken. In the AIAG method, it is the simple ratio of GRR and TV ($\sigma$) which give the exaggerated number (see Table 2).
(14) Zhou, L.; Socha, J. M.; Vogt, F. G.; Chen, S.; Kord, A. S. *Am. Pharm. Rev.* **2010**, *13*, 74.
(15) For detailed discussion on ANOVA, understanding the sum of squares (SS) and mean sum of square (MSS), see ref 9.
(16) In layman terms, if the variation from the MS is small (with respect to the total variation), then it can differentiate all three samples, but on the other hand if the variation from the MS is large or equal to the total variation, it cannot judge the difference between the samples, and hence the MS would be unsuitable for the analysis as shown in Figure 6. This is indicated by the *F*-ratio in ANOVA table. Usually if *F*-ratio is > 4, it implies that the contribution from (say sample or analyst) is significant towards variance.
(17) http://asq.org/quality-progress/2006/03/measure-for-measure/improved-gage-rr-measurement-studies.html, March 2006.
(18) http://asq.org/quality-progress/2006/05/measure-for-measure/appraiser-variation-in-gage-rr-measurement.html, May 2006.
(19) Range = maximum value − minimum value in given data set.
(20) One should avoid the use of reference standard for calculating gage R&R as it would not give any part-to-part variation.
(21) It appears that the intra- and interanalyst variability are high for UPLC, especially considering the statement that the low variations in sample indicate they are coming from a robust process. However, recalculating the standard deviation, it was found that intra- and inter-analyst variation in UPLC was less, as shown below.

| | | Standard Deviation | |
|---|---|---|---|
| | | UPLC | Potentiometer |
| Intra Analyst | Analyst-1 | 0.18 | 0.22 |
| | Analyst-2 | 0.24 | 0.46 |
| | Analyst-3 | 0.17 | 0.45 |
| Intra Sample or Inter Analyst | Sample-1 | 0.05 | 0.66 |
| | Sample-2 | 0.06 | 0.76 |
| | Sample-3 | 0.04 | 0.57 |

(22) The value of zero for variance looks strange, but the reproducibility data of Sample 2 was analyzed manually and found that the variance was 0.002, i.e., 0.2%, which is too small to be considered. This may be due to the fact that there are very few mistakes that an analyst can commit during HPLC analysis, for example, the weighing of the sample and the concentration of the sample to be injected, which can really affect the assay result. As a part of the GRR study, there is a prerequisite that the analyst must also be trained on the analysis before performing the actual GRR study. This can explain this low value of reproducibility. Variability can also be caused by the mobile phase prepared by different analysts, but that can only affect the RT of peaks and not the assay results.