

**June 2010**  
**Acceptable Variability for Sampling Recovery Studies**

Several months ago (January 2010), my favorite statistician, Lynn Torbeck, published an article in *Pharmaceutical Technology* entitled “%RSD: Friend or Foe”. The article was basically about the misuse of statistics. In it Mr. Torbeck made the statement that applying a percent relative standard deviation (%RSD) criterion to percentage recovery values in recovery studies was not statistically valid because the values themselves were already percentages. Now, it is common practice in cleaning validation programs to include a criterion for %RSD for recovery percentages in sampling recovery studies (such as swab recovery studies). For example, companies might specify for swab recovery studies that the data collected for a given spiked level have a %RSD of  $\leq 15\%$ .

This got me thinking. Is Lynn right, or are most of the pharmaceutical companies (as well as yours truly, who has taught the use of a %RSD criterion for recovery studies) right? Furthermore, if we abandon the %RSD criterion, what criterion do we use to measure variability in a sampling recovery study? And, finally, does it really make that much difference? In situations like this, I often try to look at practical data to see the impact.

Here is my first example, which in comparison to a second case, illustrates that %RSD is not necessarily a good measure of variability in a sampling recovery study. For simplicity, I am only going to illustrate this with swab sampling, involving one operator who performs three replicates. In Case A below, the 100% recovery value is  $2.03 \mu\text{g}/\text{cm}^2$ . The values obtained, the average, the standard deviation and the %RSD is given in the Table A below. For simplicity, the data on recovery values will omit the units ( $\mu\text{g}/\text{cm}^2$ ).

Table A Data

	Data values	% Recovery Values
Replicate 1 value	1.93	95
Replicate 2 value	1.96	97
Replicate 3 value	2.01	99
Average	1.97	97
Standard deviation	0.040	2.0
%RSD	2.1	2.1

Now, we have a second operator with the following data for the exact same sampling situation. Table B has the data for that operator in this situation.

Table B Data

	Data values	% Recovery Values
Replicate 1 value	1.03	51
Replicate 2 value	1.06	52
Replicate 3 value	1.11	55
Average	1.07	53
Standard deviation	0.040	2.0
%RSD	3.8	3.8

Now admittedly this is an extreme case. With one operator getting recoveries of 97% and a second operator getting recoveries of 53%, I would suspect something is wrong. However, that doesn't change the statistical evaluation. If I use %RSD for each operator, then it appears that the variability of the data with operator B (% RSD = 3.8) is much greater than the variability of operator A (%RSD = 2.0). However, when one looks at the data itself, and specifically at the standard deviations, the standard deviation in each case is the same, suggesting the variability in each case is the same.

If it doesn't make sense to use %RSD as a measure of variability, what can we use instead? If we look at the data in Table A and Table B, perhaps we fall back to using just the standard deviation of the values themselves as a measure of variability. However, while that works in those two specific situations, what will happen in a significantly different case? Let's suppose the 100% recovery value is not 2.03  $\mu\text{g}/\text{cm}^2$ , but is 20.3  $\mu\text{g}/\text{cm}^2$ . Table C has one possible data set for that situation.

Table C Data

	Data values	% Recovery Values
Replicate 1 value	19.3	95
Replicate 2 value	19.6	97
Replicate 3 value	20.1	99
Average	19.7	97
Standard deviation	0.40	2.0
%RSD	2.1	2.1

Note that the 100% value in Table C is 10 times higher than the 100% value in Table A. In addition, the data for the three replicates are 10 times higher than the data in Table A. In this case, if we look at the actual standard deviation values, we see that they are significantly different (0.040 for Table A vs. 0.40 for Table C), and we conclude that perhaps in this situation the %RSD is a better indication of variability of the data.

What we are faced with is a dilemma. What gives a better indication of variability of replicates, the standard deviation values (which appear a better measure in comparing A and B), or the %RSD values (which appear to give a better measure in comparing A and C).

One possible solution is to base the measure of variability on the standard deviation of the data itself as a percentage of the 100% recovery value. What this does is normalize the data, so that the variation in Table A and Table B are the same, but also the variation in the data in Table A and Table C are the same. The data expressing the standard deviation as a percentage of the 100% value is given below for each of the three cases covered above:

Table A Case:  $100 (0.040/2.03) = 2.0\%$

Table B Case:  $100 (0.040/2.03) = 2.0\%$

Table C Case:  $100 (0.40/20.3) = 2.0\%$

In looking at the data in the three situations, the variability of each operator seems to be the same. This measure (the standard deviation of the values as percentage of the 100% recovery value) appears to reflect that similarity. Furthermore, in cases where the data values might be more divergent, it would also appropriately reflect that greater variability.

Where does that leave us? Am I expecting people to start using this measure of variability in place of the commonly used %RSD for sampling recovery studies? Probably not. Part of the reason is that a variability criterion for sampling recovery studies is typically set at a relatively high level (15%-20% RSD), reflecting the high variability of recovery studies. Furthermore, if percent recoveries are relatively high (>80%), the difference between the proposed new measure and the conventional %RSD is somewhat minor. Finally, if percent recoveries are low but still acceptable (such as 50-65%), then the %RSD measurement will give a higher measure of variability, thus reflecting a worst case. So, while this proposed measure may provide a more scientific basis for the degree of variability, the existing method is not terribly wrong (particularly for something as variable as a swab recovery study).