Comparison of $1/x^2$ vs $1/x$ weighting for Standard Curves of Pharmaceutical PK/TK Data

Dale Schoener, Khanh Nguyen, Michael Buonarati

Alta Analytical Laboratory, El Dorado Hills, CA

Introduction

According to pharmaceutical industry guidance set by the FDA[1] and Best Practice guidance documents[2], calibration curves should use the simplest model that adequately describes the concentration-response relationship. Similarly the EMEA states a relationship which can simply and adequately describe the response of the instrument with regard to the analyte should be applied[3].

A non-weighted linear regression does not work because this model assumes equal variance over the entire range of samples. In a typical calibration curve the variance (SD squared) increases as the concentration increases.

The solution to improve the accuracy for the lowest standards is of course to use weighting. A linear regression with $1/x^2$ weighting is the most common regression and weighting used. The reason the $1/x^2$ weighting works so well is that it draws a regression line that minimizes the sum of the vertical deviation of the standard points from the regression line on a percentage basis. If the % difference from the expected results off the standard curve is summed for all the data points, this sum will be zero for a $1/x^2$ weighted linear regression. Given that acceptance criteria in this industry for standards and QCs is a constant percentage over nearly all the concentration range, then a $1/x^2$ weighting, which minimizes the sum of the deviation of the standard curve points on a percentage basis, would be the most logical weighting to use. Still people will look at a $1/x$ weighting in the belief that if this works, and given that it is "simpler" since there is no squared term, then it should be used.

The choice of weighting is typically taught as taking the sum of the absolute value of the relative error of several QC data points at each of several concentrations over the entire concentration range. This summation is done for the various weighting factors[4-6]. The least amount of weighting that minimizes this error is chosen. This method will provide for the best regression/weighting for the 3-4 validation runs. This same regression/weighting will then have to be used for all subsequent sample analysis runs where over time factors typical for LC/MS runs such as higher than expected variance or non-linearity are experienced. Therefore the weighting decision needs be based on more than the validation results only and typically the most rugged weighting needs to be chosen.

The purpose of this study is to compare $1/x$ weighting and $1/x^2$ weighting using simulated peak area ratios to determine which weighting is the most rugged. Low variance data along with real world LC/MS situations such as poor precision and high variance will be used to determine ruggedness.

Method

All data was simulated . Random normally distributed peak area ratios (PAR) were generated in Excel using a given standard deviation at the standard or QC level. A slope of one was assumed. The Excel formula used to generate a PAR was =(NORMSINV(RAND()*std dev)) + std /QC level, where RAND() returns a random number from 0 to less than 1 and NORMSINV returns the inverse of the standard normal cumulative distribution with a mean of zero and a standard deviation of one. A 1000 fold standard curve range was used along with three QC levels at the ULOQ, half of the ULOQ, and the LLOQ.

The standard deviations (low/normal variance scenario) at each standard curve point (range 1 to 1000 ng/mL) were taken from an actual LC/MS/MS validation. The CV's from LLOQ to the ULOQ were 6.8, 3.5, 4.6, 2.3, 1.6, 1.3, 2.7, and 2.9% respectively, for the 1, 2, 4, 20, 100, 400, 800, and 1000 levels. The QC CV was set at 6.8% at the LLOQ and 2.0 and 2.9 at the 1/2 ULOQ and ULOQ levels respectively. Two standards curve points were used at each level and 6 QCs were used at each level.

For the high variance scenario an 8% CV was used at all standard curve and QC points except for the LLOQ, where a 15% CV was used.

To simulate non-linearity the PARs were calculated at 90% of the 400, 800, and 1000 std level, and at 90% of the 500 and 1000 QC level, but accuracy was based on the unadjusted level.

For each simulation the same randomly generated standard curve PARs were regressed using a $1/x^2$ weighting and a $1/x$ weighting and the slopes and y intercepts calculated. Using these slopes and intercepts QC concentrations were calculated from the same randomly generated QC PARs. This was done 30 times for each of the following scenarios: normal (low) variance, high variance, normal variance with non-linearity, and high variance with non-linearity.

The parameters to determine ruggedness include: 1) grand mean accuracy along with the min and max accuracy, 2) the number of QC means (n=6) outside 15% (20% at the LLOQ), 3) the number of standard curve points dropped, and 4) the sum of the absolute error at each QC level.

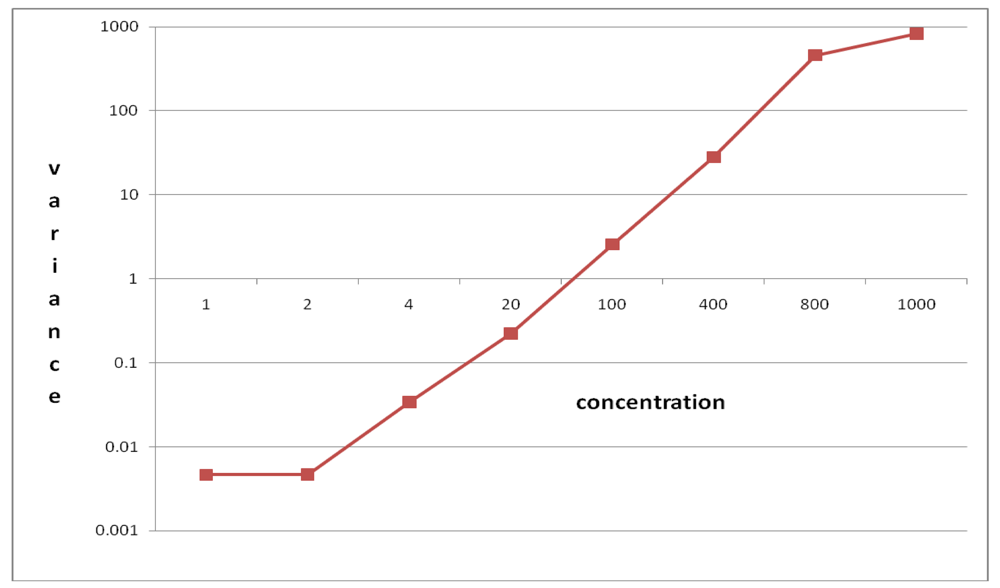

Figure 1. The change in variance vs. concentration over a 1000 fold standard curve range.

$1/x^2$ weighting

x	PAR	concentration	% accuracy	% difference from expected	absolute % difference from expected
1	0.93256	0.91985	92.0	-8.0	8.0
1	1.22208	1.20567	120.6	20.6	20.6
2	1.95067	1.92492	96.2	-3.8	3.8
2	1.97841	1.95230	97.6	-2.4	2.4
4	3.63813	3.59075	89.8	-10.2	10.2
4	4.08286	4.02978	100.7	0.7	0.7
20	20.13218	19.87346	99.4	-0.6	0.6
20	20.57412	20.30973	101.5	1.5	1.5
100	101.05592	99.76025	99.8	-0.2	0.2
100	100.61440	99.32438	99.3	-0.7	0.7
400	398.35802	393.25274	98.3	-1.7	1.7
400	400.30136	395.17118	98.8	-1.2	1.2
800	807.86195	797.50934	99.7	-0.3	0.3
800	846.12669	835.28376	104.4	4.4	4.4
1000	1009.82123	996.88073	99.7	-0.3	0.3
1000	1035.06064	1021.79672	102.2	2.2	2.2
			totals:	0.0	58.9

$1/x$ weighting

concentration	% accuracy	% difference from expected	absolute % difference from expected
0.93661	93.7	-6.3	6.3
1.21968	122.0	22.0	22.0
1.93201	96.6	-3.4	3.4
1.95913	98.0	-2.0	2.0
3.58181	89.5	-10.5	10.5
4.01662	100.4	0.4	0.4
19.70781	98.5	-1.5	1.5
20.13988	100.7	0.7	0.7
98.82576	98.8	-1.2	1.2
98.39409	98.4	-1.6	1.6
389.49368	97.4	-2.6	2.6
391.39365	97.8	-2.2	2.2
789.85969	98.7	-1.3	1.3
827.27057	103.4	3.4	3.4
987.31232	98.7	-1.3	1.3
1011.98853	101.2	1.2	1.2
	totals:	-6.1	61.5

Figure 2. Error comparison in $1/x^2$ and $1/x$ standard curves with the same PAR's.

Normal (Low) Variance Summary

$1/x^2$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	99.5	99.8	99.7	
Min accuracy	94.9	97.3	96.9	
Max accuracy	102.8	102.2	102.8	
Sum absolute error	59.5	34.4	31.1	Total absolute error over all three QC levels: 125.0

Number of standard curve points dropped over 30 runs (480 standard curve points total): 1

$1/x$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	99.1	100.0	100.0	
Min accuracy	91.8	96.9	96.1	
Max accuracy	105.8	103.7	103.6	
Sum absolute error	77.8	38.6	43.1	Total absolute error over all three QC levels: 159.5

Number of standard curve points dropped over 30 runs (480 standard curve points total): 1

Table 1: n=30 runs, n= 6 replicates per QC level 1000 fold range, high QC is at ULOQ, mid QC ½ of the ULOQ , 8 standard curve levels, each level in duplicate CV for standards and QCs: 1.3 - 4.6% CV for non-LLOQ standards, 6.8% CV at the LLOQ. Mid QC 2% CV, ULOQ QC 2.8% CV.

High Variance Summary

$1/x^2$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	99.8	99.8	100.9	
Min accuracy	87.7	91.7	93.5	
Max accuracy	114.7	111.8	108.1	
Sum absolute error	128.6	90.6	81.2	Total absolute error over all three QC levels: 300.4

Number of standard curve points dropped over 30 runs (480 standard curve points total): 30

$1/x$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	99.4	100.0	101.1	
Min accuracy	76.6	90.6	93.6	
Max accuracy	121.2	114.4	108.8	
Sum absolute error	188.8	119.4	97.1	Total absolute error over all three QC levels: 405.3

Number of standard curve points dropped over 30 runs (480 standard curve points total): 28

Table 2: n=30 runs, n= 6 replicates per QC level 1000 fold range, high QC is at ULOQ, mid QC ½ of the ULOQ , 8 standard curve levels, each level in duplicate.
CV for standards and QCs: 8% CV for non-LLOQ standards and QCs, 15% CV at the LLOQ.

Non Linearity and Low Variance Summary

$1/x^2$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	103.9	93.8	93.3	
Min accuracy	98.7	90.4	90.6	
Max accuracy	108.8	95.8	95.6	
Sum absolute error	123.4	187.0	200.1	Total absolute error over all three QC levels: 510.6

Number of standard curve points dropped over 30 runs (480 standard curve points total): 0

$1/x$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	92.6	100.4	100.0	
Min accuracy	83.4	95.3	95.1	
Max accuracy	98.4	103.2	102.8	
Sum absolute error	222.3	38.0	39.9	Total absolute error over all three QC levels: 300.3

Number of standard curve points dropped over 30 runs (480 standard curve points total): 12

Table 3: n=30 runs, n= 6 replicates per QC level 1000 fold range, high QC is at ULOQ, mid QC ½ of the ULOQ , 8 standard curve levels, each level in duplicate.CV for standards and QCs: 1.3 - 4.6% CV for non-LLOQ standards, 6.8% CV at the LLOQ. Mid QC 2% CV, ULOQ QC 2.8% CV. To simulate non-linearity the PARs were calculated at 90% of the 400, 800, and 1000 std level, and at 90% of the 500 and 1000 QC level, but accuracy was based on the unadjusted level.

Non-Linearity and High Variance Summary

$1/x^2$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	103.3	94. 8	94.6	
Min accuracy	86.8	87.0	86.5	
Max accuracy	123.4	101.0	102.4	
Sum absolute error	200.0	168.1	173.9	Total absolute error over all three QC levels: 542

Number of standard curve points dropped over 30 runs (480 standard curve points total): 69

$1/x$ weighting	LLOQ QC	MID QC	HIGH QC	
Grand Mean accuracy	89.9	102.7	102.4	
Min accuracy	69.6	91.7	92.7	
Max accuracy	121.0	115.6	114.6	
Sum absolute error	376.1	148.2	118.2	Total absolute error over all three QC levels: 643

Number of standard curve points dropped over 30 runs (480 standard curve points total): 96

Table 4: n=30 runs, n= 6 replicates per QC level 1000 fold range, high QC is at ULOQ, mid QC ½ of the ULOQ, 8 standard curve levels, each level in duplicate. CV for standards and QCs: 8% CV for non-LLOQ standards, 15% CV at the LLOQ. To simulate non-linearity the PARs were calculated at 90% of the 400, 800, and 1000 std level, and at 90% of the 500 and 1000 QC level, but accuracy was based on the unadjusted level.

Results and Discussion

In a typical calibration curve the variance (SD squared) increases as the concentration increases. See Figure 1 which is actual data showing how variance changes exponentially over a 1000 fold range. A 1000 fold standard curve range can have over a 150,000 fold change in variance between the low and high standard and this is the reason a non-weighted curve does not work. A non-weighted regression is drawn to minimize the sum of the squared vertical distances of all the standard points from the regression line. This favors the higher variance of the higher standards leading to good accuracy for the highest standards but very poor accuracy for the lowest standards.

Figure 2 shows the sum of the error for the standard curve points for a $1/x^2$ and $1/x$ weighted linear regression. Note that the sum of the % difference from the expected is 0 for the $1/x^2$ curve.

Table 1 displays the mean accuracy and range for the low variance (1.3-4.6% CV for non-LLOQ standards, 6.8% CV at the LLOQ) scenario over 30 simulated runs. Both weightings show good accuracy and a narrow accuracy range at all three QC levels.

Table 2 displays the mean accuracy and range for a high variance scenario (8% CV for non-LLOQ standards, 15% CV at the LLOQ) over 30 simulated runs. The grand mean accuracies are essentially the same for both weightings and the accuracy spread is the same at the mid and high QC levels. For both weightings 1 standard curve point was eliminated per run on average. The LLOQ accuracy range is 88% to 115% for the $1/x^2$ weighting and 77% to 121% for the $1/x$ weighting. For the $1/x$ weighting the mean of two of the 30 LLOQ sets are outside the 20% range. In this high variance situation the $1/x^2$ weighting provides better results at the LLOQ. The total absolute error over all three QC levels is lower for the $1/x^2$ weighting.

Table 3 displays the mean accuracy and range for the low variance scenario but with non-linearity over 30 simulated runs. The accuracy at the mid and high QC levels are 100,100 % for $1/x$ and 93, 94% for $1/x^2$ weighting. The accuracy at the LLOQ level averages 104%, range 99% to 108%, for the $1/x^2$ weighting and averages 93%, range 83% to 98% for the $1/x$ weighing. No standard curve points were rejected for the $1/x^2$ weighting and 12 points (average 0.4 per run) were rejected for the $1/x$ weighting. Here the $1/x$ weighting provides for better accuracy in the upper end of the curve, but lower accuracy with a broader accuracy range at the LLOQ.

Table 4 displays the mean accuracy and range for high variance and non-linearity conditions over 30 simulated runs. For the $1/x^2$ weighting the mid and high QCs all pass with mean accuracies of 95% and 95%, with ranges of 87% to 101% and 86% to 102 % for the mid and high QC levels respectively. All accuracies are within 15%. The same QCs with $1/x$ weighting have mean accuracies of 103% and 102% , with ranges of 92% to 116% and 93% to 115% for the mid and high QC levels respectively. One of mid QC means was outside 15% accuracy. The accuracy range is better at the mid and high QC levels with $1/x^2$ weighting. At the LLOQ level the accuracy for the $1/x^2$ weighting is 103%, range 87% to 123%. One of the LLOQ means is greater than 20% accuracy. For the $1/x$ weighting the mean accuracy was 90% with range of 70% to 121%. Eight of the thirty means show greater than 20% accuracy. An average of 2.3 standard curve points were rejected for the $1/x^2$ weighting and an average of 3.2 standard curve points rejected for the$1/x$ weighting. The $1/x^2$ weighting provides for much better results at the LLOQ level. The total absolute error over all three QC levels is less with the $1/x^2$ weighting.

Conclusion

If the extraction and LC/MS performance are well behaved and the variance is low (1.3-4.6% CV, 6.8% CV at the LLOQ), then either a $1/x^2$ or $1/x$ weighting appears to work equally well. When the variance increases (8% CV, 15% CV at the LLOQ) the $1/x^2$ weighting is more rugged. In non-linearity situations with low variance the $1/x$ weighting provides for better accuracy in the upper end of the curve but less accuracy at the LLOQ level. For high variance and non-linearity situations the $1/x^2$ weighting shows much better results at the LLOQ level and a lower total absolute error.

When the precision or linearity is no longer well behaved the $1/x^2$ weighting provides for much better results at the LLOQ level. $1/x$ weighting does provide for more accurate results in the upper end of the curve if the results are relatively precise but non-linear. If it is known at the time of validation that most study sample results will be in the upper end of the standard curve and the method may show some non-linearity but expected to remain low, then this is a reason to choose a $1/x$ weighting. Otherwise there does not appear to be any benefit to using a $1/x$ weighting.

References

1) Guidance for Industry. Bioanalytical Method Validation. United States Food and Drug Administration, Washington, DC, 2001.
2) Bansal, S. and DeStefano, A., AAPS Journal, 9(1) Article 11, 2007.
3) Guideline on Validation of Bioanalytical Methods, EMEA, 2009.
4) Almeida, A.M., et al., Journal of Chromatography B, 774, pg 215-22, 2002.
5) J.W. Dolan, LCGC, 22(2), 112-117, 2004.
6) J.W. Dolan, LCGC, 27(7), 534-540, 2009.