Selecting the Correct Weighting Factors for Linear and Quadratic Calibration Curves with Least-Squares Regression Algorithm in Bioanalytical LC-MS/MS Assays, and Impacts of Using Incorrect Weighting Factors on Curve Stability, Data Quality and Assay Performance

Huidong Gu*, Guowen Liu, Jian Wang, Anne-Françoise Aubry and Mark E. Arnold

Bioanalytical Sciences, Research & Development, Bristol-Myers Squibb Company
Route 206 & Province Line Road, Princeton, NJ 08543

* Corresponding author: E-mail: huidong.gu@bms.com, Tel: 609-252-7636, Fax: 609-252-6171

<u>Supporting Information A:</u> Linear and quadratic regression with least-squares algorithm revisit

Linear and quadratic regression with least-squares algorithm

In statistics, regression is a technique for modeling and analyzing several variables to determine the relationship between a dependent variable and one or more independent variables. Linear and parabolic (quadratic) regressions with the least-squares algorithm are the two commonly used equations for calibration curve fitting for LC-MS/MS assays. The term "least-squares" indicates that, for the equation utilized, the sum of squared residuals for all observed values utilized in the regression is a minimum. A residual is the difference between an observed value and the value predicted by the selected equation[5].

It is not difficult to find the solution for the linear or quadratic regression coefficients for a given set of data, (x_1, y_1), (x_2, y_2), ... (x_n, y_n) using the least squares algorithm[5]. For the linear fit, a straight line (1) is used to approximate the given set of data so that the sum of the squared residuals (Π) is a minimum (2).

$$y = a + bx \tag{1}$$

$$\Pi = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2 \tag{2}$$

The Π_{min} is found when the unknown coefficients *a* and *b* yield zero first derivatives (3-4):

$$\frac{\partial \Pi}{\partial a} = 2\sum_{i=1}^{n} [y_i - (a + bx_i)] = 0 \tag{3}$$

$$\frac{\partial \Pi}{\partial b} = 2\sum_{i=1}^{n} xi[y_i - (a + bx_i)] = 0 \tag{4}$$

Therefore, *a* and *b* can be calculated as below (5-6):

$$a = \frac{(\sum y_i)(\sum x_i^{\,2}) - (\sum x_i)(\sum x_i y_i)}{n\sum x_i^{\,2} - (\sum x_i)^2} \tag{5}$$

$$b = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^{\,2} - (\sum x_i)^2} \tag{6}$$

Where $\sum$ stands for $\sum_{i=1}^{n}$

Similarly, for the quadratic fit, a quadratic equation (7) is used to approximate the given set of data so that Π is a minimum (8).

$$y = a + bx + cx^2 \tag{7}$$

$$\prod = \sum_{i=1}^{n} [y_i - (a + bx_i + cx_i^2)]^2 \tag{8}$$

The Π_{min} is found when the unknown coefficients *a*, *b* and *c* yield zero first derivatives. Therefore, *a*, *b* and *c* can be obtained similarly as for the linear fit. For both equations, the estimates for *a*, *b*, and *c* are unique.

Linear and quadratic regression with weighted least squares algorithm

The assumptions for the linear and quadratic regressions with least-squares algorithm described above are that the errors are independent from each other (random), uncorrelated with the independent variable (x), have equal variance (σ^2) or standard deviation (σ) across the different levels of x (homoscedasticity[14]). For LC-MS/MS bioanalytical assays, a calibration curve is established between concentration (x) and instrument response (y). It is reasonable to assume that the instrument response errors of different acquisitions from samples with the same concentration are uncorrelated with each other and have a normal error distribution. However, a quick review of any set of real life standard (STD) or quality control (QC) data from assay validation or sample analysis will reveal that the instrument response errors at different concentration levels are actually correlated with the independent variable - concentration, and have a larger variance (σ^2) or standard deviation (σ) at higher concentrations as long as the sample size at each concentration level is large enough to provide a rough estimate of variance. This is a clear indication that the instrument responses have heteroscedatic error[15].

To address the heteroscedatic errors, a modified least squares algorithm, weighted least squares (9-10), is warranted to establish calibration curves in LC-MS/MS assays. The idea is to assign an appropriate weight at each concentration level which reflects the different measurement uncertainties at different concentration levels.[17] When the measurements are uncorrelated and have different uncertainties, Aitken demonstrated that the best unbiased estimates could be obtained if each weight used is equal to the reciprocal of the variance of the measurement.[18,19] This is easy to understand since the importance of each squared residual at all concentration levels are normalized/equalized with the reciprocal of the variance at the corresponding concentration level[18,20,21].

$$\prod = \sum_{i=1}^{n} W_i [y_i - (a + bx_i)]^2 = \min \tag{9}$$

$$\prod = \sum_{i=1}^{n} W_i [y_i - (a + bx_i + cx_i^2)]^2 = \min \tag{10}$$

Where $W_i = 1/\sigma_i^2$ and σ_i is the standard deviation of the instrument responses at the corresponding concentration level. The standard deviations for instrument responses are calculated separately at different concentration levels.

The weighting factor is used to equalize the importance or influence of the errors at different concentration levels on the estimation of the unknown coefficients *a* and *b* (for linear regression) or *a*, *b* and *c* (for quadratic regression). It is easy to see that the least squares algorithm is just a special case of the weighted least squares algorithm when the instrument responses at different concentrations have equal variances or standard deviations ($\sigma_1 = \sigma_2 = \cdots \sigma_i = \cdots = \sigma_n$), and, therefore, the weighting factor, *Wi*, is cancelled out ($W_i = 1$ at each concentration). By using the weighted least squares with weights which are inversely proportional to the variance at each level of the independent variable, the most precise parameter estimates could be obtained.[18-20] This is the beauty of the weighted least-squares algorithm. However, this is also its major disadvantage since the algorithm is based on the assumption that the

weights are known exactly. In real applications, including LC-MS/MS bioanalytical assays, this is almost never the case. Therefore, estimated weights have to be used. Fortunately, it was reported that normally small variations of estimated weights from their "true" weights do not affect the regression results. However, the results could be badly impacted if the weights are estimated from small data sets without enough representative samples[20]. In other words, data sets with reasonable size (≥5) should be used to make close estimation although the exact weights are not needed. One thing worth mentioning is that, in some cases, for the qualitative and quantitative approaches discussed in the paper, STD data from multiple LC-MS/MS runs may not be suitable for both approaches if the between run variance is too large to show the true relationship between σ and x. In this case, QC data within a run should be used to calculate σ as long as there are enough different QC concentration levels (≥5) and enough replicates (≥5) at each concentration level.

<u>Supporting Information B:</u> Figures for the theoretical relationships between σ and x, and σ^2 and x to justify the selection of 1, 1/x or $1/x^2$ weighting factors

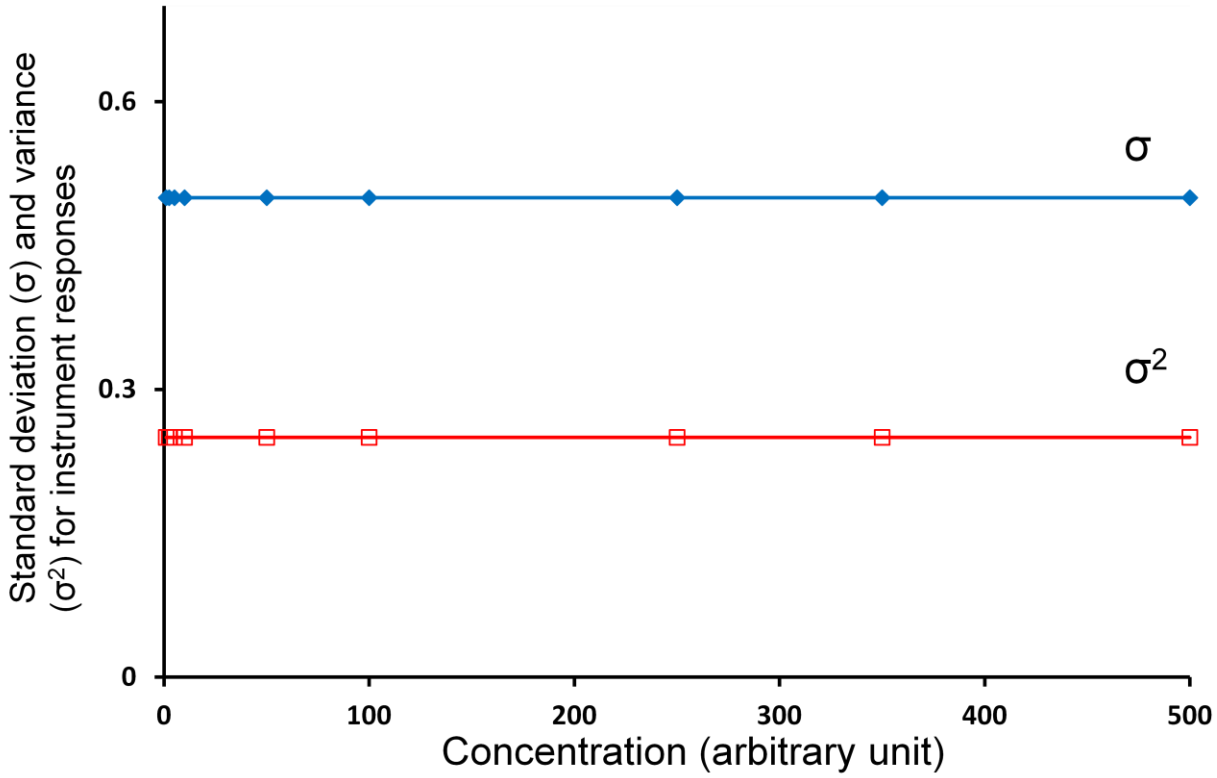

Figure SI-1a. Theoretical relationship between σ and x, and between σ^2 and x to justify the selection of *1* as the weighting factor (STD curve range: 1 - 500)

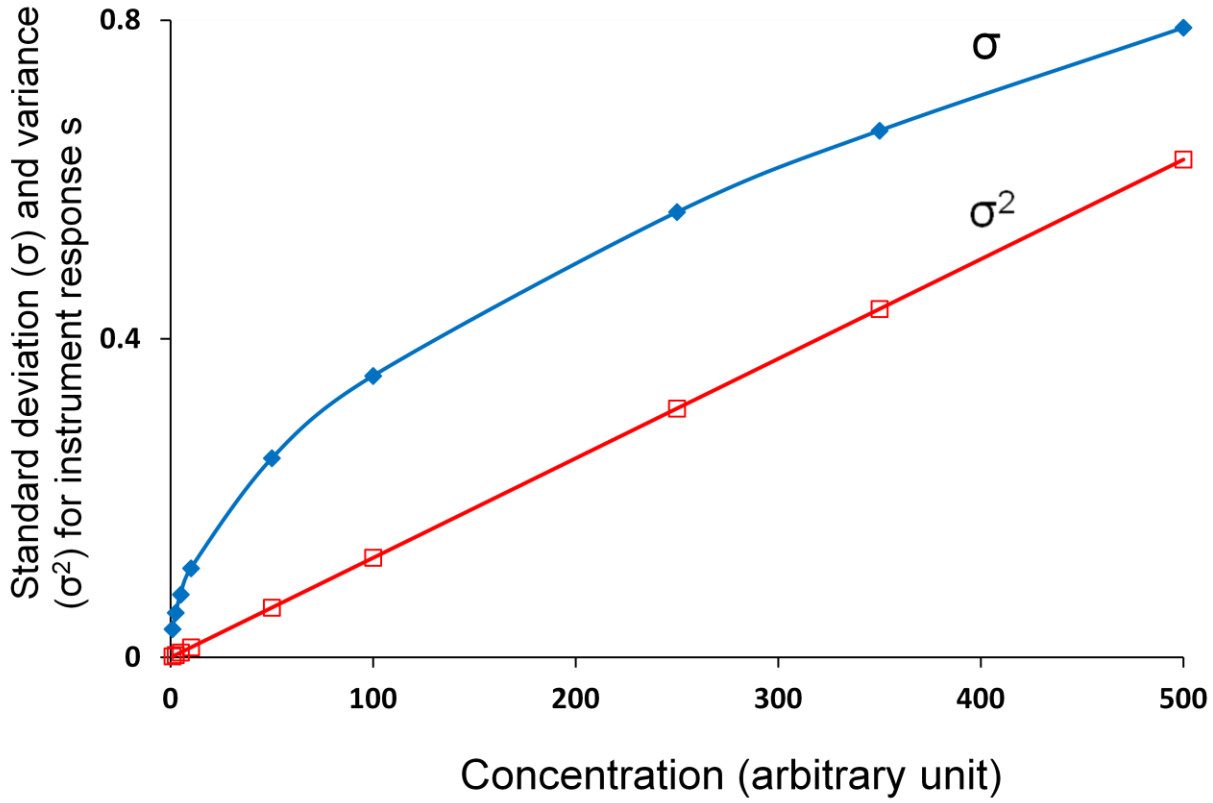

Figure SI-1b. Theoretical relationship between σ and x, and between σ^2 and x to justify the selection of *1/x* as the weighting factor (STD curve range: 1 - 500)

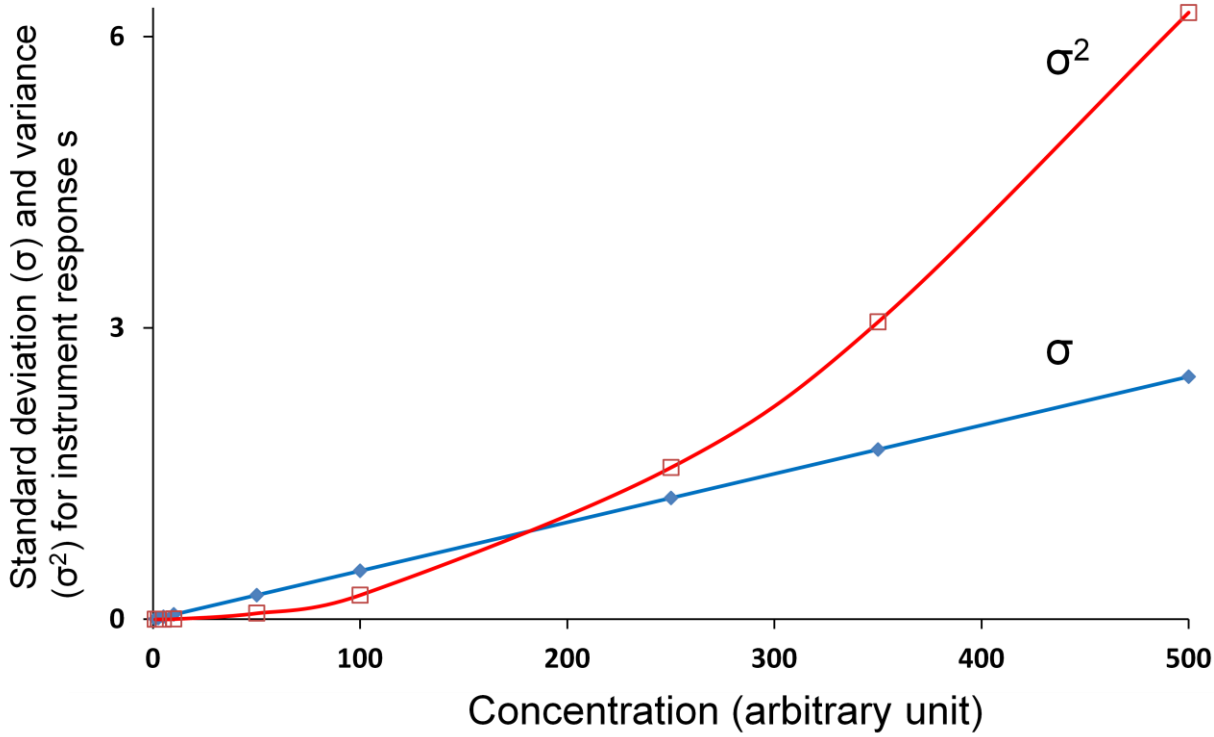

Figure SI-1c. Theoretical relationship between σ and x, and between σ^2 and x to justify the selection of $1/x^2$ as the weighting factor (STD curve range: 1 - 500)

<u>Supporting Information C:</u> Impact of STD calibrator spacing on curve stability

For the example used in the discussion of curve stability in this paper, the spacing between calibrators is not equal, with narrower spacing at low end and much wider spacing at higher end. To investigate the effect of calibrator spacing on the curve stability, the adjusted STD instrument responses with equal spacing calibrators shown in Table SI-1a were obtained by proportionally adjusting the instrument responses for run 1 in Table 3a with the corresponding concentrations. The curve stability was then examined in the same way as discussed in the curve stability section, and the findings are listed below:

- When the correct weighting factor of $1/x^2$ was used, the curve was always stable regardless the calibrator spacing. Actually the curves generated with the original validation data (Table 3a, Run 1) and the adjusted data (Table SI-1a) were overlapped nicely, further demonstrating that the calibrator spacing has no impact on the curve when the correct weighting factor is used.
- As expected, the curves were not stable when the wrong weighting factor 1 or $1/x$ was used. For the example discussed here, none of three runs could pass with the equal spacing (Table SI-1a, SI-1b and SI-1c, Run 2 and Run 3 results not shown) when the weighting factor of 1 was used as the lowest STDs were totally off. The exact impact from the different calibrator spacings on the curve stability is complicated and related to the exact calibrator arrangement. Fortunately, this is not an issue when a correct weighting factor is used.

Table SI-1a, Adjusted STD instrument responses with equal spacing calibrators

Conc. (ng/mL)	0.2	28.7	57.2	85.7	114.2	142.8	171.4	200
Run 1	0.045576	6.48455	12.82531	18.99024	24.62583	29.71189	35.32429	41.30729
	0.048468	7.00309	12.96788	19.39751	25.56505	30.61196	36.98769	42.26123

Note: The adjusted STD instrument responses were obtained by proportionally adjusting the instrument responses for run 1 in Table 3a with the corresponding concentrations. For example: 0.090377*28.7/0.4 = 6.48455, which is the first STD instrument response for 28.7 ng/mL.

Table SI-1b, Linear regression results for STD curve using adjusted instrument responses with equal spacing calibrators

Conc. (ng/mL)	Run 1 (Dev%*)		
	Weighting factor		
	1	1/x	$1/x^2$
0.2	-1866.4	-17.2	-3.3
0.2	-1859.4	-10.5	3.2
28.7	-4.1	5.0	3.0
28.7	4.7	13.4	11.2
57.2	1.6	4.2	2.2
57.2	2.8	5.4	3.4
85.7	2.5	3.0	1.0
85.7	4.8	5.2	3.2
114.2	0.8	0.3	-1.7
114.2	4.7	4.1	2.1
142.8	-2.2	-3.2	-5.1
142.8	0.8	-0.3	-2.2
171.4	-2.8	-4.1	-6.0
171.4	1.9	0.4	-1.6
200.0	-2.2	-3.9	-5.8
200.0	0.1	-1.7	-3.6

* Results expressed in percentage deviation from nominal concentrations

Table SI-1c, Curve stability test for STD curve using adjusted instrument responses with equal spacing calibrators

Conc. (ng/mL)	Run 1 (Dev%*)		
	Weighting factor		
	1	1/x	$1/x^2$
0.2	-3625.4	-24.5	-3.4
0.2	-3618.0	-17.6	3.3
28.7	-10.1	7.3	4.3
28.7	-0.8	15.9	12.7
57.2	1.7	6.6	3.5
57.2	3.0	7.8	4.7
85.7	4.6	5.4	2.3
85.7	7.0	7.7	4.5
114.2	3.7	2.6	-0.4
114.2	7.9	6.5	3.4
142.8	1.1	-1.0	-3.9
142.8	4.3	2.0	-1.0
171.4	0.9	-1.9	-4.8
171.4	5.9	2.7	-0.3
200	1.8	-1.7	-4.6
200	-15.6 (34.5)	-17.9 (34.5)	-20.3 (34.5)
Curve stability	Extremely unstable	Unstable	stable

* Results expressed in percentage deviation from nominal concentrations

Note: instrument response bias listed in parentheses, please see Table SI-1a for the original instrument responses

<u>Supporting Information D:</u> QC performance

Table SI-2a, Calculated QC concentrations using original STD curves in run 3 (1, 1/x and $1/x^2$ weighting), showing that "good" STD curves generate "good" QC

Weighting factor for the STD curve	STD Curve Pass or Fail (Table 3b)	QC Pass or Fail	Calculated QC concentration using original STD curves (Table 3a, 3b)											
			LLOQ QC		Low QC		GM QC		Mid QC		High QC		Dilution QC	
			0.200 ng/mL	%Dev	0.600 ng/mL	%Dev	8.00 ng/mL	%Dev	100 ng/mL	%Dev	160 ng/mL	%Dev	10000 ng/mL	%Dev
1	Pass with good accuracy	Pass with good accuracy	0.243	21.5	0.623	3.8	8.14	1.8	102	2.0	158	-1.3	9531	-4.7
			0.217	8.5	0.631	5.2	8.04	0.5	103	3.0	157	-1.9	9365	-6.4
			0.214	7.0	0.650	8.3	8.01	0.1	101	1.0	161	0.6	9122	-8.8
			0.213	6.5	0.633	5.5	7.98	-0.2	98	-2.0	158	-1.3	9460	-5.4
			0.217	8.5	0.584	-2.7	8.11	1.4	101	1.0	157	-1.9	9334	-6.7
			0.222	11.0	0.634	5.7	8.12	1.5	101	1.0	158	-1.3	9497	-5.0
1/x	Pass with good accuracy	Pass with good accuracy	0.228	14.0	0.609	1.5	8.13	1.6	102	2.0	158	-1.3	9531	-4.7
			0.203	1.5	0.617	2.8	8.03	0.4	103	3.0	157	-1.9	9366	-6.3
			0.200	0.0	0.635	5.8	8.00	0.0	101	1.0	161	0.6	9123	-8.8
			0.199	-0.5	0.619	3.2	7.97	-0.4	98	-2.0	158	-1.3	9461	-5.4
			0.203	1.5	0.570	-5.0	8.09	1.1	101	1.0	157	-1.9	9335	-6.7
			0.208	4.0	0.620	3.3	8.11	1.4	101	1.0	158	-1.3	9498	-5.0
$1/x^2$	Pass with good accuracy	Pass with good accuracy	0.232	16.0	0.611	1.8	8.09	1.1	102	2.0	157	-1.9	9480	-5.2
			0.206	3.0	0.618	3.0	7.99	-0.1	103	3.0	157	-1.9	9315	-6.9
			0.203	1.5	0.637	6.2	7.96	-0.5	100	0.0	160	0.0	9073	-9.3
			0.202	1.0	0.620	3.3	7.93	-0.9	97	-3.0	157	-1.9	9410	-5.9
			0.207	3.5	0.572	-4.7	8.05	0.6	101	1.0	156	-2.5	9284	-7.2
			0.212	6.0	0.621	3.5	8.07	0.9	100	0.0	157	-1.9	9446	-5.5

Table SI-2b, Calculated QC concentrations using STD curves in run 3 (1, 1/x and $1/x^2$ weighting) with one small (~1.5%) instrument response bias for one STD sample, showing that "good" STD curves generate "good" QC and "bad" STD curves generate "bad" QC.

Weighting factor for the STD curve	STD Curve Pass or Fail (Table 3c)	QC Pass or Fail	Calculated QC concentrations using STD curves with one small (~1.5%) instrument response biases (Table 3a, 3c)											
			LLOQ QC		Low QC		GM QC		Mid QC		High QC		Dilution QC	
			0.200 ng/mL	%Dev	0.600 ng/mL	%Dev	8.00 ng/mL	%Dev	100 ng/mL	%Dev	160 ng/mL	%Dev	10000 ng/mL	%Dev
1	Fail badly	Fail badly	0.151	-24.5	0.534	-11.0	8.09	1.1	103	3.0	159	-0.6	9568	-4.3
			0.126	-37.0	0.542	-9.7	7.99	-0.1	104	4.0	158	-1.3	9402	-6.0
			0.123	-38.5	0.560	-6.7	7.95	-0.6	101	1.0	162	1.3	9157	-8.4
			0.122	-39.0	0.544	-9.3	7.93	-0.9	98	-2.0	159	-0.6	9497	-5.0
			0.126	-37.0	0.495	-17.5	8.05	0.6	102	2.0	158	-1.3	9370	-6.3
			0.131	-34.5	0.544	-9.3	8.07	0.9	101	1.0	159	-0.6	9534	-4.7
1/x	Pass with good accuracy	Pass with good accuracy	0.227	13.5	0.609	1.5	8.16	2.0	103	3.0	158	-1.3	9564	-4.4
			0.201	0.5	0.617	2.8	8.05	0.6	103	3.0	158	-1.3	9397	-6.0
			0.198	-1.0	0.636	6.0	8.02	0.2	101	1.0	162	1.3	9154	-8.5
			0.197	-1.5	0.619	3.2	7.99	-0.1	98	-2.0	159	-0.6	9493	-5.1
			0.201	0.5	0.570	-5.0	8.12	1.5	102	2.0	157	-1.9	9366	-6.3
			0.207	3.5	0.620	3.3	8.13	1.6	101	1.0	159	-0.6	9530	-4.7
$1/x^2$	Pass with good accuracy	Pass with good accuracy	0.232	16.0	0.611	1.8	8.10	1.3	102	2.0	157	-1.9	9494	-5.1
			0.206	3.0	0.619	3.2	8.00	0.0	103	3.0	157	-1.9	9329	-6.7
			0.203	1.5	0.637	6.2	7.97	-0.4	100	0.0	160	0.0	9087	-9.1
			0.202	1.0	0.621	3.5	7.94	-0.7	98	-2.0	158	-1.3	9424	-5.8
			0.207	3.5	0.572	-4.7	8.07	0.9	101	1.0	156	-2.5	9298	-7.0
			0.212	6.0	0.621	3.5	8.08	1.0	101	1.0	157	-1.9	9460	-5.4

Table SI-2c, Calculated QC concentrations using STD curves in run 3 (1, 1/x and $1/x^2$ weighting) with three big (~20%) instrument response biases* for three STD samples, showing that "good" STD curves generate "good" QC and "bad" STD curves generate "bad" QC.

Weighting factor for the STD curve	STD Curve Pass or Fail (Table 3c)	QC Pass or Fail	Calculated QC concentrations using STD curves with three big (~20%) instrument response biases (Table 3a)*											
			LLOQ QC		Low QC		GM QC		Mid QC		High QC		Dilution QC	
			0.200 ng/mL	%Dev	0.600 ng/mL	%Dev	8.00 ng/mL	%Dev	100 ng/mL	%Dev	160 ng/mL	%Dev	10000 ng/mL	%Dev
1	Fail badly	Fail badly	-2.515	-1357.4	-2.064	-444.0	6.841	-14.5	118.341	18.3	184.208	15.1	11007.031	10.1
			-2.545	-1372.6	-2.055	-442.5	6.718	-16.0	119.313	19.3	183.592	14.7	10810.924	8.1
			-2.549	-1374.3	-2.033	-438.8	6.681	-16.5	116.366	16.4	187.797	17.4	10523.182	5.2
			-2.550	-1375.0	-2.052	-442.1	6.650	-16.9	113.137	13.1	184.601	15.4	10923.780	9.2
			-2.545	-1372.5	-2.110	-451.7	6.797	-15.0	117.219	17.2	182.954	14.3	10774.132	7.7
			-2.539	-1369.3	-2.052	-441.9	6.813	-14.8	116.831	16.8	184.265	15.2	10967.556	9.7
1/x	Fail badly	Fail badly	0.179	-10.5	0.611	1.8	9.136	14.2	115.883	15.9	178.942	11.8	10796.461	8.0
			0.150	-25.0	0.619	3.2	9.018	12.7	116.814	16.8	178.352	11.5	10608.714	6.1
			0.147	-26.7	0.640	6.7	8.983	12.3	113.991	14.0	182.378	14.0	10333.238	3.3
			0.145	-27.4	0.622	3.6	8.953	11.9	110.901	10.9	179.318	12.1	10716.759	7.2
			0.150	-25.0	0.566	-5.6	9.094	13.7	114.808	14.8	177.741	11.1	10573.490	5.7
			0.156	-21.9	0.622	3.7	9.109	13.9	114.437	14.4	178.996	11.9	10758.669	7.6
1/x2	Pass with good accuracy	Pass with good accuracy	0.229	14.7	0.630	5.0	8.543	6.8	107.622	7.6	166.152	3.8	10027.288	0.3
			0.203	1.3	0.638	6.4	8.434	5.4	108.486	8.5	165.605	3.5	9853.027	-1.5
			0.199	-0.3	0.658	9.6	8.401	5.0	105.867	5.9	169.341	5.8	9597.339	-4.0
			0.198	-0.9	0.640	6.7	8.373	4.7	102.998	3.0	166.501	4.1	9953.311	-0.5
			0.203	1.3	0.589	-1.9	8.504	6.3	106.625	6.6	165.038	3.1	9820.334	-1.8
			0.208	4.1	0.641	6.8	8.518	6.5	106.280	6.3	166.202	3.9	9992.211	-0.1

*: Instrument biases for three STD samples (Please see Table 3a for the original instrument responses)

1. The second STD at 160 ng/mL: 28.9
2. The first STD at 200 ng/mL: 34.6
3. The second STD at 200 ng/mL: 37.5

<u>Supporting Information E:</u> Weighting factor for quadratic and other non-linear curves

In theory, for any curve regression using least-square algorithm, the weighting factor is only related to the standard deviation of the instrument responses at each concentration (or response) level. The standard deviation of the instrument responses at each concentration level comes from the propagation/accumulation of uncertainties from sample preparation to MS detection, including the uncertainties from sample transfer, extraction, non-specific adsorption, concentration dependent recovery and matrix effect, instrument or ionization saturation, MS detection etc., at each concentration level. Therefore, for quadratic curves used in LC-MS/MS assays (or even 4-PL, 5-PL curves in ligand binding assays), $1/y^2$ should be used as long as there is a linear relationship between the standard deviation of the instrument responses and the instrument response. In most cases, such linear relationship should still hold for both linear and non-linear curves in LC-MS/MS assays. For example, one assay (omeprazole in human urine) in the survey showed very obvious quadratic bending due to detection saturation and a quadratic curve was used for this assay. Using either the proposed qualitative or quantitative approach, a linear relationship was observed between the standard deviation of instrument responses and the instrument response, and therefore, the use of $1/x^2$ ($1/y^2$ is better) was justified. Comparing to the corresponding linear indicators generated by using the quantitative approach when $1/x^2$ was used as the weighting factor, smaller linear indicators were obtained when $1/y^2$ was used as the weighting factor for all of three LC-MS/MS assays with quadratic bending surveyed in this work, further demonstrating $1/y^2$ should be used as the weighting factor in LC-MS/MS assays.

Although all of the assays discussed in this paper had constant relative responses, several assays with quadratic curves were also surveyed in this work, and the relative responses for these assays were not constant. For the omeprazole in human urine assay discussed above, the relative response was about 0.0092 per ng/mL at LLOQ of 1 ng/mL, and the relative response gradually decreased to 0.0056 per

ng/mL at ULQ of 1000 ng/mL. For these assays without constant relative responses, a linear relationship was still found between standard deviation of instrument responses and the instrument response, and the use of $1/y^2$ was justified.

However, when curves are severely bended due to ionization or detection saturation, the standard deviations of the instrument responses will not increase linearly/close linearly (or in some extreme cases, it will decrease a little bit) with the increase of the instrument responses. In these situations, the selection of $1/y^2$ may then be questionable. Therefore, the correct weighting factor should be selected using the proposed procedures in this manuscript. Fortunately, severely bended quadratic curves are rarely used in LC-MS/MS assays.

Note: Please refer to the original paper for the table and literature references.