Taylor & Francis
Taylor & Francis Group

Journal of Liquid Chromatography & Related Technologies

ISSN: 1082-6076 (Print) 1520-572X (Online) Journal homepage: https://www.tandfonline.com/loi/ljlc20

A note on the use of workstation software programs for quantification

William Craig Byrdwell

To cite this article: William Craig Byrdwell (2019) A note on the use of workstation software programs for quantification, Journal of Liquid Chromatography & Related Technologies, 42:17-18, 570-574, DOI: 10.1080/10826076.2019.1637350

To link to this article: https://doi.org/10.1080/10826076.2019.1637350

JOURNAL OF LIQUID CHROMATOGRAPHY & RELATED TECHNOLOGIES
2019, VOL. 42, NOS. 17–18, 570–574
https://doi.org/10.1080/10826076.2019.1637350

Taylor & Francis
Taylor & Francis Group

A note on the use of workstation software programs for quantification

William Craig Byrdwell

Agricultural Research Service, Food Composition and Methods Development Lab, Beltsville Human Nutrition Research Center, Beltsville, MD, USA

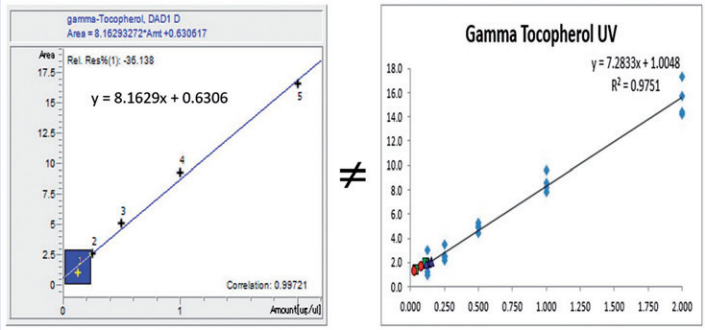
ABSTRACT

Most chromatographic and mass spectrometric instruments include workstations that allow data processing and target compound quantification. However, the software used for quantification does not all use common approaches for statistical treatment of data. Presented here is a brief description of three commonly used workstation software packages (WSPs) and the degree to which they obey widely accepted approaches to statistical treatment of data. The Thermo Fisher Scientific software packages Xcalibur and TraceFinder provided calibration lines and accompanying parameters that were similar or identical to those obtained by generic treatment using the Microsoft Excel "linest()" function, whereas the Agilent OpenLab ChemStation software did not. It is recommended to always perform calculations manually via spreadsheets for at least a few representative samples to allow better statistical treatment of data and to confirm whether the WSP used employs the commonly accepted approach to linear calibration.

KEYWORDS

Calibration curve; coefficient of determination; correlation coefficient; least squares; regression

GRAPHICAL ABSTRACT



Introduction

The ability to produce and use calibration curves (lines) for quantification of target compounds is built into many chromatography and mass spectrometry (and other spectroscopic instruments) workstation software programs (WSPs) or components. Calibration curves are normally constructed using simple least-squares (LS) regression to model the relationship between a dependent variable (i.e. the integrated peak area or detector response) versus an independent variable (i.e. the known concentration of a standard solution) as a line. As the name implies, the LS regression model minimizes the sum of the squares of the differences between observed values of detector response or integrated areas and the calculated values from the linear model. The equation for the resultant line is normally expressed in slope-intercept form as $y = mx + b$, where m is the slope and b is the intercept, with y being the signal response and x being the known concentration of a standard. Then, the concentration of a target compound can be calculated from its signal response, y, by solving for the concentration variable $x = (y-b)/m$.[1]

When reporting results from quantification by calibration curve, authors often list the line equation, in $y = mx + b$ form and the value of the coefficient of determination (CoD), r^2, which is the square of the Pearson correlation coefficient (Galton's function),[2,3] r, known simply as the correlation coefficient (CC). The CC has values from +1 (positive slope) to −1 (negative slope), while the CoD ranges from 0 to +1. Only calibration curves with positive slopes (increasing signal with increasing concentration) are discussed here. r^2 is one of the most commonly cited parameters to indicate a good fit of the data to a linear model, with values closer to 1 being assumed to be better.

CONTACT William Craig Byrdwell C.Byrdwell@ars.usda.gov Agricultural Research Service, Food Composition and Methods Development Lab, Beltsville Human Nutrition Research Center, U.S.D.A, 10300 Baltimore Ave, Beltsville, MD 20705, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ljlc.

There is much debate over the value and usefulness of r^2 as a descriptive parameter, with some requiring that it be cited[4] and others recommending against its use (or the corresponding use of r[5]) or pointing out that it is a poor measure of the curve fit.[6] It is not the goal of this communication to participate in that debate. Instead, we simply acknowledge the pragmatic fact that r^2 is a very widely used parameter in the analytical and chromatographic sciences to indicate a good fit of a set of calibration points to a line; that it is the parameter commonly used by WSPs to indicate linearity and that it is often reported by authors as a means of method validation. Given this reality, this note intends to point out instances when WSPs provide values equivalent to those that are generically calculated and when they do not. What precipitated this investigation was a calibration line from a WSP that looked "too good". It showed no spread of the points and gave a very high CoD. This caused us to manually calculate the results via spreadsheet to see if the same values could be replicated.

Materials and methods

Fat-soluble standards and three samples of standard reference material (SRM) 3278 from the National Institute of Standards and Technology (NIST) were analyzed in triplicate for tocopherols and neutral lipids (diacylglycerols and triacylglycerols) using the comprehensive two-dimensional liquid chromatography with quadruple parallel mass spectrometry (LC1MS2 × LC1MS2 = LC2MS4) approach recently reported.[7] The only data discussed here are data from the UV detector at 297 nm and fluorescence detector (FLD) at 330 nm for tocopherols, and selected ion monitoring (SIM) and selected reaction monitoring (SRM) by atmospheric pressure chemical ionization (APCI) mass spectrometry (MS). Details for acquisition parameters, chromatographic conditions, and detector settings were given in the Supporting Information to our previous report.[7]

All peaks from each detector were manually integrated using the WSP on each instrument. The raw integrated areas were output as a .pdf file from the Agilent OpenLab ChemStation (OLCS) and Trace Finder (TF) WSPs and as an "Excel Short Report" from the Xcalibur Quan Browser (XQB) WSP. Although the Agilent OLCS data were acquired and processed on software version C.01.07, the same files have been tested and the process and results have been found to be identical on the latest version, C.01.09. The data from the .pdfs were copied and pasted into a Microsoft Excel (ME) spreadsheet for further calculations. The data from the ME Short Report was incorporated directly into the calculation spreadsheet by copying and pasting the group of worksheets into the calculation worksheet template.

SRM 3278 contained α-tocopherol, which was not differentiated from the d_6-α-tocopherol by UV detection, so the external standard (ES) approach for γ-tocopherol was used for UV data. The internal standard (IS) method was used for MS data because normal and deuterium-labeled α-tocopherol were easily differentiated by mass. Since comparisons were made between WSP results and ME results for ES and IS approaches separately and ES results are not compared to IS results, the calibration method was irrelevant to the principles demonstrated and conclusions drawn from the effect of WSP processing.

Results

Figure 1 shows calibration lines from the UV and FLD detectors for quantification of γ-tocopherol by the external standard method using the Agilent OLCS WSP and from ME. Figure 2 shows APCI-MS analysis using the XQB WSP for SIM and SRM data and the corresponding ME spreadsheet plots. Figure 3 shows analysis of the same APCI-MS data using the TF WSP for SIM and SRM data and the corresponding ME spreadsheet plots. Ideally, the calibration line equations and r^2 values in the second columns in each figure would exactly match those parameters in the first column, indicating that the results by the WSP were exactly reproduced by manual calculations using the ME linest() function. The third column in each figure represents the effect of averaging the values at each concentration level before constructing the calibration line, essentially eliminating uncertainty in the points.

Figures 1B and 1E show that the ME plots, which included all raw data points, demonstrated the expected spread within the points, and gave CoDs that were lower (poorer) than the plot from the Agilent OLCS WSP (Figures 1A and 1D), which showed no spread in the data points. Figure 1B shows a CoD of 0.9751, which we would consider marginal for method validation, compared to 0.99721 from the OLCS WSP, which would be considered quite acceptable for method validation, and suitable for publication. Further calculations revealed that if the data at each point were averaged prior to construction of the calibration line (CL) (Figures 1C and 1F), it gave CoDs close to those produced by the OLCS WSP. It became evident that the OLCS did not use all data for the construction of the CL, but instead used average values at each concentration to construct the CL, hence the lack of a spread in the points in Figures 1A and 1D. Private communication with experts at Agilent confirmed that this is the case. However, simple averages were not used by the OLCS WSP, since the CL equations and CoDs in Figures 1A and 1D did not match the CL equations and CoDs in Figures 1C and 1F, respectively. Instead, OLCS uses a proprietary approach that is not readily replicated. Thus, CLs produced from the OLCS WSP cannot be directly reproduced by the generic ME linest() function. Unfortunately, using only averages to construct the CLs means that those CLs do not appear to be statistically rigorous and suitable for publication.

In all cases, the CoDs from OLCS CLs (Figures 1A and 1D) and from ME CLs produced by using the point averages (Figures 1C and 1F) were higher (better) than the CoDs produced by the conventional least-squares approach that included all raw points (Figures 1B and 1E). Although the CoDs were higher, the values calculated from the FLD approach (Figures 1D–1F) were similar between the OLCS and ME results, whereas UV data (Figures 1A–1C) differed

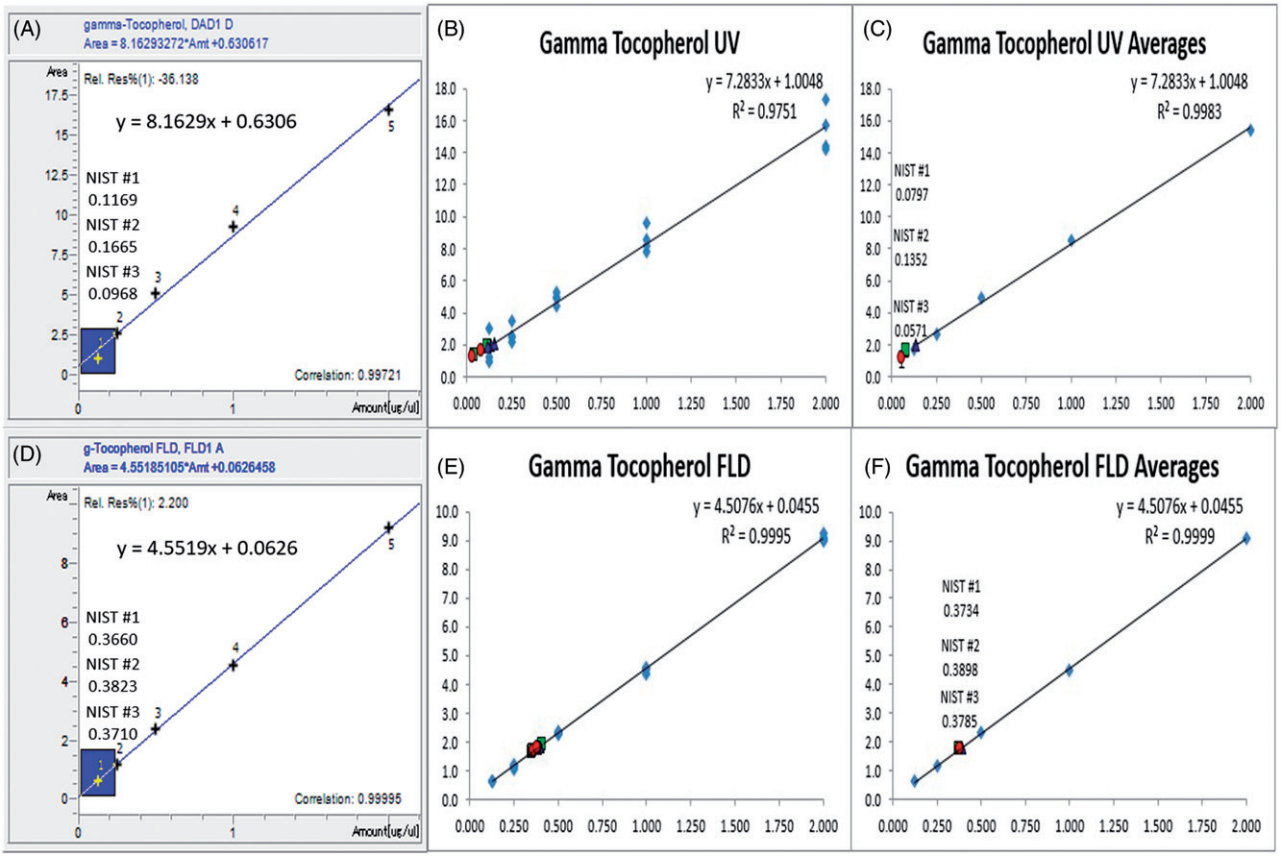


Figure 1. External standard calibration lines for γ-tocopherol obtained from Agilent OpenLab Chemstation (OLCS) and Microsoft Excel (ME) linest() function. (A) Agilent OLCS calibration line by UV at 297 nm; (B) ME UV calibration line showing all points; (C) ME UV calibration line from calibration level averages; (D) Agilent OLCS calibration line by fluorescence detection (FLD) at 330 nm; (E) ME FLD calibration line showing all points; (F) ME FLD calibration line from calibration level averages.

more substantially between OLCS and WE approaches. Thus, the primary effect of using the OLCS WSP for calculating CLs was that the uncertainty in the values was not represented (no spread in the data points) and the CoDs were overestimated. To properly represent the uncertainty and the CoDs, the CLs should be calculated via spreadsheet using the linest() function instead of using the OLCS WSP and the CoD should be reported from the ME calculations. This has other advantages, since the linest() function also provides other useful parameters, such as the standard error for the slope and intercept, standard for the estimate, the "f" statistic (to determine whether the observed relationship between the dependent and independent variables occurs by chance), as well as the regression sum of squares and the residuals sum of squares.

Comparisons of data processed using the Thermo Fisher Scientific XQB and TF WSPs to results from ME are shown in Figures 2 and 3. The XQB and TF WSPs, Figures 2A, 2D, 3A, and 3D, respectively, provided CLs with similar appearance (showing the spread of values at each point), identical calculated values and identical CoDs as the CLs produced by ME, Figures 2B and 2E and 3B and 3E, indicating that the WSPs produced statistically rigorous results equivalent to those calculated manually by spreadsheet. In the case of TF results, Figure 3, the equations for the CLs were identical to those produced by the generic ME approach. Interestingly, the XQB CLs in Figures 2A and 2D gave CL equations with slopes and intercepts that were exactly one-half of those from ME. On further investigation, it was found that this was because the IS areas appearing in the "ISTD Area" column in the XQB window were exactly two times the integrated areas of the IS for each sample, even though no dilution factor or other factor of 2 was stipulated. Thus, this appeared to be an error in the software. Nevertheless, since this was consistent throughout, the CoD of the CL was identical and the amounts calculated were identical to the results from the ME calculations, only the equation for the CL differed. Thus, the XQB and TF WSP results (calculated values and CoD) could be reported "as is". However, since additional regression parameters were produced by the linest() function, it is still beneficial to perform the calculations via spreadsheet.

These data were selected because they were less than ideal and demonstrated the dramatic effect of averaging values before constructing CLs. After normalizing for sample weight, only APCI-MS/MS SRM data produced values close to the certified value for γ-tocopherol, which is 111.5 ± 5.8 µg/g. The value produced by SRM using XQB was 139.8 ± 17.2 µg/g (uncertainty as the square root of the sum of the squares of the uncertainty in triplicate analyses of the three samples) and the value produced by SRM using TF was 139.1 ± 11.3 µg/g.

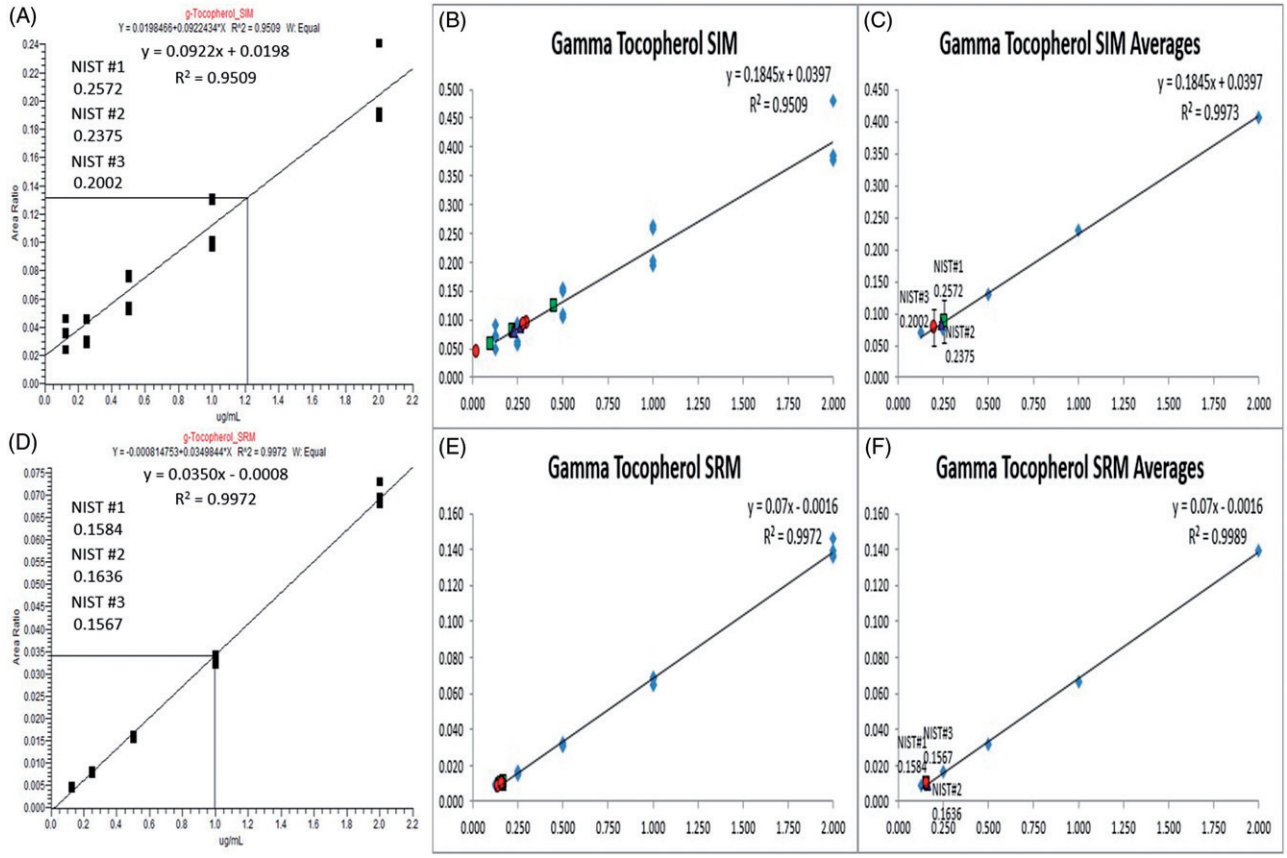


Figure 2. Thermo Fisher Scientific TSQ Vantage EMR APCI-MS calibration lines for γ-tocopherol obtained from Xcalibur Quan Browser (XQB) and Microsoft Excel (ME) linest() function. (A) XQB calibration line by selected ion monitoring (SIM) at *m/z* 416.37 + 417.37; (B) ME SIM calibration line showing all points; (C) ME SIM calibration line from calibration level averages; (D) XQB calibration line by selected reaction monitoring (SRM) for *m/z* 417.373→*m/z* 151.133; (E) ME SRM calibration line showing all points; (F) ME SRM calibration line from calibration level averages.

In all figures, the effect of using average values to construct the CLs instead of all raw values can be seen in comparison of the second columns of Figures 1–3 to the third columns. In all cases, the CoDs were higher, sometimes much higher, than those from the CLs that used the raw values. For instance, in Figure 1, the results from the UV detector (in which the vitamin D_3 peak overlapped the γ-tocopherol peak and required judicious manual peak splitting) showed a substantial increase in the CoD from $r^2 = 0.9751$ to $r^2 = 0.9983$ (the proprietary OLCS CoD was 0.9972). The FLD did not exhibit the peak overlap with vitamin D3 that was seen in the UV (since vitamin D3 does not fluoresce), so there was less uncertainty in the points and the CoDs were all higher. Similarly, APCI-MS SIM results for XQB (Figures 2A–2C) and TF (Figures 3A–2C) showed more spread in the points than SRM results (Figures 2D–2F and Figures 3D–3F), respectively, because SRM is more specific than SIM and less subject to noise and interfering peaks at the SIM mass. The CoDs showed dramatically better values by using point averages, as they went from r^2=0.9509, which we would find unacceptable for method validation, to r^2=0.9973 for the SIM results processed on the XQB WSP. Similarly, the results on the TF WSP went from r^2=0.9505 to 0.9976 for SIM results after point averaging. Thus, results that would be deemed unacceptable became acceptable by averaging the points prior to constructing the CLs, which is not a statistically rigorous practice for reporting results. Thus, there is a clear advantage for WSP providers to average values first, before calculating the CLs, because this gives higher CoDs, even though it is not a statistically rigorous approach. All results by SRM calculated using XQB (Figures 2D–2F) and TF (Figures 3D–3F) gave acceptable CoDs before point averaging.

Discussion

Some detection methods (e.g. UV, SIM) produce substantially more spread in the points than others (e.g. FLD, SRM). UV and SIM data clearly show that averaging the points first, to eliminate uncertainty at each point, caused data that should not be acceptable (i.e. r^2<0.98) to appear acceptable and valid (r^2>0.99). FLD and SRM provided more specificity, so there was less uncertainty in the detection method and so less difference between the properly treated data and the averaged-point data, making the improper treatment harder to identify. A key "red flag" that users should look for is the absence of any spread in data points in the CL. Even good data, such as the FLD data in Figure 1E, show a small amount of spread in the points, in contrast to Figure 1D.

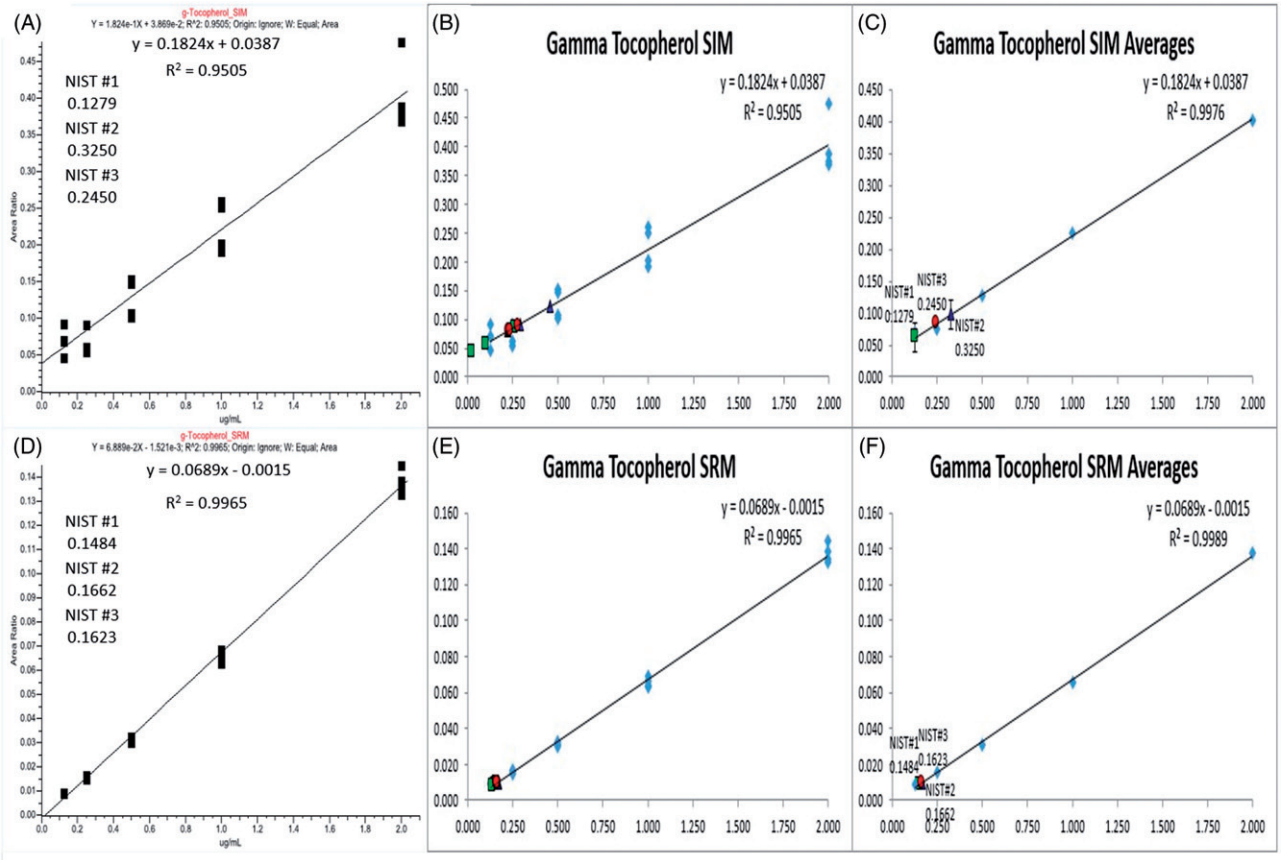


Figure 3. Thermo Fisher Scientific APCI-MS internal standard calibration lines for γ-tocopherol obtained from TraceFinder (TF) and Microsoft Excel (ME) linest() function. (A) TF calibration line by selected ion monitoring (SIM) at *m/z* 416.37 + 417.37; (B) ME SIM calibration line showing all points; (C) ME SIM calibration line from calibration level averages; (D) TF calibration line by selected reaction monitoring (SRM) for *m/z* 417.373→*m/z* 151.133; (E) ME SRM calibration line showing all points; (F) ME SRM calibration line from calibration level averages.

The important message of this note is that WSP results cannot be assumed to be accurate and statistically rigorous. In other words, the WSP cannot be treated as a "black box" that automatically produces publishable results. Instead, results should be calculated manually for at least a subset of samples and the results from whichever WSP is used should be verified by manual replication via spreadsheet calculation. Only then can the user know for sure whether the results they are producing are statistically valid. Whether the CoD is the best measure of good calibration or not, it is a pragmatic reality that users cite it as a parameter to attest to the quality of their CLs and they want the highest CoD they can get. Therefore, there is an incentive for WSP providers to use approaches that give the highest CoDs. It is up to users to verify that the results that they report follow widely accepted statistical practices.

Funding

This work was supported by the USDA Agricultural Research Service. Mention or use of specific products or brands do not represent or imply endorsement by the USDA.

ORCID

William Craig Byrdwell http://orcid.org/0000-0001-8241-428X

References

[1] Harris, M. C. *Quantitative Chemical Analysis*, 5th ed.; W.H. Freeman and Company: New York, **2001**.

[2] Pearson, K. Notes on Regression and Inheritance in the Case of Two Parents. *Proc. Royal Soc. London.* **1895**, *58*, 240–242.

[3] Galton, F. Regression Towards Mediocrity in Heredity Stature. *J. Anthropol. Inst. Gr. Brit. Ire.* **1886**, *15*, 246–263. DOI: 10.2307/2841583.

[4] Lang, T. A.; Altman, D. G. Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines". *Int. J. Nurs. Stud.* **2015**, *52*, 5–9. DOI: 10.1016/j.ijnurstu.2014.09.006.

[5] Analytical Methods Committee. Uses (Proper and Improper) of Correlation Coefficients. *Analyst.* **1988**, *113*, 1469–1471.

[6] Kiser, M. M.; Dolan, J. W. Selecting the Best Curve Fit. *LC GC N. Am.* **2004**, *22*, 112–117.

[7] Byrdwell, W. C. Comprehensive Dual Liquid Chromatography with Quadruple Mass Spectrometry (LC1MS2 x LC1MS2 = LC2MS4) for Analysis of *Parinari Curatellifolia* and Other Seed Oil Triacylglycerols. *Anal. Chem.* **2017**, *89*, 10537–10546. DOI: 10.1021/acs.analchem.7b02753.