Nonlinear regression improves accuracy of characterization of multiplexed mass spectrometric assays

January 16, 2018

Cyril Galitzine¹, Jarrett D. Egertson², Susan Abbatiello³, Clark M. Henderson⁴, Lindsay K. Pino², Michael MacCoss², Andrew N. Hoofnagle⁴,⁵, and Olga Vitek¹,⁶,*

¹College of Science, Northeastern University, Boston, MA USA
²Department of Genome Sciences, University of Washington, Seattle, WA USA
³Thermo Fisher Scientific, Cambridge, MA USA
⁴Department of Laboratory Medicine, University of Washington, Seattle, WA USA
⁵Department of Medicine, University of Washington, Seattle, WA USA
⁶College of Computer and Information Science, Northeastern University, Boston, MA USA

*To whom correspondence should be addressed.

Running title: Nonlinear regression for MS protein assay characterization

Contact: Email: o.vitek@neu.edu. Phone: +1 (617) 373 21 94
Abbreviations

DIA   data-independent acquisition
LOB   limit of blank
LOD   limit of detection
LLOQ  lower limit of quantitation/quantification
PRM   parallel reaction monitoring
SRM   selected reaction monitoring
Summary

The need for assay characterization is ubiquitous in quantitative mass spectrometry-based proteomics. Among many assay characteristics, the limit of blank (LOB) and limit of detection (LOD) are two particularly useful figures of merit. LOB and LOD are determined by repeatedly quantifying the observed intensities of peptides in samples with known peptide concentrations, and deriving an intensity versus concentration response curve. Most commonly, a weighted linear or logistic curve is fit to the intensity-concentration response, and LOB and LOD are estimated from the fit. Here we argue that these methods inaccurately characterize assays where observed intensities level off at low concentrations, which is a common situation in multiplexed systems. This manuscript illustrates the deficiencies of these methods, and proposes an alternative approach based on non-linear regression that overcomes these inaccuracies. We evaluated the performance of the proposed method using computer simulations, and using eleven experimental datasets acquired in Data-Independent Acquisition (DIA), Parallel Reaction Monitoring (PRM), and Selected Reaction Monitoring (SRM) mode. When the intensity levels off at low concentrations, the non-linear model changes the estimates of LOB/LOD upwards, in some datasets by 20-40%. In absence of a low concentration intensity leveling off, the estimates of LOB/LOD obtained with non-linear statistical modeling were identical to those of weighted linear regression. We implemented the non-linear regression approach in the open-source R-based software MSstats, and advocate its general use for characterization of mass spectrometry-based assays.
1 Introduction

Assay characterization is important in quantitative bottom-up mass spectrometry-based proteomics. By relating the observed intensities to the concentration of the protein or peptide, it enables quantification of peptides or proteins in clinical research and diagnostics [1, 2], or in systems biology research [3]. An important aspect of assay characterization is the estimation of its limit of blank (LOB), limit of detection (LOD) and lower limit of quantification (LLOQ), which quantify the sensitivity and the quantitative reproducibility of the assays. The estimation of these figures of merit in mass spectrometry has been necessary, e.g. for drug of abuse testing [4] or for clinical diagnostics, e.g. with PSA [5]. They are particularly important when the detection limits are near the clinical decision limit, as in the case of troponin [6]. They are also important for evaluating various mass spectrometric workflows [7]. Much technological research aims at estimating and minimizing these values for various experiment types. As of early 2017, over 700 publications indexed on PubMed reported LOB and LOD for mass-spectrometry based experiments. However, there is currently no commonly accepted experimental design or statistical guidelines for deriving these figures of merit, in particular in multiplexed mass spectrometric investigations.

Assay characterization involves spiking peptides or proteins in known concentrations in a simple or complex background, or titrating a peptide or protein mixture of known composition. After digestion, separation and spectral acquisition, the observed peak intensities of a peptide are summarized, and related to the true concentrations using statistical models. The statistical models used for this task is the main focus of this manuscript.

Most statistical models rely explicitly or implicitly on the canonical calibration curve [8], which contains three distinct regimes (Figure 1(a)). The noise regime, i.e. the concentrations where peptide intensities are indistinguishable from the chemical and electronic noise. These intensities are referred to as background noise. The linear regime, i.e. the concentrations where the canonical calibration curve is an increasing straight line (referred to as response curve). The slope of the line is the efficiency of the assay. Finally, the saturation regime, i.e. the concentrations where an increase in concentration does not translate into an increase in the intensity of the analyte. The three regimes are defined by the parameters Change and \( \text{Change}_{\text{Sat}} \) in Figure 1(a). Since most experiments do not include concentrations in the saturation regime, this manuscript focuses on the noise and on the linear regime. However, the approach proposed in the manuscript can be easily extended to the full canonical calibration curve.

The concepts of LOB and LOD of a peptide are best defined with respect to the canonical calibration curve, as illustrated in Figure 1(b) [9]. Assuming that the peptide intensities fall in the linear regime, the LOB is defined as the concentration at which the upper bound of the prediction interval of the background noise intersects the mean intensity of samples with the peptide. The LOD is defined as the concentration at which the upper bound of the prediction interval of the background noise intersects the lower bound of the prediction interval of samples with the peptide. The prediction interval, i.e. the confidence interval of a new intensity (as opposed to the confidence interval of the mean response curve) must be used to accurately reflect the performance of a future assay. The lower limit of quantification or quantitation (LLOQ) is loosely defined as the lowest concentration at which the measurements are acceptable for the intended use [10]. Since multiple definitions of LLOQ exist [11], this manuscript only presents results for the LOB and LOD. However, the discussion can readily be extended to LLOQ.

When mass spectrometric experiments characterize a limited number of peptides, the concentrations are fine-tuned, so that the intensities fall well inside the linear regime, and the canonical
calibration curve is observed. Following the canonical calibration curve, statistical analyses fit a linear regression that relates the concentrations and the observed intensities on the original (i.e., not log-transformed) scale. Many methods for fitting linear regression and estimating the LOB, LOD have been proposed [12, 13, 14, 15], compared [9, 16, 17, 18], and implemented in open-source software such as Quasar [19] and Skyline [20]. Most methods acknowledge that the variance of the observed intensities increases with the strength of the intensity, and specify regression with unequal variance, or use a robust estimation procedure.

Increasingly however, mass spectrometric assay characterization is multiplexed, and simultaneously quantifies hundreds or even thousands of peptides using, e.g. Selected Reaction Monitoring (SRM), or more recently Parallel Reaction Monitoring (PRM) or Data-Independent Acquisition (DIA). In these experiments it is often impossible to fine-tune the concentrations of each peptide, and some may fall outside of the linear regime. When the number of concentrations or replicates is relatively small, the intensities from the lower portion of the linear regime are similar to the intensities of the background noise. This creates uncertainty in the shape of the response, and the intensities of the peptide flatten out and produce the apparent calibration curve. The apparent calibration curve is illustrated with the dashed line in Figure 1 (a) and using an experimental dataset in Figure 2. Statistical analysis of multiplexed experiments must appropriately deal with these situations, and a linear model is not appropriate.

To account for the non-linearity of the apparent calibration curve, alternative approaches fitting a logistic response curve [21, 22, 23] have been proposed. However, as illustrated in Figure 2 they do not fully capture the intensities at low concentrations. Moreover, their estimation fails for peptides that are in linear regime, and for which the intensities do not contain both change points. As a result, the general strategy for the statistical analysis of the entire multiplexed assay characterization experiment is often unclear.

This manuscript proposes to remedy these challenges in highly multiplexed assay characterization experiments. It generalizes the approaches based on linear and logistic curve fits, and enables a more general but intuitive assay calibration. When the concentrations fall entirely in the linear portion of the dynamic range (i.e., in range of concentrations where the canonical and the apparent calibration curves agree), the output of the proposed approach is identical to that of weighted linear regression. When the concentrations deviate from this range and the observed intensities level off, the proposed approach appropriately represents the canonical calibration curve, while accounting for the uncertainty in the estimation. We evaluated the proposed approach using computer simulations, and compared the resulting figures of merit on 11 experimental datasets acquired in SRM, PRM and DIA mode. The evaluations demonstrated that the proposed approach results in a more accurate, and often more conservative, estimation of the figures of merit.

2 Experimental procedures

We evaluated the performance of the proposed approach using eleven experimental datasets, acquired in Selected Reaction Monitoring (SRM), Parallel Reaction Monitoring (PRM), or Data-Independent Acquisition (DIA), as well as using computer simulation. The datasets are summarized in Table 1. All the experimental datasets were processed with Skyline [20].

**Assay development and evaluation experiments** Experiments SRM1–SRM8 consisted of heavy isotope-labelled peptides spiked into a complex background in known and varying concentrations, and systematically quantified together with the endogenous peptides. Experiment SRM1 was a
multi-laboratory study of SRM performance for human proteins (CPTAC study 7) [24]. Targeted SRM experiments SRM2–SRM7 were all the assays available on the CPTAC assay portal [25] at the time of writing. The DIA experiment is described in detail in Supplementary Section 1.2.

In each dataset, areas under the chromatographic curves corresponding to fragments or transitions of a peptide in a run were summed, and subjected to natural logarithm transformation. In the case of the DIA dataset, only the most intense fragment ions are considered, as described in Supplementary Section 1.2. The summed and log-transformed areas were then normalized across runs, by first equalizing the medians of these values among all the endogenous peptides, and then applying the same shifts to the values of all the heavy peptides. Finally, the normalized values were converted back to the original scale. In the following, we refer to these summed and normalized peak areas as the intensity of the peptide in the run.

Assay used in large clinical studies Experiment CL [2] was a mass spectrometric assay of vitamin D binding globulin (VDBG) in human serum. A human serum sample containing VDBG was diluted into a background of chicken serum lacking VDBG. Heavy-labeled standard peptides were spiked in at a fixed concentration in each run. The assay quantified two peptides of the protein, VLEPTLK and ELPEHTVK. The peak area responses and concentration of these two peptides was calculated using a five-point calibration curve. The concentration values of the peptides were averaged to determine the concentration of VDBG. The intensities were normalized to the heavy-labeled standard according to the same procedure as above.

Simulated datasets In the experimental datasets, the true LOB/LOD are unknown, and therefore we can only evaluate the impact of different statistical approaches on their estimation. In contrast, simulated datasets can help us evaluate the accuracy of the results. We simulated two datasets, SIM1 and SIM2 with known figures of merit, as described in Supplementary Section 1.4.1.

3 Background

To place the proposed approach in the context of existing work, in this section we describe the definitions, notation, and current approaches for assay characterization based on linear [9] and logistic [21] curve fit with unequal variance. We assume that the characterization is performed at the peptide level (i.e., after summarizing all the transitions or fragments of the peptide), and describe the model for a single representative peptide.

3.1 Input data and notation

In mass spectrometry-based assay characterization experiments, each peptide is spiked or titrated in \(L\) samples in known and distinct concentrations \(C_1 < \ldots < C_L\). The experiments also typically include blank samples used to quantify the background noise. We denote the blank samples by the true concentration \(C_0 = 0\). The \(L + 1\) samples are analyzed with a mass spectrometer in technical replicates (i.e., runs). The spectral peaks corresponding to the peptide in a run are detected and quantified (e.g., by integrating the area under their chromatographic curve, or by any other method of choice) using a data processing tool such as Skyline [20]. The peak intensities are then summarized (e.g., with the sum of the chromatographic peak areas of all the peptide transitions or fragments) and normalized, to produce a single intensity value per replicate run.

The output of data processing and summarization is the input to peptide assay characterization. It consists of \((C_i, Y_{ij})\) pairs of true peptide concentrations \(C_i\), \(i = 0, \ldots, L\) and their summary
intensities $Y_{ij}$ in the replicate run $j$, $j = 1, \ldots, N_i$ on the original (i.e., not logarithm transformed) scale. The total number of intensities for a peptide is $N = \sum_{i=0}^{L} N_i$.

3.2 Existing approaches for multiplexed assays

3.2.1 Linear curve fit

When the concentrations are in the linear portion of the dynamic range, many assay characterizations rely on a linear curve fit on the original (i.e., not log-transformed) scale. Most models explicitly or implicitly specify a two-part statistical model, where one part is for the blank samples, and the other for the samples containing the peptide. For the blank samples, the approaches assume that the background noise randomly fluctuates around a constant. For the samples with the peptide, the approaches assume a linear response curve, that relates the observed intensities to the spiked or titrated concentrations. The fluctuations around this expected relationship are assumed to be Normally distributed, with variance depending on the concentration. In the notation of this manuscript, a typical model is written as:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2), \quad i = 0, \ldots, L, \quad j = 1, \ldots, N_i, \quad \text{and}$$

$$\mu_i = \text{Intercept} + \text{Slope} \times C_i$$

The existing regression-based approaches differ in their assumptions regarding the variance. Some specify a distinct variance parameter $\sigma_i^2$ for each concentration, as in Equation (1) [9]. A common alternative [12, 26] assumes that the variance is proportional to $C_i^2$, i.e.

$$\sigma_i^2 = \sigma_0^2 + \sigma_0^2 \times C_i^2.$$  \hfill (2)

The first term of the sum models the variance of the chemical and electronic noise at low concentrations, while the second models its increase with concentration. The $C_i^2$ scaling implies the constant signal to noise ratio in the linear regime. Since the pattern in Equation (2) was not observed for the datasets in this manuscript, the results below follow the approach in Equation (1) and in [9].

3.2.2 Logistic curve fit

Approaches based on logistic curve fit are similar in spirit to the approaches based on a linear curve fit above. They model the mean observed intensities $\mu_i$ in Equation (1) with a logistic curve. In our notation, a five-parameter logistic curve [27, 28] is expressed as:

$$\mu_i = a - b + \frac{b}{1 + \left(\frac{C_i}{c}\right)^d}$$

It contains five parameters [29]: $a$ is the intensity value at zero concentration and in the noise regime, $b$ is the intensity value of the saturation regime, and $c$ is the midrange concentration between $a$ and $b$. Coefficients $d$ and $g$ control the extent to which the response curve moves from the noise regime to the saturation regime. A four-parameter logistic curve sets $g = 1$ in Equation (3), and restricts the curve to be symmetric around the midpoint $\left(C = c, \mu_i = a + \frac{c}{2}\right)$. In the following, we will compare the proposed approach to the approach based on the five-parameter logistic curve, as it is most
widely used and often generates better quality fit.

### 3.2.3 Figures of merit

The *limit of blank (LOB)* of a peptide is defined as the highest concentration, at which the *expected* mass spectrometric intensity in presence of the peptide is indistinguishable from the *possible* background noise \([10, 12]\). The possible background noise is defined as the \(100 \times (1 - \alpha)\) % one-sided *prediction interval* of the intensity from the blank sample, which we denote as \(\text{Noise}_\alpha\). Therefore, in the notation of Equation (1), the LOB is the concentration \(C\) at which \(\text{Noise}_\alpha\) intersects the response curve \(\text{Intercept} + \text{Slope} \times C\). This definition formalizes the fact that at the LOB, the probability of false positive peptide quantification is \(\alpha\), and the probability of false negative peptide quantification is 0.5. The LOB is illustrated with a dot in Figure 1(b).

The *limit of detection (LOD)* of a peptide is defined as the highest concentration, at which the *possible* mass spectrometric intensity in presence of the peptide is indistinguishable from the *possible* background noise \([10, 12]\). The possible mass spectrometric intensity in presence of the peptide is defined as the lower bound of the \(100 \times (1 - \beta)\) % one-sided prediction interval at concentration \(C_i\). At the LOD, the lower bound intersects the possible background noise \(\text{Noise}_\alpha\). This definition formalizes the fact that intensities from any true concentration above the LOD must not be mistaken for background noise. At the LOD, the probability of false positive peptide quantification is \(\alpha\). The probability of false negative peptide quantification is \(\beta\). The LOD is illustrated with a triangle in Figure 1(b).

The definitions are easily extended to a logistic curve. In this case, the LOB uses the logistic response curve fit, and the LOD uses the lower bound of the prediction interval of the logistic response curve. In all the cases, the parameters \(\alpha\) and \(\beta\) are set in advance. Smaller values of \(\alpha\) and \(\beta\) lead to wider prediction intervals \([30]\), and to more conservative figures of merit. In the following we systematically set \(\alpha = \beta = 0.1\) as is commonly done \([10]\).

An alternative definition of the LOD \([13, 31]\) is the concentration \(C\) at which the response curve has intensity \(\text{Noise}_\alpha + \sigma_1 \frac{1}{1 - \frac{1}{\Phi(z_{1 - \beta})}}\), where \(z_{1 - \beta}\) is the \((1 - \beta)\times100\) quantile of the standard Normal distribution, \(\sigma_1\) is the standard deviation of the sample with the peptide at the lowest concentration. Since this definition does not have a straightforward expression of prediction intervals, we do not use it in what follows.

### 3.2.4 Parameter estimation and prediction intervals

The figures of merit rely on unknown quantities that must be estimated from the data. For the blank samples, an estimate of the expected background noise \(\mu_0\) is obtained as the average intensity \(\bar{Y}_0 = \frac{1}{N_0} \sum_{j=1}^{N_0} Y_{0j}\). The upper bound of the \((1 - \alpha)\times100\) prediction interval of the background noise is then determined as \([32]\):

\[
\hat{\text{Noise}}_\alpha = \bar{Y}_0 + \hat{\sigma}_0 t_{1 - \alpha, N_0 - 1} \sqrt{1 + \frac{1}{N_0}},
\]

where \(t_{1 - \alpha, N_0 - 1}\) is the \((1 - \alpha)\times100\) % percentile of the Student’s t distribution with \(N_0 - 1\) degrees of freedom.

For the samples with the peptide, parameters Slope and Intercept in Equation (1) are estimated via weighted least squares \([30]\), as follows. First, the variance parameters \(\sigma_i^2\) are estimated as the
sample variance at each spiked or titrated concentration \[9\]. To increase the resolution beyond the spiked/titrated concentrations, the concentrations are discretized to a finer grid \(C_1 < C_i < C_{L+L'}\). This finer grid contains the original concentrations, augmented with \(L' = 100\) concentrations equally distributed in log space between the minimum and the maximum of the spiked/titrated concentration values.

The gridded variance parameters \(\sigma_i^2\) are estimated by linear interpolation between estimates from the closest observed concentrations. Next, the concentration-specific weights are defined as \(w_i = \frac{1}{\hat{\sigma}^2_i}\). Finally, the parameters in Equation (1) are estimated by plugging in the weights. The lower bound of the one-sided \((1 - \beta) \times 100\%\) prediction interval is obtained with the closed-form expression \[13, 9\]:

\[
\hat{\Pi}_{\beta, i'} = \hat{\mu}_{i'} - t_{1-\beta, N-2} \sqrt{\frac{1}{w_{i'}} + \frac{1}{\sum_{i=1}^{L'} \sum_{i=1}^{N_i} w_i} + \frac{(C_{i'} - \bar{C})}{SCC_w}}
\]

where

\[
C_w = \frac{\sum_{i=1}^{L'} C_i N_i w_i}{\sum_{i=1}^{L'} N_i w_i}
\]

and

\[
SCC_w = \frac{\sum_{i=1}^{L'} N_i w_i (C_i - \bar{C})}{\sum_{i=1}^{L'} N_i w_i}.
\]

Here \(t_{1-\beta, N-2}\) is the \((1 - \beta)\) quantile of the Student distribution with \(N - 2\) degrees of freedom.

Unfortunately, no closed-form analogy to Equation (5) exists for the logistic curve fit. Therefore, the lower bound of the prediction interval \(\hat{\Pi}_{\beta, i'}\) is obtained via bootstrap \[33\]. The bootstrap is an iterative procedure with a large number \(B\) of iterations, say \(B = 2000\). At each iteration, the pairs \((C_{i}^*, Y_{ij}^*)\) are obtained by sampling with replacement from the observed pairs \((C_i, Y_{ij})\).

The estimates \(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}^*\) and \(\hat{\mu}_{i'}\) in Equation (3) are obtained via weighted least squares, while plugging in the full-sample weights \(w_i = \frac{1}{\hat{\sigma}^2_i}\).

With the logistic curve fit, the lower bound of the one-sided \((1 - \beta) \times 100\%\) prediction interval for a concentration \(C_{i'}\) is defined as the \(\beta \times 100\%\) quantile of the mixture of \(B\) Normal distributions \(N(\hat{\mu}_{i'}, \hat{\sigma}^2_{i'})\) with equal weights. The quantile can be approximated by simulation \[34\], or evaluated analytically on a grid of augmented concentrations above. Here we chose the grid-based evaluation as it is more computationally efficient, as detailed in Algorithm 1 in Supplementary Section 2.1.

Finally, the estimate of LOB is the concentration \(C_{i'}\) at which the estimated upper bound of the background noise \(\hat{\text{Noise}}_{\alpha}\) intersects the fitted curve \(\hat{\mu}_{i'}\) (obtained with either linear or logistic curve fit). The estimate of LOD is the concentration \(C_{i'}\) at which the estimated upper bound of the background noise \(\hat{\text{Noise}}_{\alpha}\) intersects the estimated lower bound of the prediction interval \(\hat{\Pi}_{\beta, i'}\), as shown in Fig 1(b).

4 Proposed statistical approach

4.1 Modeling the canonical calibration curve

The proposed approach assumes that the systematic relationship between the true concentrations and the observed intensities follows the canonical calibration curve in the noise and linear regimes, as shown in Fig 1(a). In other words, it generalizes the linear response \(\mu_i\) in Equation (1) by introducing the parameter \(\text{Change} \geq 0\). The intensities from all the concentrations below \(\text{Change}\)
(including the blank samples) are viewed as the background noise, and the intensities of peptides
with concentrations above Change fluctuate around the linear response curve. This bilinear model
has a "hockey stick" appearance, and is formally written as:

\[ Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2), \quad i = 0, \ldots, L, \quad j = 1, \ldots, N_i, \quad \text{and} \]

\[ \mu_i = \begin{cases} 
\mu_0, & \text{when } 0 \leq C_i < \text{Change} \\
\mu_0 + \text{Slope} \times (C_i - \text{Change}), & \text{when } \text{Change} \leq C_i 
\end{cases} \]

The parameter Change defines the functional form of the response curve. When Change \( \leq C_1 \), all
the spiked or titrated concentrations of a peptide are within the linear regime, and Equation (6)
is simplified to the linear model in Equation (1). When Change > C_1 some of the concentrations
are below the linear regime, the observed intensities from these concentrations are indistinguishable
from the background noise, and the three-parameter "hockey stick" pattern is appropriate for the
peptide. Therefore, the model can accurately represent all the peptides in a multiplexed experiment.

The model in Equation (6) is easily extended to situations where the spiked or titrated con-
centrations span noise, linear and saturation regimes in Figure 1(a). Such extension is detailed in
Supplementary Section 2.2 and results are shown in Supplementary Figure 4.

4.2 Parameter estimation and model selection

The parameters \( \mu_0 \), Slope and Change in Equation (6) are unknown, and must be estimated from the
data. The model is first fit to all the intensities of the peptide using the same weighted least squares
estimation procedure as in Section 3.2.4. When the linear model is appropriate, the non-linear
expression in Equation (6) is too complex (i.e., is over-parametrized), and results in a substantial
uncertainty in the estimate of Change. We quantify this uncertainty with the \( (1 - \gamma) \times 100\% \)
confidence interval. We set \( \gamma = 0.2 \) in what follows.

The confidence interval can be derived approximately using large-sample approximation [35], or
by a bootstrap procedure. Since the number of replicate runs in assay characterization experiments
is relatively small, here we use bootstrap as described in Supplementary Section 2.1. If the confidence
interval is located above \( C_1 \), we conclude that the change point is present. Otherwise, we conclude
that the observed intensities are consistent with the linear response, and that the simpler linear
model in Equation (1) is appropriate. Finally, the selected (linear or "hockey stick") model is refit
to the data using weighted least squares.

4.3 Prediction intervals and figures of merit

The prediction intervals from this model must reflect the uncertainty in (1) selecting the appropriate
model, (2) estimating the unknown parameters of the selected response curve, and (3) random
fluctuations of the individual intensities around the response. The combined effect of these three
sources cannot be expressed analytically. However, the intervals can be obtained using the bootstrap
procedure, similar to that in Section 3.2.4, but with modifications.

At each bootstrap iteration, the pairs \( (C_i^*, Y_{ij}^*) \) are obtained by sampling with replacement from
the observed pairs \( (C_i, Y_{ij}) \). Parameters in Equation (6) are estimated via weighted least squares,
while plugging in the full-sample weights \( w_i = \frac{1}{\hat{\sigma}_i^2} \). To reflect the uncertainty due to model selection,
the model selection is performed as described in Section 4.2 and either a linear or a “hockey stick" model is fit to the resampled dataset.

Figure 3 shows that the bootstrap iterations can vary substantially in both the selected model and its parameter estimates. Therefore, the response curve is obtained by averaging the estimated response curves across the bootstrap iterations. As can be seen from the figure, the resulting response curve may be highly non-linear, and may lack the characteristic “hockey stick" pattern. This non-linear fit reflects the uncertainty in the selection and in the estimation of the parameters of the canonical calibration curve.

Finally, the bounds of the prediction interval are obtained as the quantiles of the mixture of the predictive distributions across the iteration, where the numeric values for each concentration are evaluated on a grid as in Section 3.2.4. The figures of merit are then obtained as in Section 3.2.4. The estimate of LOB is the concentration $C_i'$ at which the estimated upper bound of the background noise $\hat{\text{Noise}}_{\alpha}$ intersects the fitted curve $\hat{\mu}_i$. The estimate of LOD is the concentration $C_i'$ at which the estimated upper bound of the background noise $\hat{\text{Noise}}_{\alpha}$ (defined in Equation (4)) intersects the estimated lower bound of the prediction interval.

4.4 Implementation

The proposed approach is implemented as part of the free and open-source R-based package MSstats v3.6 and subsequent versions (http://msstats.org/assay-characterization) [36]. The pseudocode for the algorithm is available in Supplementary Section 2.1. The computational cost of the estimation procedure is fairly limited, with the most expensive step being the calculation of bootstrap samples. For instance, Figure 5(b) was obtained on a standard workstation with 3.5 GHz CPU in about one minute. The largest dataset in this manuscript, SRM2, was analyzed in approximately 6 hours. The algorithm can be easily parallelized, and the runtime can be shortened by deploying it on a multi-core computer or on a cluster. The runtime can also be shortened by replacing the bootstrap-based confidence interval of the parameter Change in Equation (6) with an asymptotic confidence interval, but at the expense of an accuracy loss.

5 Evaluation

5.1 The proposed approach accurately fit the data

As illustrated in Figure 2 both linear and logistic curve fits can fail to simultaneously capture the flattening of peptide intensities at low concentrations, and the linearity of the response curve. In contrast, the proposed approach relies on the canonical calibration curve, and on bootstrap-based estimation, to capture both aspects. Figure 3 shows the proposed curve fit for the same peptide as in Figure 2. The curvature of the fit explicitly reflects the uncertainty in the concentration Change, at which the noise regime transitions into the linear response.

Additional examples of a better fit and better interpretation of the proposed approach are shown in Figure 4 for the SRM2 experiment, and in Figure 5 for the DIA experiment. As can be seen, the proposed approach led to higher figures of merit as compared to linear or logistic curve fit. This reduced optimism in the performance of the assays may have important implications for their downstream use.

The proposed approach did not always alter the figures of merit. Supplementary Figure 1 illustrates the fact that, when the concentrations fall entirely within the linear portion of the dynamic
range, the proposed approach was identical to that of weighted linear regression. The figure also indicates that the proposed approach was relatively robust to deviations from the linear regime at larger concentrations. The larger concentrations were associated with higher variance, and were therefore given a smaller weight during the estimation procedure. Supplementary Figure 2 presents a case where the proposed approach lowered the estimated figures of merit as compared to an existing approach.

5.2 The proposed approach improved the accuracy of LOB/LOD estimates

As we have seen in Figure 4 and in Figure 5, the differences in the model fit between the existing and the proposed approaches impacted the estimates of the figures of merit. Since the true figures of merit in these experiments are unknown, we evaluated the accuracy of the estimates using the simulated dataset SIM1.

Figure 6(a) illustrates the accuracy of fit in one simulation instance. It shows that deviations from linear patterns increased the intercept and decreased the slope of the linear fit, as compared to the exact response curve. Although the logistic fit captured the non-linearity of the intensities at low concentrations, it deviated from the linear pattern in the linear regime. These situations led to smaller (i.e., less accurate, too optimistic) figures of merit.

The accuracy of the proposed approach over all the simulation instances is reported in Table 2 and illustrated in Figure 6(b). Both linear and logistic curves dramatically underestimated the true figures of merit. On average, for the simulated dataset, the linear curve underestimated the LOB by 63% and the LOD by 25%. The 5-parameter logistic curve fit underestimated the LOB by 53% and the LOD by 6%. In contrast, the proposed approach was slightly conservative, on average it overestimated the LOB by 1% and the LOD by 4%. However, it substantially reduced both the magnitude and the variability of the estimation error. This indicates that the estimates by the proposed approach for the experimental datasets are likely closer to the true values than those by linear and logistic fits.

Supplementary Figure 3 shows that when the number of grid points used to discretize the spiked or titrated concentrations was sufficiently large, it had a relatively small impact on the accuracy of the results. The insufficient number of discretized concentrations can be diagnosed using plots such as Supplementary Figure 3(a), which display overly jagged curves. As shown in Supplementary Figures 3(b) and (c), this issue can be corrected by using a finer grid.

Finally, Supplementary Figure 4 uses simulated datasets SIM2 and SIM3 to illustrate that the proposed approach accurately estimates the figures of merit for assays with various combinations of noise, linear and saturation regimes.

5.3 The proposed approach produced more conservative figures of merit in the assay development and evaluation experiments

Figure 7 compares the figures of merit obtained with linear, logistic and proposed approaches for all the assay development and evaluation experiments. Since the true values for these datasets are unknown, the figure plots ratios of the estimates by the logistic and the proposed approaches to the estimates of the linear fit.

As compared to the linear fit, the proposed approach substantially reduced the optimism regarding the figures of merit for all the datasets. As compared to the logistic curve fit, it reduced the optimism for many datasets. The extent of the changes depended on the number of peptides
for which intensities deviate from the linear pattern, and on the type of these deviations. A larger increase in the LOB and LOD was associated with a larger proportion of peptides with non-linear patterns. For example, in the PRM dataset 84% of peptides had non-linear patterns (i.e., more than 50% of the bootstrap iterations selected the "hockey stick" model with the proposed approach). The proposed approach increased the LOB and LOD by on average 40% as compared to the linear fit. In contrast, dataset SRM1 had only 14% of peptides with non-linear patterns, and the proposed approach did not produce substantial changes in the figures of merit.

At the same time, the proposed approach did not always produce more conservative figures of merit. Supplementary Figure 2 illustrates one such a case for a peptide from dataset SRM7. The proposed approach produced smaller (i.e., more optimistic) figures of merit in this case. This behavior may be associated with intensities that deviate substantially from the canonical calibration curve, e.g., in presence of interferences or other data quality issues. In the assay development and evaluation experiments, such situation was observed on average in 10% of the peptides over all the datasets.

Supplementary Figure 5 illustrates these patterns for the DIA experiment. It shows that the non-linearity of a response curve of a peptide is strongly associated with the difference between the proposed and linear model-based estimates of LOB and LOD.

5.4 The proposed approach produced conservative figures of merit in the clinical research assay

We applied the proposed approach to the clinical research assay experiment. This dataset is representative of clinical research assays in that it is not multiplexed, contains many different concentrations, and a large number of replicates as compared to the assay development and evaluation experiments described previously in this manuscript.

The linear fit for this dataset is illustrated in Figure 8(a). The linear curve appears to fit well overall. However, a zoom into the lowest concentrations shows that it failed to accurately represent this particular region, which is of our main interest. The logistic curve fit and the fit by the proposed approach are illustrated in Figure 8(b). Due to the lack of intensities in the saturation range, the logistic curve fit was unstable. It failed to adequately capture the leveling off at low concentrations, and resulted in an overly wide prediction interval. In contrast, the proposed approach accurately represented the data, facilitated the interpretation in the context of the canonical calibration curve, and provided reasonable estimates of the figures of merit.

Other quantities, in addition to LOB and LOD, are often of interest in clinical research assays. These include lower limit of quantification (LLOQ). In this manuscript we used a definition of LLOQ, which is based on a linear curve fit, and is limited to the concentrations in the experiment. The definition and the details of the calculation for this dataset are described in Supplementary Section 3.2. The procedure resulted in LLOQ=1 μg/ mL. Although this quantity is not necessarily related to the LOD, we note that it was below the value obtained by the proposed approach (LOD=1.03), and may therefore be potentially be too optimistic.

6 Discussion

We introduced a novel statistical approach, which is specifically designed for characterization of multiplexed mass spectrometry-based assays. It expands upon the linear curve fit between the
observed peptide intensities and known concentrations by directly incorporating the concept of the canonical calibration curve.

When peptide concentrations fall entirely within the linear portion of the dynamic range, the fit by the proposed approach is identical to that of weighted linear regression. However, when peptide concentrations are small and their intensities flatten out, the proposed approach results in a more accurate and a more interpretable fit. This is particularly important in multiplexed experiments, where it is impossible to ensure that the concentrations of all the assayed peptides are in the linear portion of the dynamic range.

The proposed approach improves the accuracy of the figures of merit as compared to the linear and to the logistic curve fits. When peptide concentrations are below the linear portion of the dynamic range, the proposed approach typically results in more conservative figures of merit. This may have important implications regarding the usability of the assay. At the same time, the concentrations above the linear portion of the dynamic range do not have a strong impact on the quality of the fit. This is due to the fact that intensities from high concentrations have higher variance, and have a smaller weight during the estimation. If needed, extensions of the proposed approach to tri-linear fits can be used to capture concentrations exceeding the linear regime.

The proposed approach is compatible with any strategy for quantifying a peptide in a run. In this manuscript, we chose to sum the intensities of all the transitions or fragments of a peptide that are reported by data processing, because this approach is simple, frequently used, and does not require additional software or manual intervention. However, other strategies are also acceptable. For example, quantification can be performed using top transitions or fragments. The transitions or fragments can be manually or automatically curated to remove interferences. If a curation procedure is used, it must be incorporated into every subsequent use of the assay to avoid overly optimistic characterization of its performance.

The proposed approach can be used for absolute quantification. A byproduct of the estimation procedure is a confidence interval for the parameter Change, which determines the lower bound of the linear portion of the dynamic range. The fit obtained by the proposed approach for concentrations above the confidence interval, or a separate linear fit limited to the concentrations above the confidence interval, can be used to relate the observed intensities of the peptide to the true concentrations.

Since the proposed approach discretizes the true concentrations, its resolution is not limited to the concentrations used in the experimental design. Therefore, it can help compare the performance of alternative assays or workflows for a same peptide. Moreover, discrepancies between the figures of merit by the proposed approach and by the linear fit, in particular in cases where the proposed approach produces lower estimated values, can be used to automatically screen the assays for interferences, or for other data quality issues, which can be subsequently manually inspected.

Overall, we believe that the proposed approach offers multiple benefits, and advocate its general use for characterization of multiplexed mass spectrometry-based assays.

7 Acknowledgments

The work was supported in part by the US National Science Foundation CAREER award DBI-1054826 to O.V. We thank Ching-Yun Chang for valuable discussions and feedback and Meena Choi for help with the implementation.
8 References

[1] J. Pannee, E. Portelius, M. Oppermann, A. Atkins, M. Hornshaw, I. Zegers, P. Hojrup, L. Minthon, O. Hansson, H. Zetterberg, K. Blennow, and J. Gobom. A selected reaction monitoring (SRM)-based method for absolute quantification of $A\beta_{38}$, $A\beta_{40}$, and $A\beta_{42}$ in cerebrospinal fluid of Alzheimer’s disease patients and healthy controls. *J Alzheimers Dis*, 33(4):1021–1032, 2013.

[2] C.M. Henderson, Lutsey P.L., Misialek J.R., Laha T.J., and J.H. Selvin E, Eckfeldt. Measurement by a novel LC-MS/MS methodology reveals similar serum concentrations of vitamin d–binding protein in blacks and whites. *Clin. Chem.*, 62(1):179–187, 2016.

[3] C. Lawless, S.W. Holman, P. Brownridge, K. Lanthaler, V.M. Harman, R. Watkins, D.E. Hammond, R.L. Miller, P. Sims, C.M. Grant, C Eyers, R.J. Beynon, and S.J. Hubbard. Direct and absolute quantification of over 1800 yeast proteins via selected reaction monitoring. *Mol. Cell Proteomics*, 15(4):1309–1322, 2016.

[4] D.A. Armbruster, M.D. Tillman, and L.M. Hubbs. Limit of detection (LOD)/limit of quantitation (LOQ): comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs. *Clin. Chem.*, 40(7):1233–1238, 1994.

[5] J. E. McDermed, R. Sanders, S. Fait, R.E. Klem, M.J. Sarno, T.H. Adams, and E.P. Diamandis. Nucleic acid detection immunoassay for prostate-specific antigen based on immuno-PCR methodology. *Clin. Chem.*, 58(4):732–740, 2012.

[6] M. Moretti, B. Pieretti, D. Sisti, M. Rocchi, and S. Gasperoni. Analytical performance and clinical decision limit of a new release for cardiac troponin I assay. *Ann Clin Biochem*, 52(1):169–172, 2014.

[7] L. C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, and R. Aebersold. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell Proteomics*, 11(6):O111.016717, 2012.

[8] H. Hubschmann. *Handbook of GC/MS*. Wiley, 2008.

[9] I. Lavagnini and F. Magno. A statistical overview on univariate calibration, inverse regression, and detection limits: Application to gas chromatography/mass spectrometry technique. *Mass Spectrom. Rev.*, 26(1):1–18, 2006.

[10] V.J. Barwick and E. Prichard, editors. *Eurachem guide: Terminology in analytical measurement introduction to VIM 3*. Wiley, 2011. Available from www.eurachem.org.

[11] J. Carlson, A. Wysoczanski, and E. Voigtman. Limits of quantitation yet another suggestion. *Spectrochim. Acta B*, 96:69–73, 2014.

[12] L.A. Currie. Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal Chem*, 40(3):586–593, 1968.
[13] M.E. Zorn, R.D. Gibbons, and W.C. Sonzogni. Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. *Anal Chem*, 69(15):3069–3075, 1997.

[14] K. Linnet. Partly nonparametric approach for determining the limit of detection. *Clin. Chem.*, 50(4):732–740, 2004.

[15] E. Desimoni and B. Brunetti. About estimating the limit of detection of heteroscedastic analytical systems. *Anal. Chim. Acta*, 655(1-2):30–37, 2009.

[16] J. Vial and A. Jardy. Experimental comparison of the different approaches to estimate LOD and LOQ of an HPLC method. *Anal Chem*, 71(14):2672–2677, 1999.

[17] D.A. Armbruster and T. Pry. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev*, 29 Suppl 1:S49–52, 2008.

[18] D.R. Mani, S.E. Abbatiello, and S. Carr. Statistical characterization of multiple-reaction monitoring mass spectrometry (MRM-MS) assays for quantitative proteomics. *BMC Bioinformatics*, 13:S9, 2012.

[19] S. E. Abbatiello, B. Schilling, D. R. Mani, L. J. Zimmerman, S. C. Hall, B. MacLean, M. Albertolle, S. Allen, M. Burgess, M. P. Cusack, M. Gosh, V. Hedrick, J. M. Held, H. D. Inerowicz, A. Jackson, H. Keshishian, C. R. Kinsinger, J. Lyssand, L. Makowski, M. Mesri, H. Rodriguez, P. Rudnick, P. Sadowski, N. Sedransk, K. Shaddox, S. J. Skates, E. Kuhn, D. Smith, J. R. Whiteaker, C. Whitwell, S. Zhang, C. H. Borchers, S. J. Fisher, B. W. Gibson, D. C. Liebler, M. J. MacCoss, T. A. Neubert, A. G. Paulovich, F. E. Regnier, P. Tempst, and S. A. Carr. Large-scale interlaboratory study to develop and analytically validate and apply highly multiplexed and quantitative peptide assays to measure cancer-relevant proteins in plasma. *Mol. Cell Proteomics*, 14(9):2357–2374, 2015.

[20] B. MacLean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, and M. J. MacCoss. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26(7):966–968, 2010.

[21] C. Holstein, Griffin M., Hong J., and Sampson P.D. Statistical method for determining and comparing limits of detection of bioassays. *Anal Chem*, 87(19):9795–9801, 2015.

[22] K.R. Lee, B. Dipaolo, and X. Ji. Calibration and LOD/LOQ estimation of a chemiluminescent hybridization assay for residual DNA in recombinant protein drugs expressed in E. coli using a four-parameter logistic model. *Drug Dev Ind Pharm*, 26(6):661–669, 2000.

[23] FDA. *Guidance for Industry Bioanalytical Method Validation*, 2013. [https://www.fda.gov/downloads/Drugs/Guidances/ucm368107.pdf/](https://www.fda.gov/downloads/Drugs/Guidances/ucm368107.pdf/)

[24] CPTAC. *CPTAC data portal*, (accessed January, 2017). [https://cptac-data-portal.georgetown.edu](https://cptac-data-portal.georgetown.edu).

[25] CPTAC. *CPTAC assay portal*, (accessed January, 2017). [https://assays.cancer.gov](https://assays.cancer.gov)
[26] M.D. Wilson, D.M. Rocke, B. Durbin, and H.D. Kahn. Detection limits and goodness-of-fit measures for the two-component model of chemical analytical error. *Anal. Chim. Acta*, 509(2):197–208, 2004.

[27] M. Baud. Data analysis, mathematical modeling. *Methods of immunological analysis*, 1:656–671, 1993.

[28] F.J. Richards. A flexible growth function for empirical use. *J. Exp. Bot.*, 10(2):290–301, 1959.

[29] P. Gottschalk and J. Dunn. The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Anal Biochem*, 343(1):54–65, 2005.

[30] M.H. Kutner, J. Neter, C.J. Nachtsheim, and W. Li. *Applied linear statistical models*. Irwin, 1990.

[31] T. A. Addona, S. E. Abbatiello, B. Schilling, S. J. Skates, D. R. Mani, D. M Bunk, C. H. Spiegelman, L. J. Zimmerman, A. J. Ham, H Keshishian, S. C Hall, S. Allen, R. K. Blackman, C. H Borchers, C. Buck, H. L. Cardasis, M. P. Cusack, N. G. Dodder, B. W. Gibson, J. M. Held, T. Hieltke, A. Jackson, E. B. Johansen, C. R Kinsinger, J. Li, M. Mesri, T. A. Neubert, R. K. Niles, T. C. Pulsipher, D. Ranshoff, H. Rodriguez, P. A. Rudnick, D. Smith, D. L. Tabb, T. J Tegeler, A. M. Varyath, L. J. Vega-Montoto, A. Wahlander, S. Waldemarson, M. Wang, J. R. Whiteaker, L. Zhao, N. L. Anderson, S. J. Fisher, D. C. Liebler, A. G. Paulovich, F. E. Regnier, P. Tempst, and S. A. Carr. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring–based measurements of proteins in plasma. *Nature Biotechnol*, 27(7):633–641, 2009.

[32] R. Walpole, R. Myers, S. Myers, and K. Ye. *Probability and statistics for engineers and scientists*. Pearson, 9th edition, 2011.

[33] N. Meade and T. Islam. Prediction intervals for growth curve forecasts. *Int. J. Forecast.*, 14(5):413–430, 1995.

[34] M. Rahman, R. Rahman, and L. Pearson. Quantiles for finite mixtures of normal distributions. *Int J Math Educ Sci Technol*, 37(3):352–358, 2006.

[35] T.S. Ferguson. *A Course in large sample theory*. Chapman & Hall, 1996.

[36] M. Choi, C.Y. Chang, T. Clough, D. Broudy, T. Killeen, B. MacLean, and O. Vitek. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Biominformatics*, 30(17):2524–2526, 2014.

[37] J. Kennedy, P. Yan, L. Zhao, R.G. Ivey, U.J. Voytovich, H.D. Moore, C. Lin, E.L. Pogosova-Agadjanyan, D.L. Stirewalt, K.W. Reding, J.R. Whiteaker, and A.G Paulovich. Immobilized metal affinity chromatography coupled to multiple reaction monitoring enables reproducible quantification of phospho-signaling. *Mol. Cell Proteomics*, 15(2):726–739, 2016.

[38] J. J. Kennedy, S. E. Abbatiello, K. Kim, P. Yan, J. R. Whiteaker, C. Lin, J. S. Kim, Y. Zhang, X. Wang, R. G. Ivey, L. Zhao, H. Min, Y. Lee, M.-H. Yu, E. G. Yang, C. Lee, P. Wang, H. Rodriguez, Y. Kim, S. A. Carr, and A. G. Paulovich. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nature Methods*, 11(2):149–155, 2014.
[39] R. M. Schoenherr, R. G. Saul, J. R. Whiteaker, P. Yan, G. R. Whiteley, and A. G. Paulovich. Anti-peptide monoclonal antibodies generated for immuno-multiple reaction monitoring-mass spectrometry assays have a high probability of supporting Western blot and ELISA. *Mol. Cell Proteomics*, 14(2):382–398, 2015.
9 Figure Legends

Figure 1: Definitions. (a) Canonical (solid line) and apparent (dashed line) calibration curves. The changes between the noise, linear, and saturation regimes occur at concentrations $C_{\text{Change}}$ and $C_{\text{ChangeSat}}$. When the number of concentrations or replicates is small, the intensities from the lower or upper range of the linear regime are similar to the intensities from the noise or of the saturation regime. This creates uncertainty in the shape of the response, and produces the apparent calibration curve. The linear portion of the dynamic range is defined as the range of concentrations where the canonical and the apparent calibration curves agree. (b) Estimation of LOB and LOD based on a linear fit, illustrated for experiment SRM7, study IIIA, site 95, peptide ESDTSYVSLK. The LOB is defined as the concentration at the intersection between the upper bound of the 95% one-sided prediction interval of the background noise, and the linear fit. The LOD is defined as the concentration at the intersection between the upper bound of the 95% one-sided prediction interval of the background noise, and the lower bound of the 95% one-sided prediction interval of the linear fit (which also corresponds to the lower bound of the 90% two-sided prediction interval). We emphasize the importance of using prediction intervals (as opposed to confidence intervals for the mean fit, which would have led to more optimistic figures of merit). The quality of fit impacts the estimates of the figures of merit.

Figure 2: Restrictions imposed by linear or logistic curves may result in poor quality fit. DIA experiment, peptide SSAAPPPPR. The linear fit failed to capture the non-linearity of the intensities at low concentrations. The logistic curves ensured a better quality fit at low concentrations, but did not fully capture the patterns, and violated the expected linearity of the response at higher concentrations. (a) Zoom into low concentrations. (b) Overall view.

Figure 3: Proposed curve fit. DIA experiment, peptide SSAAPPPPR. For every bootstrap iteration, the approach performs model selection, and fits the selected linear or “hockey stick” curve (grey lines) to the resampled dataset. The fitted curves vary substantially between the bootstrap iterations, indicating the uncertainty in the model and in the parameter $C_{\text{Change}}$. The final curve fit (solid red line) is obtained by averaging the bootstrap fits at every discretized concentration. The 90% two-sided prediction intervals (dashed red lines) are the percentiles of the mixture of the predictive distributions over the bootstrap fits. The resulting curve more accurately fits the data, and better approximates the canonical calibration curve than the linear or logistic fits in Figure 2. (a) Zoom into low concentrations. (b) Overall view.

Figure 4: Comparison between the linear fit and the proposed approach. SRM2 experiment, peptide GFSVVADTPELQR. The proposed approach fit the data more accurately, and better approximated the canonical calibration curve than the linear fit. This increased the LOB estimate by a factor of 1.3, and the LOD estimate by a factor of 1.8. (a) Linear curve fit. (b) Proposed approach.

Figure 5: Comparison between the logistic curve fit and the proposed approach. DIA experiment, peptide SSAAPPPPR. The proposed approach fit the data more accurately, and better approximated the canonical calibration curve than the logistic fit. This increased the LOB estimate by a factor of 1.45, and the LOD estimate by a factor of 1.65. (a) 5-parameter logistic curve fit. (b)
Proposed approach.

Figure 6. Comparative accuracy of the linear, 5-parameter logistic, and proposed approaches. Simulated dataset SIM1. (a) Visualization of one simulation instance, and of the fits by the three methods (zoomed into the low concentrations). The fit by the proposed approach closely approximates the true model. The numeric values of the figures of merit, averaged over 100 simulations, are detailed in Table 2. (b) Visualization of accuracy and the error of the LOB and LOD estimation by the three methods, summarized over 100 simulations. The %Error is defined as 
\[ \left( \frac{\text{estimated LOB,D} - \text{true LOB,D}}{\text{true LOB,D}} \times 100 \right) \]. The proposed approach was slightly conservative, and substantially reduced both the magnitude and the variability of the estimation error.

Figure 7. Summary of LOB and LOD estimates by the linear, 5-parameter logistic, and proposed approaches. Each box summarizes the ratios between the estimates by the 5-parameter logistic curve fit (green) or the proposed fit (red) to the estimates by the linear fit, over all the peptides in each assay development and evaluation experiment. The horizontal line indicates a perfect agreement with the conclusions from the linear fit. Values above the horizontal line indicate estimates that are more conservative than the linear fit. As compared to the linear fit, the proposed approach substantially reduced the optimism regarding the figures of merit in all the datasets. As compared to the logistic curve fit, it reduced the optimism for many datasets. (a) LOB. (b) LOD. The LOB plot shows for each data set the % of peptides where non-linearity was present in more than 50% of bootstrap iterations. A greater percentage indicates a greater number of peptides with a non-linear characteristic in a particular dataset. Whenever this percentage is large, the linear method substantially underestimates the LOB and LOD at the dataset level compared to the proposed.

Figure 8. Estimation of the figures of merit in an assay used for large clinical research studies (a) Estimation based on the linear fit. Insert: zoom into low concentrations. (b) Estimation based on the 5-parameter logistic and on the proposed fit. Due to the lack of intensities in the saturation range, the logistic curve fit was unstable. It failed to adequately capture the leveling off at low concentrations, resulted in an overly wide prediction interval, and could not produce an estimate of LOD. In contrast, the proposed approach accurately represented the data, and provided estimates of both LOB and LOD. While the estimates of LOB were the same for the two fits for this particular peptide, this is not generally the case and depends on patterns of variation in the data.
10 Tables

| Experiments | Acquisition | # peptides/# proteins | # replicates | # blank samples | # concentrations |
|-------------|-------------|-----------------------|--------------|-----------------|------------------|
| DIA [Suppl. Sec. 1.2] | DIA | 96/6 | 3 | 3 | 10 |
| PRM [25] | PRM | 43/37 | 3 | 4 | 8 |
| SRM1 [19] | SRM | 115/34 (6 sites) | 3 | 3 | 10 |
| SRM2 [37] | SRM | 343/79 | 3 | 5 | 17 |
| SRM3 [38] | SRM | 154/83 | 3 | 3 | 9 |
| SRM4 [38] | SRM | 153/79 | 3 | 3 | 9 |
| SRM5 [37] | SRM | 45/22 | 3 | 5 | 9 |
| SRM6 [39] | SRM | 40/22 | 3 | 5 | 9 |
| SRM7 [31] | SRM | 11/11 (3 repeats) | 4 | 4 | 10 |
| SRM8 [Suppl. Sec. 1.1] | SRM | 15/2 | 4 | 5 | 10 |
| CL [2] | SRM (Clinical) | 5/1 | 5 | 5 | 17 |
| SIM1 [Suppl. Sec. 1.4.1] | Simulated | 1/1 (100 repeats) | 4 | 5 | 10 |
| SIM2 [Suppl. Sec. 1.4.2] | Simulated | 1/1 | 4 | 5 | 10 |
| SIM3 [Suppl. Sec. 1.4.3] | Simulated | 1/1 | 4 | 5 | 10 |

Table 1 Datasets used to evaluate the proposed approach.

| Method | LOB | LOD |
|--------|-----|-----|
|        | Value | Error | CV | Value | Error | CV |
| True   | 166  | -63% | 34% | 183   |       |     |
| Linear | 62   | -53% | 25% |       |       |     |
| Logistic | 124 | -25% | 18% | 171   | -6%  | 12% |
| Proposed | 169 | 1%   | 8%  | 190   | 4%   | 3%  |

Table 2 Accuracy of assay characterization, evaluated on the simulated data set SIM1. The %Error is calculated as $\left(\frac{\text{estimated LOB,D} - \text{true LOB,D}}{\text{true LOB,D}} \times 100\right)$. The % coefficient of variation CV is defined as $\left(\frac{\text{stand. dev of estimated LOB,D}}{\text{mean of estimated LOB,D}} \times 100\right)$.
Figure 1

Figure 2
Figure 3

Figure 4
Figure 5

(a) Spiked Concentration vs. Intensity with 95% noise, 90% prediction, and mean prediction. LOB = 128, LOD = 202.

(b) Concentration vs. Intensity with 95% noise, 90% prediction, and mean prediction. LOB = 177, LOD = 350.

Figure 6

(a) Concentration vs. Intensity for Exact, Linear, Logistic, and Proposed methods. % Error for Linear, Logistic, and Proposed methods.

(b) Box plots of % Error for Linear, Logistic, and Proposed methods with LOB and LOD.
LOD estimates, proposed or logistic fit / LOD estimates, linear regression

Figure 7
**Figure 8**

(a) Linear fit: \( I = 0.0052 + 0.0041 \, C \)

LOB= 0.93, LOD=1.03

(b) Logistic, LOB=0.96, LOD=N/A

Proposed, LOB=0.93, LOD=1.03

95% Noise