This is the postprint version of the following article: Mona M. Khamis, Nancy Klemm, Darryl J Adamko and Anas El-Aneed (2018). Comparison of accuracy and precision between multipoint calibration, single point calibration and relative quantification for targeted metabolomic analysis. Anal Bioanal Chem. 2018 Sep;410(23):5899-5913. Published in its final version at doi: 10.1007/s00216-018-1205-5.

Comparison of accuracy and precision between multipoint calibration, single point calibration and relative quantification for targeted metabolomic analysis

Mona M. Khamis[1], Nancy Klemm[2], Darryl J Adamko[3] and Anas El-Aneed[1*]

[1]College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada.
[2] Brandenburg University of Technology Cottbus-Senftenberg, Germany
[3]Department of Pediatrics, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada.
* Phone number: (306) 966-2013, email address: anas.el-aneed@usask.ca

Abstract

Targeted metabolomics requires accurate and precise quantification of candidate biomarkers, often through tandem mass spectrometric (MS/MS) analysis. Differential isotope labeling (DIL) improves mass spectrometric (MS) analysis in metabolomics by derivatizing metabolites with two isotopic forms of the same reagent. Despite its advantages, DIL-liquid chromatographic (LC)-MS/MS can result in substantial increase in workload when fully validated quantitative methods are required. To decrease the workload, we hypothesized that single point calibration or relative quantification could be used as alternative methods. Either approach will result in significant saving in resources and time. To test our hypothesis, six urinary metabolites were selected as model compounds. Urine samples were analyzed using a fully-validated multipoint dansyl chloride-DIL-LC-MS/MS method. Samples were reprocessed using single point calibration and relative quantification modes. Our results demonstrated that the performance of single point calibration or relative quantification was inferior, for some metabolites, to multipoint calibration. The lower limit of quantification failed in the quantification of ethanolamine in most of participant samples using single point calibration. In addition, its precision was not acceptable in one participant during serine and ethanolamine quantification. On the other hand, relative quantification resulted in the least accurate data. In fact, none of the data generated from relative quantification for serine was comparable to that obtained from multipoint calibration. Finally, while single point calibration showed an overall acceptable performance for the majority of the model compounds, we cannot extrapolate the findings to other metabolites within the same analytical run. Analysts are advised to assess accuracy and precision for each metabolite in which single point calibration is the intended quantification mean.

Keywords: metabolomics; single point calibration; validation; quantification; differential isotope labeling, dansyl chloride

1. Introduction

Metabolomics encompasses the identification and quantification of all possible endogenous and exogenous metabolites in a biological sample [1-3]. Biomarker discovery is a key application in metabolomics that aims to improve disease diagnosis and prognosis while also assessing therapeutic efficacy and toxicity [1-4]. Metabolomics studies flow sequentially from untargeted to targeted platforms [5-7]. Unlike screening in the untargeted approach, targeted quantification of preselected metabolites requires a robust methodology usually with multiple reaction monitoring (MRM) acquisition [6,8,5,7]. Targeted analysis is typically conducted by coupling tandem mass spectrometry (MS/MS) to gas chromatography (GC) or liquid chromatography (LC) [6,8,5,7].

The lack of well-established analytical frameworks for absolute quantification is a bottleneck in the biomarker qualification process [9-13]. Contrary to method validation for xenobiotics, available guidelines [14,15] are not well suited to address all challenges with endogenous metabolite quantification. Consequently, there has been confusion on the extent of method validation needed to meet regulatory requirements [9-12]. A "fit-for-the-purpose" approach, achieving minimum validation criteria, has been deemed adequate for preliminary biomarker discovery experiments. However, methods designated for clinical decision making require full validation to ascertain accuracy, precision, specificity, robustness and stability [9-12].

In comparison with multipoint calibration, single point calibration involves the use of a single reference solution for the quantification of the target analyte. It assumes that the response of the analyte and the detector is linear to the analyte' s concentration and the hypothetical calibration curve has a zero y-intercept [16,17]. Single point calibration has been frequently used [18-25] and it represents a compromise between validation rigidity, workload and speed of data acquisition, thus providing semi-quantitative data [18]. Inaccuracies from single point calibration have been reported, citing low analytes concentration [19] or nonlinear calibration models [20], as possible reasons. On the contrary, it has demonstrated comparable performance to multipoint calibration for quantifying analytes in biological fluids [21-23], pharmaceutical preparations [24] and tissue extracts [25,22].

Differential isotope labeling (DIL) in metabolomics, pioneered by L. Li, has been exploited to address many typically encountered challenges, including matrix interference and internal standard availability [26-28]. Using DIL, the derivatizing reagent is synthesized into light and heavy (deuterium or ^{13}C-labeled) forms [29,27,30,28,31-35]. The simultaneous derivatization of the analyte with both forms generates two isotopologue products that when mixed are detected as a peak pair by mass spectrometry (MS) [29,33,27,34,32,28,31,30]. Derivatizing reagents targeting specific submetabolomes, such as dansyl chloride (DNS-Cl) (alcohols, amines and phenols) [27,29], *p*-dimethylaminophenacyl bromide (acids) [28,30], dansyl hydrazine (carbonyl) [31] and bromoacetonylquinolinium bromide (thiols) [32] are continuously introduced to DIL technique, increasing its usefulness in metabolomics. DIL is used for both absolute and semi quantification purposes (Figure 1) [29,33,30,27,34,32,28,35,31]. However, absolute quantification [28,35-37] has been an unfavorable option for researchers seeking rapid quantitative metabolomics data. This can be attributed to the additional workload for optimizing different reaction conditions along with the routine optimization of chromatographic/MS systems and the extensive validation needed prior to sample analysis [15,14]. Accordingly, DIL has been mostly exploited in semi quantification, or "*relative quantification*" as described within the metabolomics society (Figure 1) [27,32].

We have recently developed an LC-MS/MS method for the absolute quantification of 19 amine- and phenol- containing urinary metabolites using DIL with $^{12}C_2/^{13}C_2$- DNS-Cl [37]. The method was fully validated according to regulatory guidelines [14,15]. However, a typical batch for clinical data acquisition would necessitate around 28 non-patient injections (around 11 hours) of calibration

standards, quality control (QC) samples, blanks and system suitability samples [38,37], leading to an increased analysis time. Accordingly, we decided to pursue a comparative study between three different quantification modes using DIL-LC-MS/MS, where the performance of single point calibration or relative quantification was compared to the validated multipoint calibration method [37]. We hypothesized that these methods could provide accurate and precise analytical data, which would subsequently result in significant reduction in the required resources for high throughput targeted DIL-LC-MS/MS metabolomics. However, our hypothesis was not accurate for the entire set of metabolites selected for this study and the opposite was proven true. This work represents the first comparison of the analytical performance of three DIL-based targeted methods for metabolomics application.

2. Experimental

2.1. Materials and reagents

$^{12}C_2/^{13}C_2$- DNS-Cl were synthesized as previously described [37]. All reagents and reference standards are detailed in our recent work [37].

2.2. Urine sample collection

Participants (n=7) were enrolled after obtaining their written informed consent as approved by the University of Saskatchewan's biomedical research ethics board (Bio# 13-89). Random mid-stream urine samples were collected from healthy male participants (25-40 years of age) currently not taking any form of medications. Urine specimen cups (Starplex Scientific Inc, ON, CA) were frozen at -80 ºC shortly after collection and samples were subjected to 1 freeze-thaw cycle in which they divided into aliquots in 1.5 mL micro centrifuge tubes (Fischer, CA). A pooled urine sample was prepared by mixing equal aliquots from 5 participants.

2.3. Standards derivatization

The preparation of underivatized individual standards, underivatized working standards mixture and derivatized standards mixture was done as previously reported [37]. Briefly, for the derivatized standards, i.e. $^{12}C_2$-DNS-analytes stock solution preparation; 50 µL from the working stock solution of the 19 standards were mixed with 30 µL bicarbonate/carbonate buffer (pH 9.4, 0.5 M) and 40 µL $^{12}C_2$-DNS-Cl (10.13 mg/mL in acetonitrile (ACN)). The mixture was heated at 60 ºC for 30 minutes and excess DNS-Cl was quenched with 10 µL 0.25 M NaOH with further heating at 60 ºC for 10 min. Seventy µL of 300 mM FA in 50% ACN were added for medium acidification [37]. These derivatization steps, with the use of the appropriate isotopic form of DNS-Cl, are common between $^{12}C_2$-DNS-analytes, $^{13}C_2$-DNS-ISs and $^{12}C_2/^{13}C_2$- urine samples.

Only for the preparation of $^{13}C_2$-DNS-ISs stock solution, used in multipoint and single point calibrations, the addition of 300 mM FA in 50% ACN was followed by further dilution with 50 µL of 50% ACN. This dilution step was previously optimized in order to achieve the validated concentration of the ISs in the final mixtures [37]. Ten µL aliquots from the aforementioned solution were used for spiking purposes in single and multipoint calibrations [37]. We have previously validated the suitability of the "surrogate urine" in simulating real derivatized urine [37]. The surrogate urine is used for the preparation of multipoint calibration standards and QC samples and it is prepared following the same derivatization steps described for $^{12}C_2$-DNS analytes solution. However, 50 µL of pooled urine sample (7 patients) were used instead of the working stock solution and 30 µL ACN substituted the derivatizing reagent [37]. This matrix is meant to compensate for the absence of metabolite-free urine needed for method validation.

For relative quantification, only $^{12}C_2$-DNS-analytes solution is needed and was prepared similar to multipoint calibration. The produced $^{12}C_2$-DNS-analytes solution was further mixed with equal volume from $^{13}C_2$-DNS-pooled urine prepared as described below.

2.4. Urine derivatization

Urine aliquots were thawed to room temperature, diluted two-fold with ACN, vortexed and centrifuged at 13,000 rpm for 10 min prior to derivatization. Multipoint and single point quantification required the preparation of $^{12}C_2$-urine samples [37]. Fifty µL of the supernatant solution of each urine sample (7 participants and one pooled urine sample) were derivatized as described for $^{12}C_2$-DNS analytes solution. Following the addition of 300 mM FA in 50% ACN, 50 µL of each reacted urine mixture were diluted with 50 µL 50% ACN containing 10 µL $^{13}C_2$-DNS-ISs**.** The solutions were transferred into HPLC vials equipped with 100 µL glass inserts for analysis [37]. For relative quantification, additional $^{13}C_2$- pooled urine was prepared. After the addition of 300 mM FA in 50% ACN in $^{12}C_2$-urine or $^{13}C_2$- pooled urine, equal portions of both isotopic solutions were mixed for HPLC analysis.

2.5. Instrumentation

The validated method developed in our previous work [37] has been partially validated on a 1290 Agilent UPLC system (Mississauga, ON, Canada) interfaced to an AB Sciex 6500 API QTRAP instrument (AB Sciex, Concord, ON, Canada). A 1290 Agilent autosampler set at 4 ºC was used to deliver 5 µL samples for separation on a Kinetex C18 column (100 mm×2.1 mm, 5 µm ID, 100 Å pore size, Phenomenex, Canada), maintained at 22 ºC. The binary mobile phase system, composed of (A) 5% acetonitrile (ACN) in 0.1% formic acid (FA) and (B) 0.1% FA in ACN, was used for metabolites separation in 25 min. At a flow rate of 250 µL/min, the gradient system was as follows; t=0, 90% A, t=6, 85% A, t=19, 35% A, t=20, 35% A, t=20.1, 90% A and t=25, 90% A.

Quantification was achieved using MRM scan mode and positive electrospray ionization (ESI). The monitored precursor ion → product ion transition for each $^{12}C_2$-DNS-analyte (dansyl derivatized metabolite), was *m/z* $[M+H]^+$→ *m/z* 170.10, while the analogues $^{13}C_2$-DNS- internal standards (ISs) were monitored at *m/z* $[M+H]^+$→ *m/z* 172.10. A qualifier diagnostic product ion was monitored for each $^{12}C_2$-DNS-analyte to confirm its identity, shown in Table 1. The following parameters were optimized: turbo spray ion source temperature= 550 ºC, ion spray voltage= 5.5 kV, curtain gas= 30, collision gas= 9, nebulizer gas= 50, heater gas= 50, entrance potential=10, collision exit potential= 13, total dwell time= 20 msec and cycle time=1.4504. The collision energy (CE) and declustering potential (DP) were separately optimized for each analyte (Table 1). Data processing was achieved on Analyst software, version 1.6.2 (AB Sciex, Concord, ON, Canada) and SPSS software (version 24).

2.6. Metabolite selection

Based on our previous experience with human urine quantitative analysis [37], we chose 4 metabolites: valine (VAL), tryptophan (TRP), serine (SER) and ethanolamine ($ETNH_2$). Each met at least three of the following criteria: (A) the metabolite is typically present in urine below its upper limit of quantification (ULOQ), and does not require additional sample dilution; (B) the metabolite is not seen at a concentration below the lower limit of quantification (LLOQ); (C) the metabolite's concentration is known to vary greatly among participant samples regardless of the participant's hydration status; and (D) the average concentration of the metabolite in pooled urine sample was in the middle range of the calibration curve.

The performance of the single point and relative quantification methods was further challenged through the inclusion of sarcosine (SAR) and lysine (LYS), which we have found to be present in urine towards the lower end of their linearity ranges. The selection criteria were biased towards the "best performing" metabolites. Such bias ensured that the variations in metabolite concentration would reflect true differences in performance across different quantification methods. Metabolites that might compromise the power of the analysis due to their trace levels (below LLOQ) were excluded, so as those that might introduce an error in the comparative study due to their additional sample preparation (above ULOQ).

2.7. Metabolite quantification

A. Multipoint calibration

Absolute quantification requires the use of calibration and QC samples (Figure 1A). Preparation of such solutions was done in accordance with our published protocol [37]. Eight-point calibration curves were generated by plotting peak area ratios of $^{12}C_2$-DNS analytes and $^{13}C_2$-DNS ISs against the corresponding $^{12}C_2$-DNS analytes concentrations (Table 1). Linear regression equations were generated using the sum of least squares with $1/x^2$ weighing; a factor previously optimized during method validation [37]. QC samples were prepared at three different levels; lower QC (LQC), middle QC (MQC) and high QC (HQC) (Table 1).

B. Single point calibration

Raw data generated from Analyst software, version 1.6.2 was reprocessed for single point calibration. Since in real practice, each calibration curve point or QC sample can serve as a calibrator for analyte quantification (Figure 1A), metabolites were quantified using all calibration points and all three QC levels. There are differences in the linearity range optimized for each metabolite based on its levels in urine [37]. As such, the concentration of the calibrators are expressed in terms of fold concentration to the LLOQ. Calibrators [200×] to [1×], represent the eight calibration solutions, where [200×] corresponds to the ULOQ and [1×] represents LLOQ (Table 1). Peak area ratios of $^{12}C_2$-DNS-analytes to $^{13}C_2$-DNS-ISs in calibrators and individual urine samples were the basis for metabolite quantification (section 2.8).

C. Semi "Relative" quantification

Relative quantification was achieved in two sequential steps as previously reported (Figure 1B) [27]. In the first step, the selected metabolites were quantified in pooled urine relative to their standards. In the second step, the quantified metabolites in pooled urine were used to determine their concentration in individual urine samples. For the quantification of metabolites in pooled urine (Step 1), 3 different volumes from $^{12}C_2$-DNS-analytes solution, producing final concentrations equivalent to calibrators [133.3×], [83.3×] and [2.67×], were properly diluted to 50 µL with 50% ACN and were separately mixed with 50 µL of $^{13}C_2$- pooled urine sample. The concentrations of the selected $^{13}C_2$-DNS-urinary metabolites were determined using absolute peak area comparison against each $^{12}C_2$-DNS-analyte calibrator (section 2.8). In step 2, 50 µL of $^{13}C_2$- pooled urine sample were mixed with an equal volume of individual $^{12}C_2$- urine sample (Figure 1B). Quantification of metabolites in individual urine samples was also achieved through absolute peak area comparison (section 2.8).

2.8. Statistical data analysis

Data processing for absolute quantification was conducted using Analyst software, version 1.6.2. Raw data was further exported to Excel (Microsoft office 2013) for the calculation of metabolite

concentrations in single point calibration mode (equation 1). Raw data in relative quantification was processed based on equations 2 and 3. The final concentrations of all metabolites were further exported to SPSS program (version 24) to create figures.

$$equation\ 1: Conc.(Urine) = \frac{Peak\ Area\ Ratio\ (Urine)\ \times\ Conc.(Calibration\ Standard)}{Peak\ Area\ Ratio\ (Calibration\ Standard)}$$

$$equation\ 2: Conc.(Pooled\ Urine) = \frac{Peak\ Area\ (Pooled\ Urine) \times Conc.(Calibration\ Standard)}{Peak\ Area\ (Calibration\ Standard)}$$

$$equation\ 3: Conc.(\ Individual\ Urine) = \frac{Peak\ Area\ (\ Individual\ Urine) \times Conc.(Pooled\ Urine)}{Peak\ Area\ (Pooled\ Urine)}$$

3. Results and Discussion

This work was designed to address four questions in a DIL-LC-MS/MS targeted metabolomic workflow used for biomarker qualification. First, *if an analytical method has been fully validated using multipoint DIL-LC-MS/MS, could metabolites be quantified with similar accuracy and precision, using single point calibration?* This would substantially reduce instrument time, the number of calibration standards, and the QC samples to be prepared. Second, *would the choice of the single calibrator point (low vs. high and/or proximity to analyte concentration in a sample) affect the accuracy?* Third, *if the DIL-LC-MS/MS method was used with relative quantification, would the knowledge of the dynamic range of the mass spectrometer along with the optimization of the derivatization reaction solely suffice for the acquisition of data with acceptable accuracy and precision?* Finally, *would the exclusion of individual urine samples from the pooled urine sample compromise the accuracy and precision of relative quantification?* This is particularly important when new participants are still being recruited during the biomarker validation process.

3.1. Multipoint calibration method

Table 1 summarizes the optimized MS parameters and the linearity ranges [37] of the selected metabolites. Figure 2 shows two representative calibration curves for $ETNH_2$ and SER. All eight calibration points for all metabolites were within 15% of their respective nominal values, indicating acceptable calibration curves [15,14] for quantification (data not shown). Table 2 compiles the accuracy and precision of the LQC, MQC and HQC samples, obtained during participant sample analysis. As seen in Table 2, all QC samples were within 15% of their respective nominal values. In addition, their relative standard deviation (RSD%) was below 15%, indicating that the analytical runs met regulatory guidelines during clinical data acquisition [15,14]. Details on the validation process of the multipoint calibration method are beyond the scope of this article, however, readers are advised to refer to our previously published work [37].

Urine samples from seven individuals and a pooled urine sample from five participants were analyzed using our previously validated multipoint calibration DIL-LC-MS/MS method [37]. Table 3, summarizes the concentrations of the metabolites in these samples. Calculated RSD% demonstrates high precision of the measurements (below 15%), with the exception of SER in participant 5 in which RSD% was 15.9% (Table 3). The concentration of TRP in participant 3 was above the ULOQ and was omitted from the calculations. Similarly, SAR in participant 1 was lower than the LLOQ and data from this participant for SAR was excluded (Table 3). As expected, the selected metabolites with the exception of SAR and LYS vary greatly in their concentration among the participant samples as well as in their distribution within their calibration ranges (Figure 3).

3.2. Single point calibration method

A. Results using QC samples

Single point calibration was achieved by mathematical manipulation of the raw dataset obtained from the multipoint calibration method (equation 1). Similar to multipoint calibration where QC samples are used for quality assessment, QC samples were also used to evaluate the performance of the eight individual calibration standards as single point calibrators (Table 4). The acceptance criteria established for bioanalytical method validation was adopted [14,15], in which accuracy values within ±15% and RSD% below 15% are deemed acceptable. The performance of single point calibration was found inferior to that of the multipoint calibration for LYS, SER and $ETNH_2$, in which 3 out of 8 calibrators failed to demonstrate acceptable accuracy in at least one QC measurement (Table 4). Calibrator [33.3×] resulted in unacceptable accuracy (115.5% and 117.8%) at the LQC level for SER and $ETNH_2$, respectively, while in LYS, calibrator [3.3×] resulted in 115.8% accuracy at the MQC level. These deviated results would have been anticipated if the observed values for calibrators [33.3×] and [3.3×] were different than the calculated values provided by the regression model (Figure 2), which was not the case. On the other hand, the accuracy of calibrator [1×] (equivalent to LLOQ) was acceptable for all metabolites with the exception of $ETNH_2$, in which accuracies less than 85% were obtained for the MQC and HQC levels (Table 4).

As for the methods' precision, the criterion in all metabolites was met at all levels, and therefore was comparable to multipoint calibration from a validation standpoint (Table 4). However, one interesting observation is that the highest RSD% variation for the LQC and HQC in SAR, VAL, TRP, $ETNH_2$ and LYS were mostly generated from calibrator [1×] (i.e. LLOQ). This information is valuable as it indicates higher chances for imprecision of this calibrator in single point calibration (Table 4). Overall, the values generated from multipoint calibration were more accurate and precise that that generated from single point calibration (Tables 2 and 4).

B. Results using participants' samples

The accuracy and precision of the participants' data generated using single point calibration was evaluated. Metabolites were quantified using individual calibrators [1×] through [200×]. In addition, the three QC levels, [166.7×], [83.3×] and [2.7×], were also used for quantification, as any standard solution can serve as a calibrator for data acquisition in single-point calibration. Concentration of the metabolites generated from multipoint calibration were used as a benchmark and compared against those generated using single point calibration. With the exception of $ETNH_2$, all calibrators resulted in acceptable accuracies, in which bias values were less than ±15% [15,14] (Figure 4). VAL concentration in participant 3, generated from calibrator [33.3×], had the highest deviation from multipoint calibration (-14.8% bias) (Figure 4). The average bias values among all participants for SAR, VAL, TRP and LYS were the highest with calibrator [3.3×]. Contrary to expectation [39], we have found that the choice of calibrator that is in close proximity to the concentration of the analytes in urine samples did not result in the most accurate results for low concentration metabolites.

Another concerning finding for accuracy using single point calibration were observed with $ETNH_2$. First, calibrator [1×; i.e LLOQ] resulted in unacceptable accuracy values among the majority of the samples, with an average bias of -18.3% (Figure 4). This can be explained by the fact that the LLOQ is usually the calibration point in which validation criteria [15,14] allows for wider accuracy acceptance limit (±20% instead of ±15%). The second observation is that 7 out of the 10 calibrators (excluding [1×]) failed to demonstrate acceptable accuracy in $ETNH_2$ measurements in pooled urine, where a maximum bias of -21.3% was observed with calibrator [2.7×] (Figure 4). We could not draw an explanation for this observation. In fact, the inadequate performance of most calibrators in measuring

$ETNH_2$ had a direct impact on the conclusion of accuracy for single point calibration. As such, we cannot exclude the probability that other metabolites [37], not investigated within this study, might not be correctly quantified when single point calibration is used. Therefore, it is extremely important to investigate the performance of each metabolite, individually, using various calibrators prior to the consideration of single point calibration for regular data acquisition.

We then looked at the precision of single point calibration, and found it inferior to multipoint calibration for $ETNH_2$, LYS and SER (Figure 5). Calibrator [1×] resulted in marginally unacceptable RSD% of 15.9% and 15.3% for $ETNH_2$ and LYS, respectively. This was not unexpected from this LLOQ calibrator, where higher RSD% values (<20%) are acceptable by the regulatory agencies during method validation of multipoint calibration [15,14]. In addition to $ETNH_2$ and LYS, the highest RSD% values (still less than 15%) were observed with calibrator [1×] in TRP and SAR (Figure 5). Again, the relative inferior precision of this calibrator is important for LYS and SAR, in which a calibration point close to the analyte in real samples might not be the optimum point for precise data generation. Regardless of the calibrator used, the RSD% values in SER were similar per participant (Figure 5). Moreover, most calibrators resulted in unacceptable precision of SER measurements in participants 2 and 5 (RSD%; 15.2%-17.1%). In general, while the eleven calibrators resulted in acceptable precision in the majority of the metabolites in the analyzed samples, the RSD% values were relatively higher in single point calibration in comparison with multipoint calibration (Figure 5)

3.3. Relative quantification

We tailored our validated LC-MS/MS method [37] to quantify metabolites in a relative mode using equations 2 and 3 [27] (Figure 1B), rather than the use of $^{13}C_2$-isotopic form as an internal standard (Figure 1A). We hypothesized that the proximity of metabolites concentrations in pooled urine sample to that of the spiked $^{12}C_2$-DNS-analytes could also influence the accuracy and precision of the measurements. $^{13}C_2$-DNS-pooled urine was first quantified against $^{12}C_2$-DNS-analytes solution. The solutions were mixed at equal volumes, however, the concentration of the standard solution was varied to produce high, medium and low calibrators; [133.3×], [83.3×] and [2.7×]. Following the quantification of metabolites, pooled urine aliquots were then separately mixed with equal volumes of $^{12}C_2$-derivatized individual urine samples (Figure 1B).

We were also concerned that the exclusion of individual urine samples from the pooled urine sample might compromise the accuracy and precision of relative quantification. This is particularly important when new participants are still being recruited during the biomarker validation process. In fact, the inability to pool all samples prior to analysis, for instance due to different times and locations of analysis, has been proposed as a potential impediment to the use of pooled control urine in large scale metabolomics [40]. As such, we created a sample pooled from five participants and left two out for later individual comparison. The pooled urine was then used to quantify all individual urine samples, including the excluded samples.

Relative quantification was very precise but showed inaccuracies in measurements more than single point calibration (Figure 6). With the exception of VAL and $ETNH_2$, all metabolites had at least one calibrator with inaccurate data (greater than ±15% deviation from multipoint calibration). Calibrators [133.3×] and [83.3×] failed to produce accurate quantitative data (i.e. within ± 15%) for SAR, LYS, and TRP in at least 3 out of 8 samples. On the other hand, the lowest calibrator, i.e., [2.7×] was the only standard producing accurate measurements for LYS, SAR and TRP in all urine samples (Figure 6). While this observation can be justified by the low concentration of LYS and SAR in individual and pooled urine, an extrapolation to TRP, a widely distributed metabolite among patients

and its linearity range (Figure 3), was not applicable. Finally, none of the tested calibrators produced accurate results in SER measurements in 6 out of 8 samples (Figure 6).

In contrast to accuracy, precision of relative quantification was comparable to that of multipoint calibration. As shown in the error bars of Figure 6, all measurements were deemed precise with a RSD% of less than 15%. An exception was observed with calibrators [2.7×] and [133.3×] that resulted in high RSD % (18.3% and 16%) in participants 7 and 2, respectively during SER quantification. Interestingly, unlike single point calibration, participant 7 consistently demonstrated the highest RSD% in all metabolites and with all calibrators. This observation could be attributed to the exclusion of this participant from the pooled urine sample. Therefore, the pooled matrix might not reflect the real matrix effect experienced within this participant. However, since this observation was not consistent with participant 6, who was also excluded, random error could also justify the higher dispersion of RSD%. The limited number of excluded participant samples hinders the decisive conclusion on such observation.

4. Conclusion

Our comparison of single point calibration and relative quantification to a multipoint validated DIL-LC-MS/MS method demonstrated shortcomings. We initially hypothesized that single point calibration would serve as an excellent substitute to multipoint calibration in the selected analytes and would not necessitate further testing of other metabolites [37]. Unfortunately, this hypothesis was not correct due to inconsistent performance across all investigated metabolites. Therefore, we recommend that if single point calibration is to be used for clinical data acquisition, it requires assessment on case-by-case basis. This would require investigation of the quality of the data for each metabolite using larger cohorts of samples. In addition, the results herein cannot be extrapolated to the same metabolites when a different derivatizing reagent or matrix is employed.

The calibration mode had more influence on the accuracy of the data than its precision. Surprisingly, we did not find the accuracy of single point calibration to be dependent on the closeness of the analyte concentration to that in the sample. Further, the inaccuracy with $ETNH_2$ emphasizes that results from one calibrator might not be applicable to other metabolites beyond this study. In general, we recommend against the use of the LLOQ, even if the metabolite of interest lies within the lower range of the linear curve. In addition, unless established via experimentation that a specific calibrator is inaccurate and/or imprecise, any calibrator could theoretically serve as an adequate single point calibration standard for quantification, regardless of its concentration proximity to the metabolite of interest.

In regard to relative quantification, this method was less accurate in comparison to single point calibration. However, It has value in preliminary semi-quantitative analysis (e.g. screening metabolites for possible biomarker targets), which then would require validation. One possible explanation is the dependence of the results on the values obtained from pooled urine. Errors within the pooled urine metabolite quantification are further augmented in individual samples. Therefore, the quantification of the metabolite in the pooled urine sample represents the most critical step, upon which the accuracy of the rest of the analyses are dependant. For instance, in SAR, LYS, VAL and TRP, the calibrator that resulted in the most accurate quantification values in pooled urine was also associated with the highest accuracy in participant samples. Using the absolute peak areas can also be another potential source of error, where the matrix in which the pooled urine sample is quantified, is different from that of the individual urine samples. Exclusion of samples from the pooled matrix can also be a potential source of lower quality data [40].

Finally, this study was biased towards metabolites whose concentrations were already known to typically fall within the constructed calibration curve. While an inconsistent difference in the investigated quantification methods was demonstrated, such difference is expected to be further amplified in situations with highly concentrated metabolites that typically require dilution (e.g. histidine, alanine and glycine) [37]. The same would apply to metabolites which frequently fall below their LLOQ (e.g. 1-methylhistamine) [37]. While a general conclusion on relative quantification or single point calibration cannot be made, based on our study findings, we advise the analytical community to exercise caution and rigor assessment when employing single point calibration or relative quantification for targeted urine metabolomics.

5. Acknowledgment

The authors acknowledge the valuable comments and discussions with Dr. Randy Purves, University of Saskatchewan, concerning the project. Ms. Khamis thanks the College of Pharmacy and Nutrition for scholarship funding. Funding for the project was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant and the QTRAP 6500 instrument was purchased through a Western Diversification Grant from the Western Economic Diversification Canada.

6. Conflict of interest

The authors declare no conflict of interest.

References

1. Zhang A, Sun H, Wu X, Wang X (2012) Urine metabolomics. Clinica Chimica Acta 414:65-69
2. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL (2011) Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chemical Society Reviews 40 (1):387-426
3. Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. TrAC Trends in Analytical Chemistry 24 (4):285-294
4. Mamas M, Dunn WB, Neyses L, Goodacre R (2011) The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of toxicology 85 (1):5-17
5. Mikami T, Aoki M, Kimura T (2012) The application of mass spectrometry to proteomics and metabolomics in biomarker discovery and drug development. Current molecular pharmacology 5 (2):301-316
6. Xiao JF, Zhou B, Ressom HW (2012) Metabolite identification and quantitation in LC-MS/MS-based metabolomics. TrAC Trends in Analytical Chemistry 32:1-14
7. Gika HG, Theodoridis GA, Plumb RS, Wilson ID (2014) Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. Journal of pharmaceutical and biomedical analysis 87:12-25
8. Roberts LD, Souza AL, Gerszten RE, Clish CB (2012) Targeted metabolomics. Current Protocols in Molecular Biology:30.32. 31-30.32. 24
9. Lee JW, Weiner RS, Sailstad JM, Bowsher RR, Knuth DW, O'Brien PJ, Fourcroy JL, Dixit R, Pandite L, Pietrusko RG (2005) Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. Pharmaceutical research 22 (4):499-511
10. Lee JW, Smith WC, Nordblom GD, Bowsher RR, Bloom J, Dean R (2003) Validation of assays for the bioanalysis of novel biomarkers: Practical recommendations for clinical investigation of new drug entities. Drugs and the Pharmaceutical Sciences 132:119-148
11. Lee JW, Devanarayan V, Barrett YC, Weiner R, Allinson J, Fountain S, Keller S, Weinryb I, Green M, Duan L (2006) Fit-for-purpose method development and validation for successful biomarker measurement. Pharmaceutical research 23 (2):312-328
12. Timmerman P, Anders Kall M, Gordon B, Laakso S, Freisleben A, Hucker R (2010) Best practices in a tiered approach to metabolite quantification: views and recommendations of the European Bioanalysis Forum. Bioanalysis 2 (7):1185-1194
13. Beger RD, Colatsky T (2012) Metabolomics data and the biomarker qualification process. Metabolomics 8 (1):2-7
14. Administration UFaD (2001) FDA guidance for industry: bioanalytical method validation. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research: Rockville, MD
15. Agency EM (2011) Committee for Medicinal Products for Human Use (CHMP). Guideline on Bioanalytical Method Validation
16. Cuadros-Rodríguez L, Gámiz-Gracia L, Almansa-López EM, Bosque-Sendra JM (2001) Calibration in chemical measurement processes. II. A methodological approach. TrAC Trends in Analytical Chemistry 20 (11):620-636
17. Cardone MJ, Palermo PJ, Sybrandt LB (1980) Potential error in single-point-ratio calculations based on linear calibration curves with a significant intercept. Analytical Chemistry 52 (8):1187-1191
18. Schymanski EL, Singer HP, Longrée P, Loos M, Ruff M, Stravs MA, Ripollés Vidal C, Hollender J (2014) Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. Environmental science & technology 48 (3):1811-1818

19. Peters FT, Maurer HH (2007) Systematic comparison of bias and precision data obtained with multiple-point and one-point calibration in six validated multianalyte assays for quantification of drugs in human plasma. Analytical chemistry 79 (13):4967-4976
20. Liu RH, McKeehan AM, Edwards C, Foster G, Bensley W, Langner J, Walia A (1994) Improved gas chromatography/mass spectrometry analysis of barbiturates in urine using centrifuge-based solid-phase extraction, methylation, with d 5-pentobarbital as internal standard. Journal of Forensic Science 39 (6):1504-1514
21. Wang S, Magill JE, Vicente FB (2005) A fast and simple high-performance liquid chromatography/mass spectrometry method for simultaneous measurement of whole blood tacrolimus and sirolimus. Archives of pathology & laboratory medicine 129 (5):661-665
22. Speed D, Dickson S, Cairns E, Kim N (2001) Analysis of paracetamol using solid-phase extraction, deuterated internal standards, and gas chromatography-mass spectrometry. Journal of analytical toxicology 25 (3):198-202
23. Bjørk MK, Nielsen MK, Markussen LØ, Klinke HB, Linnet K (2010) Determination of 19 drugs of abuse and metabolites in whole blood by high-performance liquid chromatography–tandem mass spectrometry. Analytical and bioanalytical chemistry 396 (7):2393-2401
24. Leito S, Mölder K, Künnapas A, Herodes K, Leito I (2006) Uncertainty in liquid chromatographic analysis of pharmaceutical product: Influence of various uncertainty sources. Journal of chromatography A 1121 (1):55-63
25. Kimura M, Rodriguez-Amaya DB (2002) A scheme for obtaining standards and HPLC quantification of leafy vegetable carotenoids. Food chemistry 78 (3):389-398
26. Xu F, Zou L, Liu Y, Zhang Z, Ong CN (2011) Enhancement of the capabilities of liquid chromatography–mass spectrometry with derivatization: general principles and applications. Mass spectrometry reviews 30 (6):1143-1172
27. Guo K, Li L (2009) Differential 12C-/13C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. Analytical chemistry 81 (10):3919-3932
28. Stanislaus A, Guo K, Li L (2012) Development of an isotope labeling ultra-high performance liquid chromatography mass spectrometric method for quantification of acylglycines in human urine. Analytica chimica acta 750:161-172
29. Zhao S, Luo X, Li L (2016) Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics. Analytical Chemistry
30. Peng J, Li L (2013) Liquid–liquid extraction combined with differential isotope dimethylaminophenacyl labeling for improved metabolomic profiling of organic acids. Analytica chimica acta 803:97-105
31. Zhao S, Dawe M, Guo K, Li L (2017) Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Carbonyl Submetabolome. Analytical Chemistry
32. Liu P, Huang Y-Q, Cai W-J, Yuan B-F, Feng Y-Q (2014) Profiling of thiol-containing compounds by stable isotope labeling double precursor ion scan mass spectrometry. Analytical chemistry 86 (19):9765-9773
33. Zhou R, Guo K, Li L (2013) 5-Diethylamino-naphthalene-1-sulfonyl Chloride (DensCl): A novel triplex isotope labeling reagent for quantitative metabolome analysis by liquid chromatography mass spectrometry. Analytical chemistry 85 (23):11532-11539
34. Guo K, Ji C, Li L (2007) Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples. Analytical chemistry 79 (22):8631-8638
35. Han J, Lin K, Sequeira C, Borchers CH (2015) An isotope-labeled chemical derivatization method for the quantitation of short-chain fatty acids in human feces by liquid chromatography–tandem mass spectrometry. Analytica chimica acta 854:86-94

36. Leng J, Wang H, Zhang L, Zhang J, Wang H, Guo Y (2013) A highly sensitive isotope-coded derivatization method and its application for the mass spectrometric analysis of analytes containing the carboxyl group. Analytica chimica acta 758:114-121
37. Khamis MM, Adamko DJ, El-Aneed A (2017) Development of a validated LC-MS/MS method for the quantification of 19 endogenous asthma/COPD potential urinary biomarkers. Analytica Chimica Acta 989:45-58
38. Briscoe CJ, Stiles MR, Hage DS (2007) System suitability in bioanalytical LC/MS/MS. Journal of pharmaceutical and biomedical analysis 44 (2):484-491
39. Preu M, Petz M (1998) Isotope dilution GC-MS of benzylpenicillin residues in bovine muscle. Analyst 123 (12):2785-2788
40. Peng J, Chen Y-T, Chen C-L, Li L (2014) Development of a universal metabolome-standard method for long-term LC–MS metabolome profiling and its application for bladder cancer urine-metabolite-biomarker discovery. Analytical chemistry 86 (13):6540-6547

Table 1: The MS and regression parameters of the investigated dansylated metabolites

Analyte*	**Sarcosine (SAR)**	**Valine (VAL)**	**Tryptophan (TRP)**	**Serine (SER)**	**Ethanolamine (ETNH2)**	**Lysine (LYS)**
Regression equation	1.63E-03*x*+ 3.12E-04	1.63E-03*x*+ 4.46E-04	7.18E-04*x*+ 3.37E-04	3.77E-04*x* + 6.20E-04	3.41E-04*x*+ 1.71E-03	3.44E-04*x*+ 1.40E-04
r^2	0.9974	0.997	0.997	0.998	0.998	0.9971
Calibration points [× of LLOQ][1]	**Concentrations ng/mL**					
ULOQ [200×]	375	750	1500	3000	3000	4800
[133.3×]	250	500	1000	2000	2000	3200
[75×]	187.5	375	750	1500	1500	2400
[66.7×]	125	250	500	1000	1000	1600
[33.3×]	62.5	125	250	500	500	800
[10×]	18.75	37.5	75	150	150	240
[3.3×]	6.25	12.5	25	50	50	80
LLOQ [1×]	1.875	3.75	7.5	15	15	24
HQC [166.7×]	312.5	625	1250	2500	2500	4000
MQC [83.3×]	156.25	312.5	625	1250	1250	2000
LQC [2.7×]	5	10	20	40	40	64
CE[2]	27	30	39	29	32	38
DP[3]	66	75	60	70	55	80
Q1	323.0	351.1	438.2	339.1	295.1	613.2
Q3[4] Qualifier product ion	157.1	336.1	130.1	324.1	157.1	234.1

* Metabolites quantified as DNS derivatives, [1][200×] corresponds to the upper limit of quantification (ULOQ) and [1×] represents the lower limit of quantification (LLOQ), [2] Collision energy, [3] Declustering potential, [4] Quantifier product ion= *m/z* 170.1

Table 2: Inter-day accuracy and precision of QC samples of the multipoint calibration method during participant sample analysis

	LQC*[1]				MQC*[2]				HQC*[3]			
	Nominal[**]	**Mean**[**] ± SD**	**RSD (%)**	**Accuracy (%)**	**Nominal**[**]	**Mean** ±SD**	**RSD (%)**	**Accuracy (%)**	**Nominal**[**]	**Mean **± SD**	**RSD (%)**	**Accuracy (%)**
SAR	5	5.01±0.25	5.03	100.2	157	163.38±2.84	1.74	104.0	323	306.63±10.19	3.32	98.0
VAL	10	10.20±0.39	3.84	102.0	313	324.25±4.35	1.34	103.5	625	617.5±26.6	4.31	98.9
TRP	20	20.00±0.67	3.37	100.0	625	658.75±16.09	2.44	105.3	1250	1245±58.02	4.66	99.8
SER	40	41.38±1.65	3.98	103.7	1250	1290.00±18.26	1.42	103.5	2500	2485±110.3	4.44	99.4
ETNH2	40	40.80±1.42	3.47	102.0	1250	1297.50±17.08	1.32	103.8	1250	2455±71.88	2.93	98.2
LYS	64	63.98±2.14	3.35	100.0	2000	2072.50±23.63	1.14	104.0	4000	4032.5±133.8	3.32	100.8

*n=4
**ng/ml
1 lower quality control level
2 middle quality control level
3 third quality control level

Table 3: Concentrations of metabolites in the analyzed urine samples using multipoint calibration

	Metabolite	SAR	VAL	TRP	SER	ETNH2	LYS
Participant 1	Mean[1]	**27.5**[2]	723.2	3882.7	14176.0	6720.0	1712.0
	SD	0.2	37.8	121.2	871.8	263.4	83.1
	RSD%	0.7	5.2	3.1	6.1	3.9	4.9
Participant 2	Mean[1]	50.7	3637.3	13557.3	34933.3	26293.3	3674.7
	SD	1.8	193.6	571.6	3841.1	937.5	202.6
	RSD%	3.5	5.3	4.2	11.0	3.6	5.5
Participant 3	Mean[1]	291.7	7056.0	**30506.7**[2]	41386.7	32640.0	17760.0
	SD	4.7	399.7	805.3	2178.2	1270.0	1269.9
	RSD%	1.6	5.7	2.6	5.3	3.9	7.2
Participant 4	Mean[1]	137.9	6080.0	20533.3	40160.0	41760.0	3114.7
	SD	0.9	296.3	184.8	3187.9	1049.2	345.2
	RSD%	0.7	4.9	0.9	7.9	2.5	11.1
Participant 5	Mean[1]	48.4	3477.3	7429.3	26933.3	16586.7	3301.3
	SD	1.8	147.8	451.5	4294.2	756.1	309.6
	RSD%	3.6	4.3	6.1	15.9	4.6	9.4
Participant 6	Mean[1]	70.5	3440.0	9888.0	18293.3	28800.0	3296.0
	SD	1.9	69.7	242.1	1854.4	576.9	194.0
	RSD%	2.8	2.0	2.5	10.1	2.0	5.9
Participant 7	Mean[1]	97.8	3018.7	5776.0	26293.3	18773.3	5130.7
	SD	3.8	92.4	146.6	488.8	402.7	124.3
	RSD%	3.9	3.1	2.5	1.9	2.1	2.42
Pooled Urine	Mean[1]	111.2	4261.3	15173.3	31146	25333.3	6277.3
	SD	6.3	197.5	456.6	2916.8	1512.3	470.4
	RSD%	5.6	4.6	3.0	9.4	5.9	7.5

[1] ng/mL, n=3, [2] values below lower limit of detection or above upper limit of quantification

Table 4: Inter-day accuracy and precision using single point calibration

			Calibrator							
Metabolite	**QC level[1]**		[200×]	[133.3×]	[75×]	[66.7×]	[33.3×]	[10×]	[3.3×]	[1×]
	LQC	% Accuracy	101.6	96.3	96.8	97.8	109.7	98.3	109.2	98.2
		RSD%	3.1	5.8	4.1	3.5	5.8	5.4	2.5	*7.6*
SAR	MQC	% Accuracy	105.1	99.6	100.1	101.2	113.4	101.6	113.0	101.5
		RSD%	1.8	2.8	1.7	1.7	2.8	2.5	4.8	4.5
	HQC	% Accuracy	98.8	93.7	94.1	95.2	106.7	95.6	106.2	95.5
		RSD%	2.6	5.4	3.7	3.0	5.5	5.1	2.1	*7.4*
	LQC	% Accuracy	105.1	99.2	98.1	101.0	111.1	100.3	113.0	97.9
		RSD%	2.4	5.5	3.8	2.6	4.8	3.7	5.1	*6.1*
VAL	MQC	% Accuracy	105.2	99.3	98.1	101.1	111.2	100.4	113.2	98.0
		RSD%	1.2	4.0	2.3	1.2	3.3	2.2	6.2	4.6
	HQC	% Accuracy	100.1	94.5	93.4	96.1	105.8	95.5	107.6	93.2
		RSD%	3.7	6.7	5.1	3.9	6.1	5.0	4.6	*7.3*
	LQC	% Accuracy	108.3	102.7	103.1	104.2	**115.5**	103.9	111.2	96.6
		RSD%	3.6	3.9	3.9	4.0	4.5	4.2	5.0	4.3
SER	MQC	% Accuracy	104.1	98.7	99.2	100.2	111.1	99.9	107.0	92.9
		RSD%	1.1	1.5	1.5	1.7	2.5	2.0	4.2	2.1
	HQC	% Accuracy	100.4	95.2	95.6	96.6	107.1	96.3	103.0	89.6
		RSD%	4.0	4.9	4.8	5.1	6.0	5.5	2.4	5.6
	LQC	% Accuracy	102.3	94.9	96.6	97.6	109.2	98.9	109.4	97.3
		RSD%	2.8	3.3	3.0	3.0	3.1	3.2	5.0	*5.6*
TRP	MQC	% Accuracy	106.7	98.9	100.7	101.8	113.8	103.0	114.1	101.4
		RSD%	2.6	1.7	2.0	3.4	1.8	1.7	6.3	3.3
	HQC	% Accuracy	100.8	93.6	95.2	96.1	107.6	97.4	107.7	96.0
		RSD%	4.7	6.2	5.4	3.9	5.8	6.1	2.2	*9.2*
	LQC	% Accuracy	111.5	106.5	106.0	105.7	**117.8**	103.8	111.9	90.1
		RSD%	3.0	2.0	2.2	2.9	2.0	2.5	3.2	*9.3*
ETNH2	MQC	% Accuracy	105.0	100.3	99.8	99.6	111.0	97.8	105.4	**84.9**
		RSD%	1.8	1.5	2.3	1.6	1.4	1.3	2.0	*10.1*
	HQC	% Accuracy	99.3	94.9	94.5	94.2	105.0	92.5	99.7	**80.5**
		RSD%	2.4	4.4	5.5	2.6	4.4	3.1	2.2	*13.3*
	LQC	% Accuracy	98.0	96.0	94.3	96.3	108.1	97.7	110.6	99.6
		RSD%	1.9	4.3	2.1	1.9	4.3	4.7	4.3	*6.5*
LYS	MQC	% Accuracy	102.6	100.5	98.7	100.8	113.2	102.2	**115.8**	104.2
		RSD%	2.7	2.0	0.9	1.4	2.1	2.5	5.9	4.3
	HQC	% Accuracy	99.7	97.7	96.0	98.0	110.1	99.4	112.5	101.5
		RSD%	2.6	5.9	3.7	3.3	6.0	6.3	3.4	*8.1*

[1]bolded values indicate accuracy measurements out of the acceptable range ±15%, italic and underlined values indicate the highest precision values among calibrators

[2] n=4

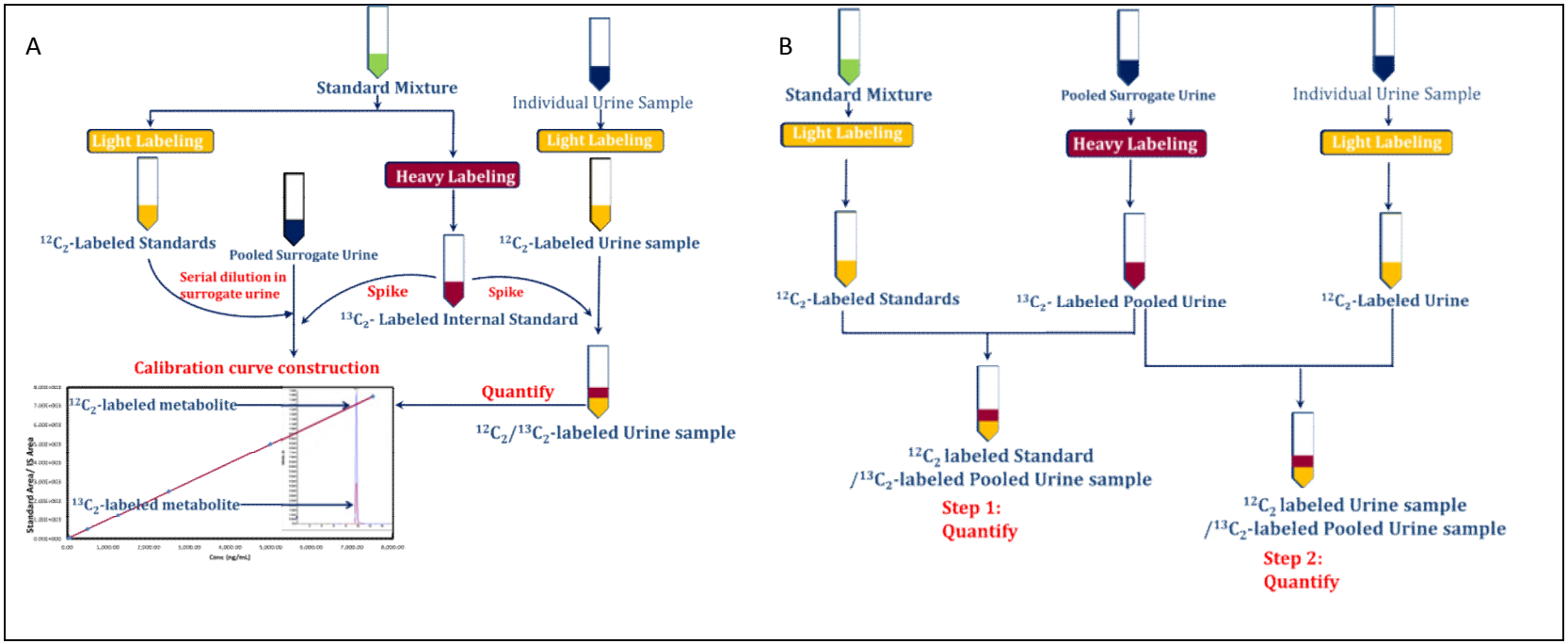

Figure 1. Quantitative platforms in DIL-LC-MS/MS; (A) single and multipoint calibration and (B) relative quantification. Pooled surrogate urine is prepared from 7 pooled urine samples processed as described under 2.3. standards derivatization.

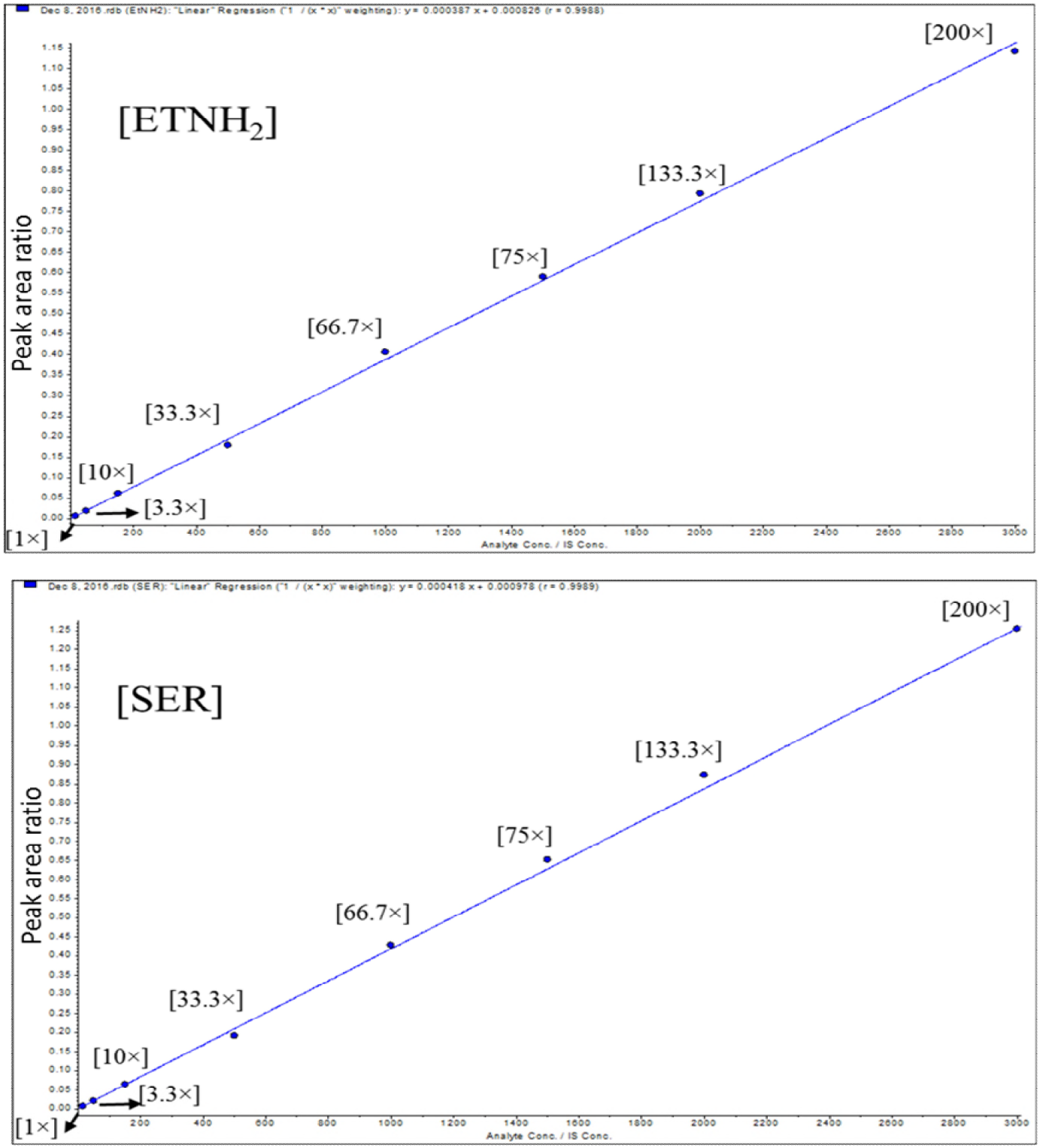

Figure 2: representative calibration curves generated from the multipoint calibration method for the quantification of $ETNH_2$ and SER

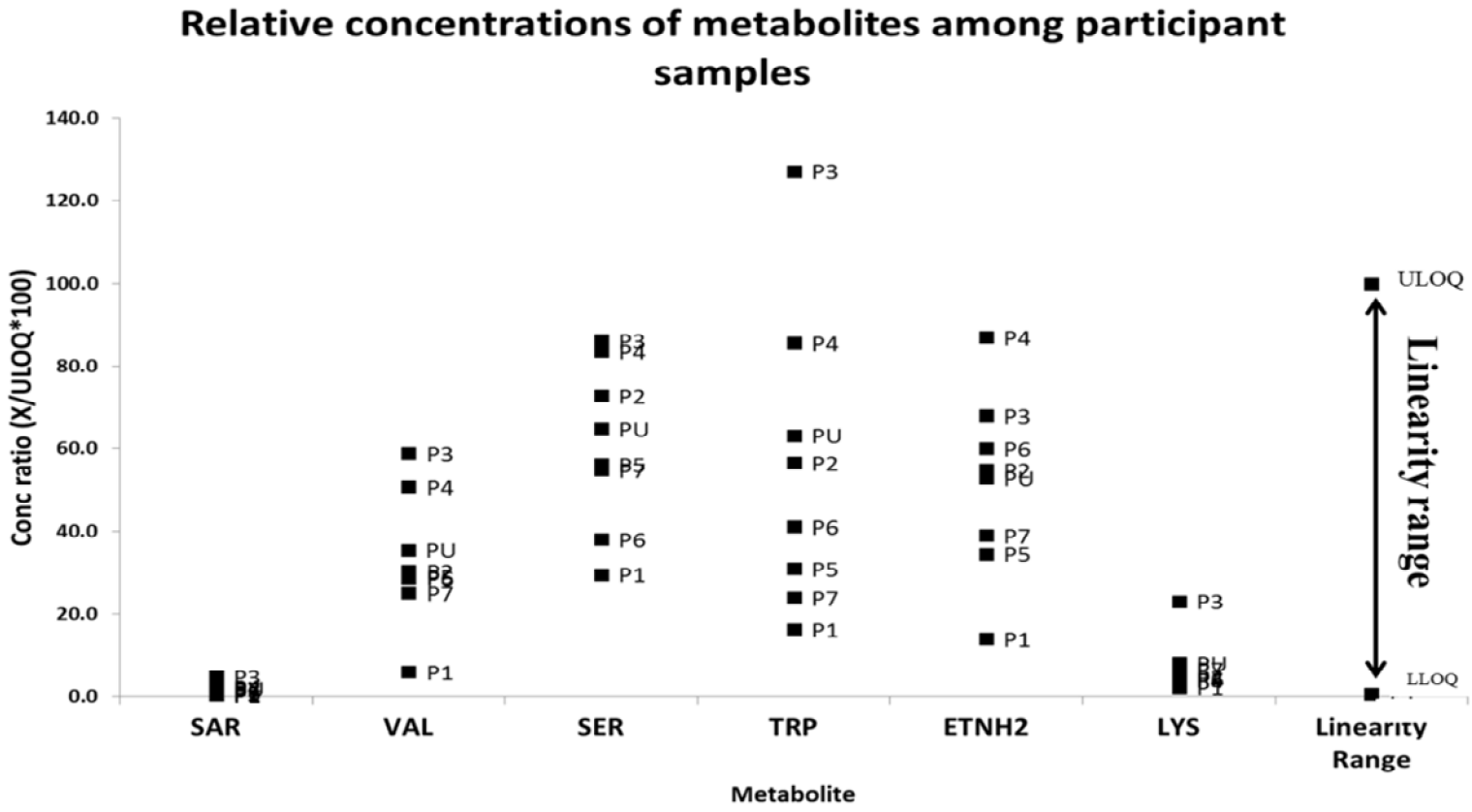

Figure 3: Concentrations of investigated metabolites in patients urine samples, data normalized to each metabolite's ULOQ. Patients are abbreviated as P followed by their ID number, PU is pooled urine sample.

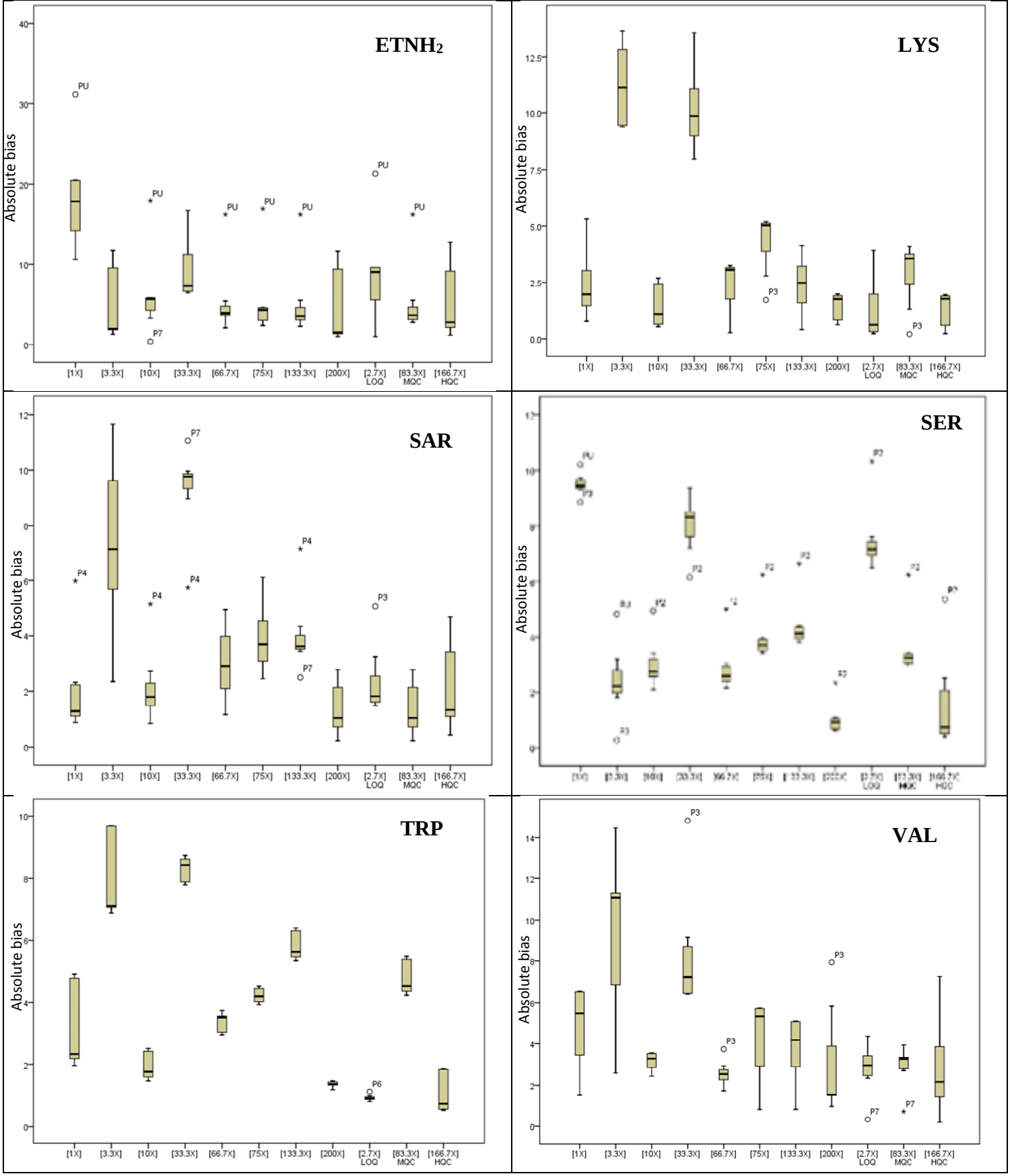
ETNH2
LYS
SAR
SER
TRP
VAL
Absolute bias
[1X]
[3.3X]
[10X]
[33.3X]
[66.7X]
[75X]
[133.3X]
[200X]
[2.7X] LOQ
[83.3X] MQC
[166.7X] HQC

Figure 4: Box and whisker plots of inter-day bias of metabolites measurements as obtained with single point calibration. Bias is defined as % deviation from multipoint calibration calculated using each calibration solution (1×-200×) and QC sample (3 levels) as single point calibrators, *equation 1*, Y axis represent the absolute bias values calculated using 8 urine samples each processed in triplicates, the upper and lower whiskers represent higher and lower bias values, respectively. The box represents the interquartile range and the median is expressed through the vertical line within in the box. Participants are represented by "P", PU is pooled urine

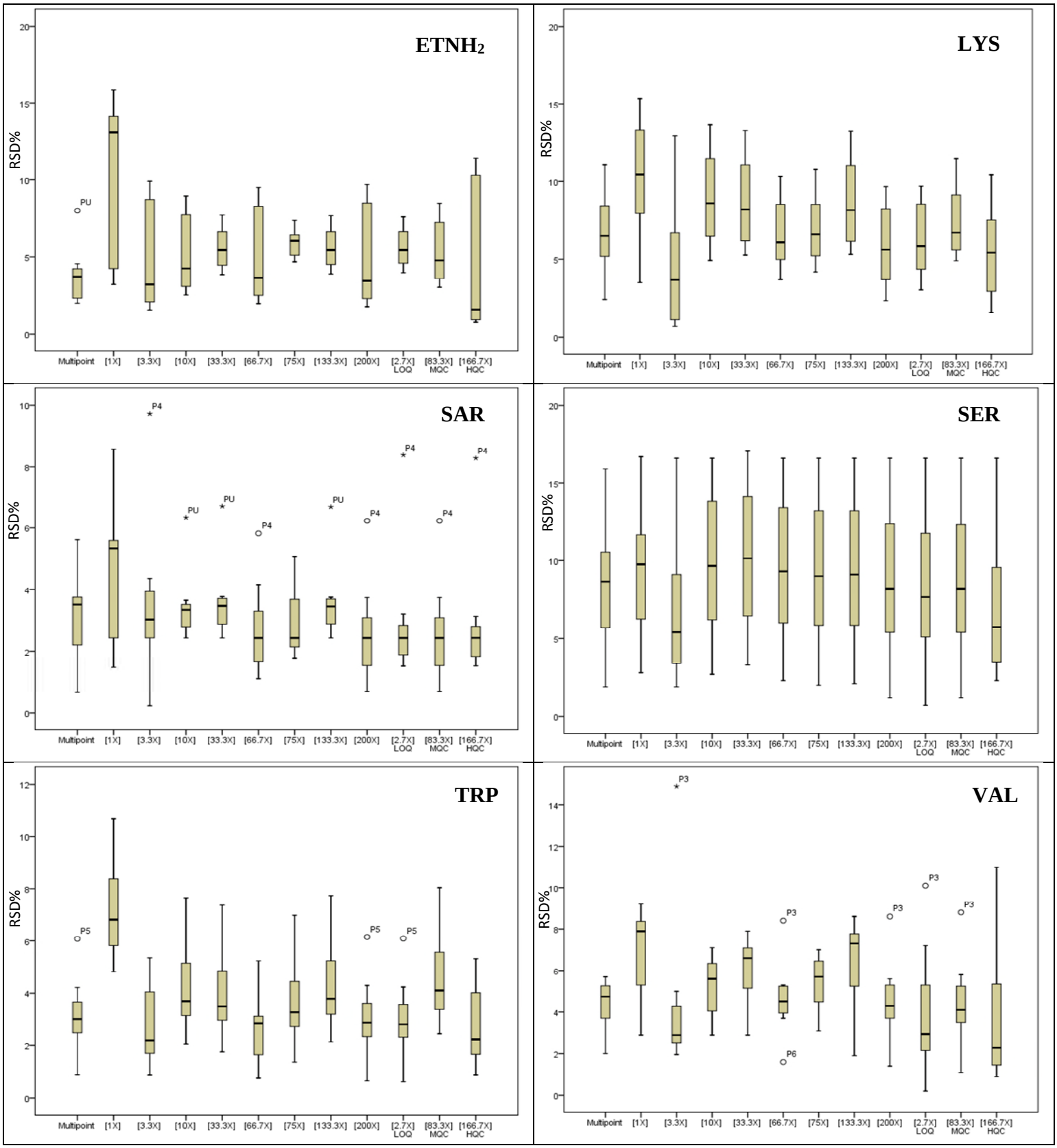

Figure 5: Box and whisker plots of metabolites measurements inter-day precision as obtained with multipoint calibration (using regression equation, *Table 1*) and single point calibration (using each calibration solution [1×] to [200×] and QC solutions (3 levels) as calibrators, *equation 1*). Y axis represent the RSD% ranges obtained from 8 urine samples each processed in triplicates; the upper and lower whiskers represent higher and lower RSD% ranges, respectively. The box represents the interquartile range of RSD% values and the median is expressed through the vertical line within in the box. Participants are represented by "P", while "PU" stands for pooled urine

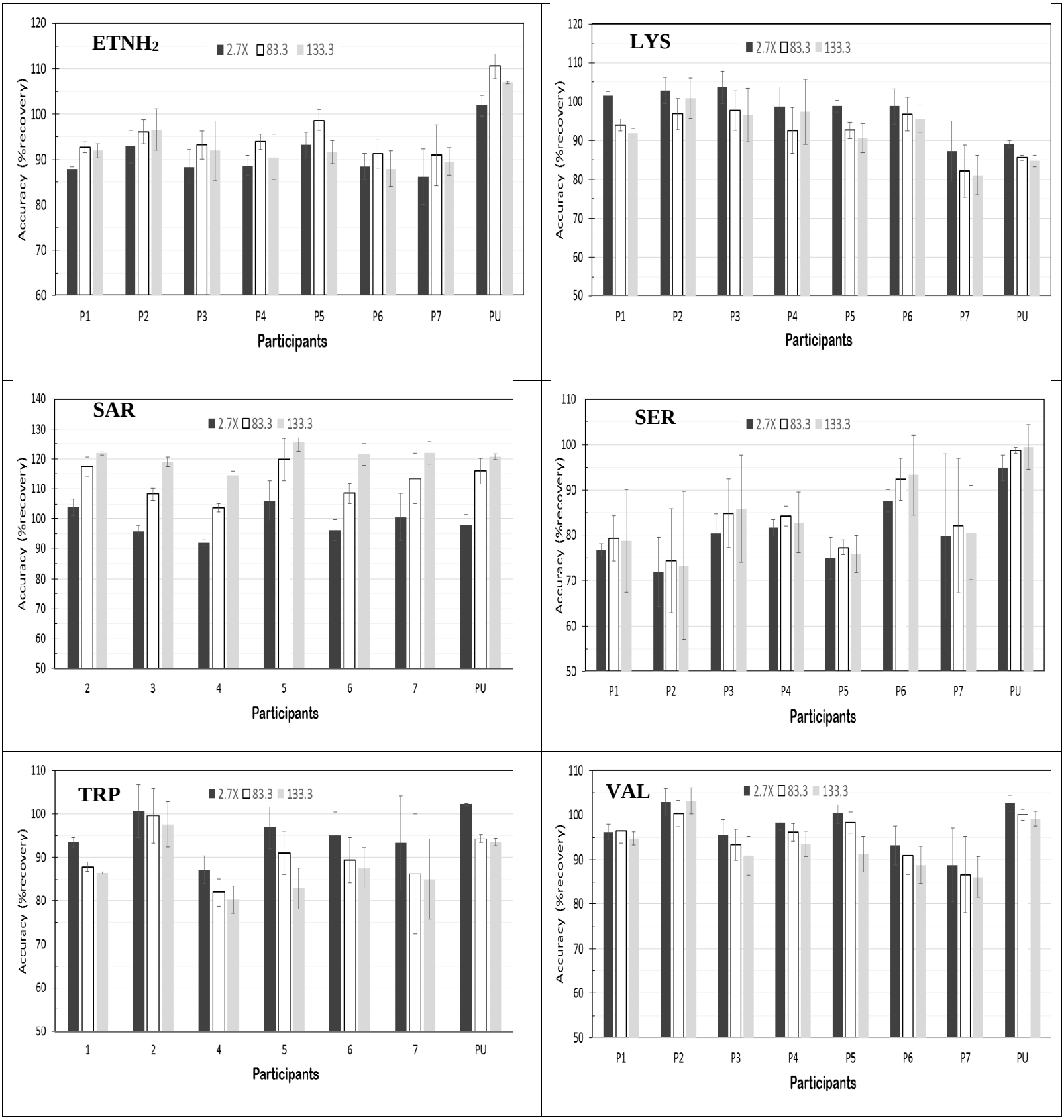

Figure 6: Accuracy and precision of relative quantificaiton, bar hight respresent method accuracy, in comparison to multipoint calibration, errror bars represent RSD%, n=3