Talanta 223 (2021) 121694

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Quantitation of phosphatidylethanol in dried blood after volumetric absorptive microsampling

Katleen Van Uytfanghe [a], Maria del Mar Ramirez Fernandez [b], Aurelie De Vos [a], Sarah MR. Wille [b], Christophe Pol Stove [a,*]

[a] Laboratory of Toxicology, Department of Bioanalysis, Faculty of Pharmaceutical Sciences, Ghent University, Ottergemsesteenweg 460, 9000, Ghent, Belgium
[b] Federal Public Service Justice, National Institute of Criminalistics and Criminology, Chaussée de Vilvorde 100, 1120, Brussels, Belgium

ARTICLE INFO

Keywords:
Phosphatidylethanol
Liquid chromatography tandem mass spectrometry
Volumetric absorptive microsampling
Direct alcohol marker
Method comparison

ABSTRACT

Background: Stimulated by the increased recognition of phosphatidylethanol (PEth) as sensitive direct marker of alcohol intake, the Ghent University's Laboratory of Toxicology and the National Institute of Criminalistics and Criminology combined their efforts to develop a quantitative method. To facilitate implementation the focus was on the use of a sampling technique which allows quick and easy blood collection, without the need of dedicated personnel at any place/any time. In the meantime the cooperation of the two labs should also allow to initiate a Belgian network of laboratories capable of quantifying PEth.

Methods: Dried blood microsamples were collected via volumetric absorptive microsampling (VAMS). PEth 16:0/18:1 was quantified after liquid-liquid extraction using two independent isotope dilution - liquid chromatography – tandem mass spectrometry methods. A systematic review of the entire process at both sites was performed before the final method comparison using samples from 59 routine toxicology cases collected within a one-year time interval.

Results: Initial differences between both laboratories were solved by focusing on important methodological aspects: (i) trueness verification of the calibration protocol focusing on the primary material, preparation of the stock solutions and adequate equilibration of calibrators and QCs, and (ii) verification of comparability of results obtained with different *m/z* transitions. Several of these aspects could only be verified by critically assessing spiked and native samples. After a final validation good average comparability of the two methods was observed. The average bias was −0.4%, with 85% of the differences within 20%. Moreover, the methods proved to be reproducible and robust within a one-year time interval.

Conclusion: This study is the first to develop a quantitative volumetric absorptive microsampling based method for PEth measurements, in addition it is the first to perform a systematic comparison of PEth measurements between two laboratories. From the discussion on the encountered pitfalls it is clear that also on a global scale, more efforts are needed to improve interlaboratory agreement.

1. Introduction

In the quest of finding markers for monitoring long term alcohol consumption/abstinence with a higher specificity and sensitivity than today's standards, the usefulness of phosphatidylethanol is increasingly being recognized [1–4]. Phosphatidylethanol is a group of phospholipids, formed via the action of the enzyme phospholipase D, but only when ethanol is present [5]. The choice of the phosphatidylethanol analogue for the research described in this manuscript was based on the fact that phosphatidylethanol 16:0/18:1 is the most abundant and most commonly measured species [1,2]. In what follows we refer to this particular species as "PEth". PEth has a slow elimination rate with a reported half-life of approximately 4–10 days [6,7]. A single drinking episode resulting in a blood alcohol concentration of 1 g/L was reported to be detectable up to 12 days afterwards [8]. Hence, PEth is a highly selective and sensitive marker for alcohol intake, with a sensitivity of 95% and a specificity of 100% to detect chronic and excessive alcohol consumption *vs*. respectively 77% and 88% for carbohydrate-deficient transferrin [9,10]. PEth has the potential to become a key biomarker for routine screening in different settings and is already used in forensic psychiatry and monitoring programs, and for judging driving ability, the

* Corresponding author.
E-mail address: christophe.stove@ugent.be (C.P. Stove).

https://doi.org/10.1016/j.talanta.2020.121694
Received 24 June 2020; Received in revised form 17 September 2020; Accepted 21 September 2020
Available online 24 September 2020

Abbreviations

dried blood spots (DBS)
external quality control (EQA)
hematocrit (hct)
internal standard (IS)
internal quality control (IQC)
Laboratory for Toxicology at Ghent University (UGent)
liquid chromatography – tandem mass spectrometry (LC-MS/MS)
lower limit of quantitation (LLoQ)
liquid-liquid extraction (LLE)
National Institute of Criminalistics and Criminology (NICC)
phosphatidylethanol (PEth)
relative standard deviation (RSD)
room temperature (RT)
standard deviation (SD)
upper limit of quantitation (ULoQ)
volumetric absorptive microsampling (VAMS)

identification of alcohol intake in specific risk groups and for neonatal screening of prenatal alcohol exposure [11].

With PEth being located at the surface of red blood cells, its measurement requires the collection of whole blood samples [12]. Quick and easy blood collection, without the need of dedicated personnel may further increase the potential use of this biomarker: samples could be drawn at any moment/place. In view of this, fingerprick sampling combined with the collection of dried blood spots (DBS), was shown to be suitable for the quantitative determination of PEth [9,13]. In addition, there is an increased interest in the use of devices to volumetrically collect dried blood microsamples [14,15]. Best known in this context are the Mitra® devices, capable of volumetric absorptive microsampling (VAMS), i.e. the volumetric collection of blood by an absorptive tip. These devices overcome some limitations related to variability in processed sample volume upon taking a sub-punch of a conventional DBS, often referred to as the hematocrit effect [16–18]. Moreover, these devices may also offer an advantage from a sampling point of view, by increasing the user friendliness, as suggested by feedback obtained from VAMS device users [18,19].

The Ghent University's Laboratory of Toxicology (UGent) and the National Institute of Criminalistics and Criminology (NICC) recognized the potential of PEth as a valuable marker for alcohol intake and set up different projects to further evaluate its use. These projects include high numbers of study participants and aim at evaluating the applicability of PEth, the evaluation of our current in-house decision limits (see Fig. 1) and the use of VAMS in different settings (from sampling by study nurses and general practitioners to home sampling). The limits are based on previous work and reports by international peers [2,9,20], taking into account a "grey zone" with respect to the cut-off suggestive for excessive alcohol use. The latter gives the benefit of the doubt to the study subject and compensates for (i) the measurement uncertainty (as is also foreseen in the Belgian legislation for measurements performed in the context of driving under the influence of drugs) and (ii) the inconsistency between current decision limits. Several variables contribute to this measurement uncertainty/inconsistency – part is related to the insufficient comparability of current PEth methods, as can be deduced from the results of the Equalis PEth external quality scheme (product code 295, rounds 2017–2020), in which reported results may differ up to 2-fold, while z-scores may still be 'acceptable' [21,22].

Here, we report on the development and application of PEth 16:0/18:1 measurement procedures in both laboratories, and on how pitfalls encountered in getting the results comparable were solved. Method comparability was validated using 59 individual donor samples.

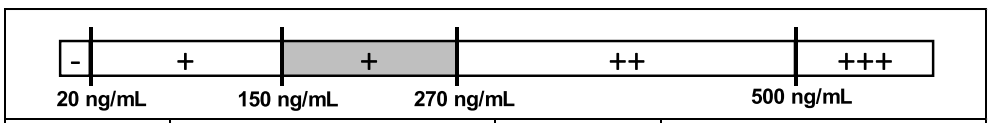
2. Materials and methods

A detailed description of the chemicals, sample preparation and liquid chromatography – tandem mass spectrometry (LC-MS/MS) procedures applied in both laboratories can be found in supplementary data. In short, PEth was extracted from 10-µL VAMS after adding 250 µL of extraction solvent and 60 µL methanol containing 25 ng/mL (0.034 µM) PEth-D5 as internal standard (IS) and shaking (1400 rpm) for 60′ at room temperature (RT). The extraction solvent consisted of 2 mM ammonium acetate 0.01% formic acid in a 2/8/0.2 water/isopropanol/formic acid mixture. Subsequently, this extract was subjected to liquid-liquid extraction (LLE) using 1 mL n-hexane and shaking (1400 rpm) for 10' at RT. The n-hexane fraction was collected, dried and reconstituted in 50 µL of injection solvent, of which 5 µL was used for the measurement with two completely different LC-MS/MS procedures. For quantification, both methods used m/z 701 → 255.

2.1. Preparation of calibrators and QCs

We refer to the supplementary data for a detailed description of the preparation of the calibrators and controls, and the verification experiments performed to establish the final protocol. Pipetting of whole blood is involved. Therefore, to keep precision/accuracy under control, only fresh blood was used and all pipetting steps were done with gravimetric control. From a master stock solution of 8 mg/mL (11 mM) PEth 16:0/18:1, working solutions in whole blank blood were prepared with PEth concentrations of 10, 20, 50, 100, 250, 500, 1000 and 2000 ng/mL (0.014, 0.028, 0.069, 0.138, 0.345, 0.670, 1.38 and 2.76 µM) and internal quality control (IQC) samples with PEth concentrations of 10, 30, 500 and 1500 ng/mL (0.014, 0.041, 1.38 and 2.07 µM). For the final protocol, spiked samples were equilibrated overnight at 4 °C before sampling on VAMS. The lower and upper limit of quantification (LLoQ and ULoQ), at 10 respectively 2000 ng/ml, were defined based on previous work [9].

External Quality Assessment (EQA) samples were obtained from Equalis (Uppsala, Sweden) - rounds 17:03, 18:01 and 18:03 [21]. To avoid possible commutability issues (more specifically, the provided



< 10 ng/mL	compatible with abstinence	150 - 270 ng/mL	social drinker, with more important alcohol use
10 - 20 ng/mL	compatible with abstinance or minor alcohol intake	270 - 500 ng/mL	suggestive for excessive alcohol use
20 - 150 ng/mL	social drinker	> 500 ng/mL	excessive alcohol use

Fig. 1. Decision limits currently used at the UGent Laboratory of Toxicology, based on Kummer et al. [9]. The indicated values, expressed in µM limits, are respectively 0.014 µM, 0.014–0.028 µM, 0.028–0.21 µM, 0.21–0.37 µM, 0.37–0.69 µM and > 0.69 µM.

blood might behave differently from fresh blood when being absorbed by the VAMS tips), EQA samples were diluted 1:5 with whole blank blood and equilibrated overnight (4 °C) before application to VAMS. Four different dilution lots were prepared over time.

2.2. Sample collection

Blood sample collection was approved by the Ethics Committee of Ghent University Hospital (EC UZG 2018/0740 – Registration n° B670201836586). Venous blood was obtained from an alcohol abstinent healthy person in citrate tubes (BD vacutainer®, 2.7 mL). Mitra™ devices, obtained from Neoteryx® (Torrance, CA), were used to generate dried blood samples from the venous whole blood. Samples were prepared by allowing the absorptive tip of the devices to wick up blood, by touching the surface of the blood, thereby taking care to prevent overfilling. After the device was completely filled with blood, contact remained for 2 more seconds before the device was placed in the accompanying plastic clamshells and dried for 2 h at RT. For longer storage the clamshells were put in a zip closure plastic bag containing a 5 g MiniPax® absorbent packet from Sigma Aldrich (Diegem, Belgium) and stored at RT. Multiple series of calibrators and IQC samples were sampled at once and stored until further analysis.

For the method comparison, samples collected from persons in a driving license regranting program or samples from routine toxicology cases were used. Capillary blood was directly obtained through a fingerprick with a BD® Microtainer (contact-activated lancet). The first drop of blood was wiped off, while the second drop was used for sampling. Samples were collected over a one year time interval, with measurements at UGent being performed within a 2–3 weeks' time interval after collection. At the NICC, the samples were measured at the end of the sample collection period. Sample storage was as described above.

2.3. Prevalidation experiments

In order to investigate initial differences between the methods at UGent and NICC, several experiments were performed. These encompassed comparison of results for the same extracts (from calibrators, 4 IQC samples in duplicate, 7 EQA samples in duplicate and 9 native samples), analyzed at both sites, to investigate whether observed differences were due to the LC-MS/MS method itself. Also the UGent and NICC stock solutions were directly compared, by generating 3 independent working solutions, which were further diluted with water to prepare 3 calibrators (2000 ng/mL (2.76 μM)). Ion ratios (area analyte/area IS) of the UGent and NICC calibrators were compared (Details can be found in supplementary data).

2.4. Method validation

The method validation covered selectivity, calibration model and homoscedasticity, trueness, precision, carry-over, matrix effect, extraction efficiency and stability. Besides these parameters, which are based on U.S. Food and Drug Administration and European Medicines Agency guidelines for bioanalytical method validation [23–25], the influence of the hematocrit (hct) on the extraction was also investigated, in line with the IATDMCT guideline on validation of DBS-based procedures [26]. The LLoQ was not determined but was pre-fixed based on the cut-off value of alcohol abstinence described in literature (see Fig. 1). Details can be found in supplementary data.

2.5. Method comparison study

A set of 59 individual donor samples was measured in both laboratories. Comparability was evaluated using the criteria generally applied for incurred sample reanalysis, i.e. at least two-thirds of the results should not deviate more than 20% of their mean [24,25]. Furthermore, results were evaluated using Passing Bablok regression and a Bland-Altman plot. For the latter, the variation in the %differences (expressed as 1.96 times the standard deviation (SD) of the % differences), should not exceed 29.9% (total error: observed average absolute bias + 1.96 * observed average total precision; see Table 3 for the numerical values). To exclude that the difference in time between the initial measurement at UGent, and the incurred sample analysis at NICC would have an influence on the results, the % deviation was also evaluated against the difference in time.

2.6. Measurement uncertainty

Measurement uncertainty was calculated following the "Handbook for calculation of measurement uncertainty in environmental laboratories" and was based on the total precision for the method and a bias component derived from the results within the EQA scheme (taking into account the bias and its standard deviation) [27,28]. The acceptance limit of 42.4% for the measurement uncertainty was propagated based on the individual limits for precision and bias (both 15%).

2.7. Data analysis

Data analysis was done using Excel and Medcalc statistical software, version 14.12.0 (Ostend, Belgium). Statistical tests were performed 2-sided, with 95% confidence – unless otherwise stated.

3. Results

3.1. Preparation of calibrators and internal quality control samples

As scouting experiments had revealed that discrepant results for authentic VAMS samples were obtained, depending on whether VAMS calibrators were prepared from freshly spiked blood vs. blood that was allowed to equilibrate with spiked PEth for some time, we systematically evaluated this variable. Results obtained with liquid whole blood calibrators analyzed immediately after spiking whole blood, or after storage for 72 h at 4 °C were compared to results obtained with VAMS-based calibrators, generated immediately after spiking whole blood, or after first equilibrating the blood at 4 °C for 1, 8, 24 or 72 h. These revealed no significant differences (95% confidence) for blood equilibrated for 1, 8 and 72 h. For the final protocol the most time efficient option (i.e. overnight equilibration) was adopted. The supplementary data contain details for both experiments. An impact of incubation time on the extractability was also observed by others [29,30].

3.2. Method validation

The methods proved to be selective, as no unacceptable interfering peaks were detected in the blank samples (i.e., the blank blood used to prepare the calibrators, 4 different donors during the course of the validation). For calibration, at UGent, quadratic curves (weighing $1/x^2$) showed the best fit, while at the NICC a linear fit (weighing 1/x) was best. The mean back-calculated PEth concentrations were within ±15% difference of the nominal values (20% for LLOQ), except for 2 values observed in eight calibrations curves at the NICC.

At UGent, from the signal:noise of the LLoQ (18.4 ± 4.7 (SD), n = 10*2), the LoD was estimated to be 1.7 ng/mL (0.002 μM). At the NICC, the LoD was not determined, as (quantitative) results are only reported from the LLoQ concentration level onwards.

There was no signal detected in blanks injected after samples with a concentration at the ULoQ.

The non IS-compensated matrix effect was 79% and 114% at UGent and NICC, respectively indicating some suppression and enhancement of ionization (Table 1). In both labs, an improvement was seen for the IS-compensated matrix effects, at 98% and 109% for UGent and NICC, respectively. Importantly, inclusion of the IS also led to a substantial improvement of the %RSD values, all being below 8%, which is well

Table 1
Non-IS compensated matrix effects and IS-compensated matrix for QC low and QC high.

	UGent		NICC	
	Low (%)	High (%)	Low (%)	High (%)
%ME (%RSD)	79.6 (21)	77.8 (16)	119.5 (24)	107.8 (14)
%MEIS (%RSD)	100 (7.9)	95.9 (5.2)	110.7 (5)	108 (6)

within the acceptance limit of 15%.

Recovery of the analyte over the entire procedure ranged from 44 to 64% and was consistent within each condition (%RSD < 15%) (Table 2). For the low QC, a slight hct dependence was observed, as the IS-compensated recovery for the low hct was not within 15% of the recovery for samples with a normal hct. For the subjects under study in the envisaged application, hct is expected to be within the normal range (36–50%) [31]. Hence, in this case there will be no impact. The efficiency of the LLE procedure was on average 75.2% (varying from 71 to 81%), with no noticeable concentration- or hct-dependence (Table 2).

PEth was stable in extracts after 72 h of storage in the autosampler at 4 °C (differences ≤ 5.1%; Supplementary Table S2). In VAMS samples, stability was demonstrated for at least one week at the three different evaluated temperatures (4°C, RT, and 45°C) and for one month at RT, as deviations did not significantly exceed ±15% (Supplementary Figure S3). Furthermore, results for the 4 EQA samples revealed that VAMS samples can be stored for at least 400 days at RT, as longitudinal analysis of EQA results revealed no discernible trend over time (slope not significantly different from 0 (P<0.0001)), the vast majority of the PEth results (87%) laying within ±15% of the normalized mean (Fig. 2). Note that, taking into account the mean total precision of the method (11%) and the bias criterion (15%) – a mean result of 2 duplicates on a single day may deviate 35% (= maximum allowable bias + z (2.58) * precision/square root of n). None of the results is outside this range.

Data for bias and precision are summarized in Table 3. Acceptance criteria were met at every concentration level, with a single exception (total precision at LLoQ for the NICC, at 20.3% narrowly exceeding the acceptance criterion of 20%). The expanded measurement uncertainty for results obtained by either of the two laboratories, calculated based on these results, was 38% (coverage factor k=2), which is below the preset specification of 42.4%.

Fig. 3 shows the results of the method comparison in a scatter plot with Passing Bablok regression analysis (A) and in a Bland-Altman Plot (B). Both show the good average comparability of the two methods, as the slope and intercept are not significantly different from 1 and zero, respectively, both having relatively narrow confidence intervals in the Passing Bablok curve. On average there was essentially no bias (-0.4%) between both methods, the bias at the decision points being 0.9% (at 20 ng/mL (0.028 μM)) and 3.6% (at 270 ng/mL (0.37 μM)). Results were above LoD for 54 samples at Ugent (> 1.7 ng/mL (0.002 μM)) and above LLoQ for 51 samples at NICC (LLoQ > 10 ng/mL (0.014 μM)). The preset acceptance limits (29.9%) for the variation of the %differences (see section 2.5) aligned reasonably well with the limits of agreement (expressed as 1.96*SD of the %differences) in the Bland-Altman plot, suggesting the fit-for-purposeness. This was also underscored by the fact that only 15% of the samples had a deviation outside the 20% limit, the highest deviations being observed for samples with a concentration close to the decision limit (20 ng/mL (0.028 μM)). In our routine practice, most samples analyzed in the context of driving license regranting have (far) higher concentrations (median 181 ng/ml; n = 716). Furthermore, Fig. 3C, which gives an overview of the samples scored between 9 and 400 ng/mL (0.013–0.55 μM), shows that, overall, taking into account the measurement uncertainty and the decision limits mentioned in Fig. 1, a consistent scoring would be obtained by both labs. Last, no influence of the difference in time between both measurements was observed (Fig. 3D).

Table 2
Recovery and LLE efficiency (mean and %RSD; n = 6).

	%Recovery (%RSD)		
	Low hct	Normal hct	High hct
Low QC	64.2 (12)	47.6 (14)	43.9 (14)
High QC	62.5 (14)	64.1 (14)	61.3 (12)
	%IS compensated recovery (%RSD)		
	Low hct	Normal hct	High hct
Low QC	61.0 (10)	46.6 (9.5)	42.5 (13)
High QC	54.5 (15)	54.0 (8.1)	58.6 (10)
	%LLE efficiency (%RSD)		
	Low hct	Normal hct	High hct
Low QC	71.1 (12)	70.6 (11)	79.4 (14)
High QC	70.8 (14)	78.1 (12)	81.1 (6)

4. Discussion

Stimulated by the increased recognition of PEth as a useful marker for alcohol consumption and as a follow-up to our previous research, UGent and NICC decided to collaborate, both implementing their own method, to initiate a Belgian network of laboratories capable of quantitatively determining PEth [1–4,9,18]. Broad applicability of the method is guaranteed by utilizing as a starting point a simplified blood collection procedure, based on a finger prick. Because VAMS allows a fixed volume sampling of blood and because of previous positive experience, this approach was chosen to collect dried blood microsamples [18]. Besides the ease of sample collection, transport and storage, dried blood microsampling approaches hold other advantages as well, which are particularly relevant for PEth: PEth is very stable in dried blood (in contrast to liquid blood) and *ex vivo* formation of PEth, because of the presence of alcohol in blood, can be avoided [32]. This may be an issue when blood is not properly stored, giving rise to falsely elevated results [22].

Both laboratories used the same extraction procedure allowing to attain an LLOQ of 10 ng/mL in a robust manner from merely 10 μL of blood. The actual measurements were done using two completely different LC-MS/MS methods (different LCs, columns and MS instrumentation). Initial trueness experiments using EQA and native samples revealed a 30% difference between the two laboratories, in contrast to comparable results for spiked samples. At first, this was attributed to the equilibration of spiked calibrators before their application on VAMS. The equilibration experiments suggest that spiked PEth should be given time to become incorporated into the membrane of the red blood cells, as in native samples. If not, the extraction efficiency of PEth from the VAMS is higher for incompletely equilibrated spiked samples compared to native samples. This could lead to an underestimation of PEth concentrations in real samples, as well as a potential bias between labs. This hypothesis is supported by the findings of others [29,30] and by the fact that for blood samples (not dried to VAMS), the results obtained for native samples are comparable, whether or not PEth spiked to calibrator samples was allowed to equilibrate with the red blood cells. Finally, calibrators were equilibrated overnight at 4°C before application to VAMS. Confidence in the applied protocol was found in the fact that, over one year and 4 different batches of calibrators, results for EQA samples were stable (see also Fig. 2).

Allowing the calibrators to equilibrate did not result in comparable data. A further search for the root cause of the observed differences pointed at the use of different quantifiers in both laboratories: the transition 701 → 255 (UGent) *vs.* the transition 701 → 281 (NICC). When quantified by the transition 701 → 281, results for native samples (but not spiked samples, as in the QC's) are a factor 1.4 (SD 0.09, n=522) higher. This remarkable difference is documented by Luginbühl et al. and attributable to (i) a different ratio of the PEth 16:0/18:1 to 18:1/16:0 isomers in the reference material *versus* the naturally occurring isomers (16:0/18:1 being more abundant in authentic samples) [33,34],

Table 3
Data of bias and precision at the two application sites.

	UGent									NICC			
	LLoQ	L	M	H	EQA1	EQA2	EAQ3	EQA4	Native sample	LLoQ	L	M	H
Concentration (ng/mL)	10	30	500	1500	735	244	465	146	15.7	10	30	50	1500
Concentration (μM)	0.014	0.041	1.38	2.07	1.01	0.337	0.641	0.201	0.022	0.014	0.041	1.38	2.07
Number of duplicates	12	15	14	14	36	28	29	33	31	8	8	8	8
% Bias[a]	5.2	−4.5	−5.5	−6.5						4.4	−10.3	−8.9	8.5
z-score in EQA					−1.6	−2.2	−1.7	−1.0					
Within-run precision (%)	7.8	6.2	7.9	5.1	6.9	7.9	8.2	8.6	9.6	15.4	6.2	9.8	7.1
Between-run precision (%)	10.9	4.8	7.8	10.1	4.8	7.5	7.1	10.8	6.6	13.3	7.0	4.6	5.0
Total precision (%)[a]	13.4	7.9	11.1	11.3	8.4	10.9	10.8	13.8	11.7	20.3	9.3	10.8	8.7

[a] Results used to calculate the acceptance limits for the 1.96SD of the differences in the method comparison.

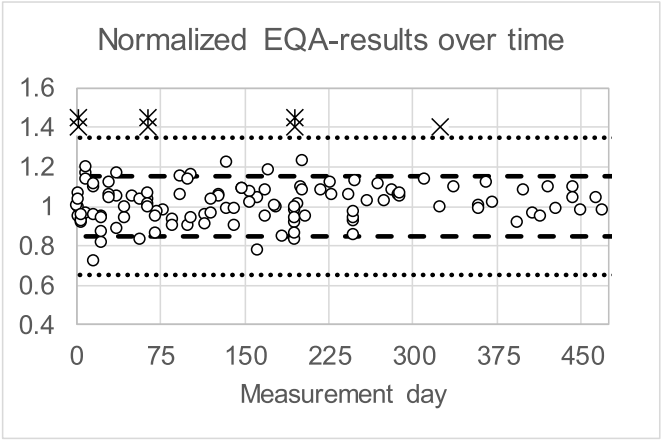


Fig. 2. Overview of EQA results over time. For each EQA sample the results are normalized to their mean. The dashed line represents the ±15% overall bias limits, the dotted line shows the acceptance limit for a daily mean of duplicate measurements (35%). (x) and (*) mark the day where respectively a new calibrator lot or new QC lot was started.

and (ii) a different fragmentation efficiency of the side chains of the two isomers (the sn-2 position being the preferred fragmentation site (*m/z* 281 for PEth 16:0/18:1)) [33,34]. As in high-throughput methods, both isomers typically aren't chromatographically separated, native samples might have an ion ratio that is distinct from that observed in calibrators, depending on the calibrator chosen [33]. This is an issue that has been largely neglected in the literature and can only be solved by reference material manufacturers. When both labs used the transition 701 → 255, analysis of the same extracts (for both calibrators and samples) resulted in an average difference of 1% between both labs, with a SD on the differences of 10%.

Yet, the implementation of this second measure still didn't result in comparable data for samples that were analyzed using independently prepared calibration curves. Hence, we directly compared the different lab's stock solutions to exclude that these would result in a calibration bias. Three independent working solutions prepared from the stock solutions of the two labs were compared. Observed isotope ratios (%RSD) for the UGent solutions were 16.9 (2.4%) vs. 10.0 (3.1%) for the NICC solutions, indicating that, although both laboratories used the same reference standards, the stock solutions deviated. Rather than searching for the root cause of this difference, new and independent stock solutions were made, using the most accurate procedure possible. Using two new vials of reference standard (one in each lab), meticulously following the protocol described in supplementary data, resolved the issue. This protocol involved weighing the original vial before and after transfer of the PEth. The rationale behind this was that it is common practice for manufacturers to overfill the vials so that the customer receives at least the amount of analyte ordered [35]. Hence, taking the nominal weight instead of the exact amount could possibly lead to a calibration bias. After applying this protocol, the results obtained by both laboratories were finally comparable, and the actual method validation and formal method comparison could start.

In both laboratories, the method for quantification of PEth in VAMS samples was fully validated and the results showed that, overall, the preset quality specifications were met for all investigated aspects. Moreover, the obtained LLoQ, recovery and trueness are comparable to most recent published PEth 16:0/18:1 methods. Although precision seems to be a little higher, it should be noted that most published methods started from larger starting volumes of whole blood. Here, the starting volume was merely 10 μL of blood, in the format of a dried VAMS sample. While we and others have previously demonstrated that conventional DBS can be used for PEth analysis, the method reported here is the first to quantify PEth, starting from 10 μL VAMS samples. [29, 36–39]. In a real-world setting, with non-supervised sampling by non-trained individuals, the %RSD after application of the entire procedure (from sampling to analysis) is somewhat higher (14%) than that observed here for lab-generated samples (11%), but still acceptable, as we reported elsewhere [19].

At UGent, the method has been implemented on a weekly basis for over a year. During this period, multiple sets of calibration curves and EQA lots were prepared and measured. This allowed us to validate the method's robustness as well as the stability of the dried VAMS samples. No lot-to-lot variation was observed when evaluating the EQA results, as both the precision and trueness proved to be stable over time. Combining the data obtained from the evaluation of different calibrator lots and EQA lots, made at different time points, also proved the long-term stability of PEth in the VAMS devices (Fig. 2). At the same time, this also demonstrated the consistent extractability over time, an important parameter when dealing with VAMS [40,41]. The excellent stability and consistent extractability could also be derived from the results of the method comparison, which can also serve as an incurred sample reanalysis experiment, with a time difference between the two measurements ranging from 30 days up to almost 400 days. In addition, the results of the method comparison demonstrated the good comparability of the results, independently obtained by both laboratories, with an average difference of -0.4%, and 85% of the samples having a difference within ±20%, thereby fulfilling the requirement of incurred sample reanalysis (i.e. 2/3 of the samples should be within 20% of the average).

The efforts that were required to achieve comparable results between only two laboratories made us realize that one should be very careful when considering and interpreting published PEth results. Some of the observations that were described above would never have been made when only using spiked QC samples, as these will always behave in exact the same way as spiked calibrators. This potential issue of non-commutability between on the one hand spiked calibrators/QCs and on the other hand native samples points to the importance of including native samples from the initial validation experiments on. In addition, as also pointed out by the recent publication of Luginbühl *et al.*, the choice

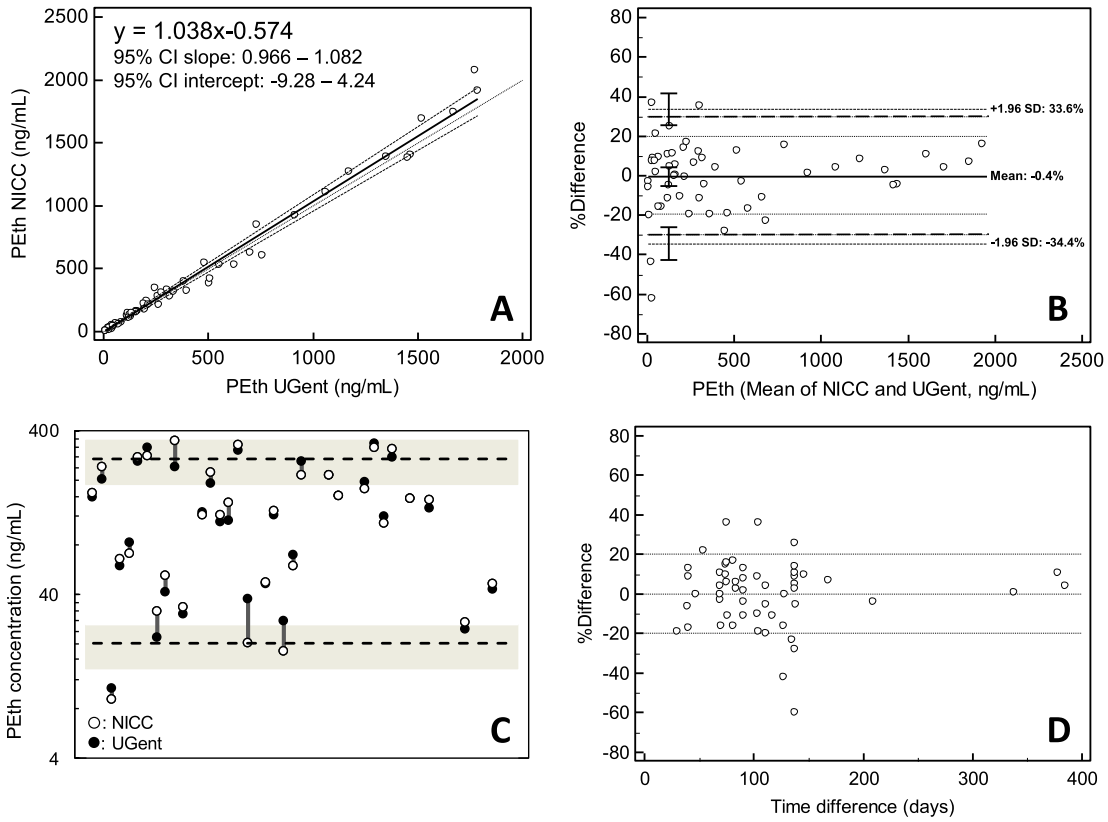


Fig. 3. Comparison of PEth data of 59 authentic samples, measured at UGent and NICC. A, Passing Bablok regression plot: the black line is the regression line, the dotted line is the line of equality and the dashed lines represent the 95% confidence interval around the regression line. B, Bland Altman plot: the mean % difference is represented by the black line, surrounded by the 95% confidence interval. The limits of agreement (1.96SD of the differences) are indicated by the dashed lines (also surrounded by their respective 95% confidence intervals). The dotted lines represent either the ±20% limits and the acceptance limits for 1.96SD of the % differences (at ±29.9%). C, % difference plot (sorted per time difference): the dotted lines represent zero or the ±20% limits. D, PEth levels for each method. Both results are connected with a line to visualize the effect the use of a different method had on the interpretation of the result. For clarity, the graph only represents all data points between 9 and 400 ng/mL (0.013–0.55 μM). Cut-off levels are indicated by dashed lines. The grey rectangles represent the area corresponding to cut-off ± 29.9% total error.

of the reference standard and the quantifier ion may have an important consequence on the numerical value that will be reported [33]. Given this discrepancy, and given the judicial framework in which we (like many labs that measure PEth) operate, we opted to use the transition 701 → 255 for quantitation purposes. By doing so, we give the benefit of the doubt to the people under investigation. However, the fact that different labs may opt to use different transitions, yielding different results, also demonstrates the urgent need to standardize PEth quantification in order to allow more reliable inter-method comparisons and to justify the use of common cut-off values. This is yet another reason to interpret the results of EQA schemes with utmost care, as we have no reference as to which calibrator and which quantifier was used by which participant.

5. Conclusion

This study is the first to implement PEth quantification based on dried blood samples after volumetric microsampling. A technique which is highly suitable in the main areas of interest for PEth determination, where sampling by untrained professionals, regardless place and time is of benefit. Moreover, it is the first to perform a systematic method comparison of PEth analysis between 2 different labs. The winding road to achieve comparable results points out the important methodological aspects that need to be tackled: from trueness verification of the calibration protocol, starting with the primary material and the preparation of the stock solutions, over adequate equilibration of the calibrators and QCs with spiked PEth, and verification of the comparability of results obtained with different *m/z* transitions. Several of these phenomena can only be verified by critically assessing spiked ánd native samples. Up to now, to the best of our knowledge, only three other groups briefly mention one of the observed phenomena [29,30,34]. The final method comparison in this report underpins the suitability of both labs' methods for the intended use. The robustness of the methods and the stability of the samples allow to conduct large-scale epidemiological studies, with comparable results regardless the time point of sample collection, the time point of measurement and even the laboratory.

Given the worldwide increased interest to use PEth as a primary marker for the follow-up of (abstinence from) alcohol use, it is essential that also on a global scale more method comparability is achieved. This will require a concerted effort, including an increased comparability between primary reference materials, and a consensus on which analyte should be measured: PEth 16:0/18:1 – with or without separation of PEth 18:1/16:0 - or other PEth-analogues, or even the sum of different PEth analogues. This will better allow future research to reach a consensus on decision limits, discern reliable PEth half-life, etc. In addition, further improvement of the reliability of PEth analysis will improve the consistency of the values obtained by different laboratories (and decrease their uncertainties) and will further strengthen the implementation of its use in a variety of contexts.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authorship contribution statement

Katleen Van Uytfanghe: Methodology – Validation - Formal analysis - Writing. Maria del Mar Ramirez Fernandez: Methodology – Validation - Formal analysis - Writing Aurelie De Vos: Validation – Formal analysis. Sarah MR Wille: Methodology – Resources – Supervision - Writing. Christophe Stove: Methodology – Resources – Supervision - Writing. All authors have approved the final article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.talanta.2020.121694.

References

[1] H. Andresen-Streichert, A. Muller, A. Glahn, G. Skopp, M. Sterneck, Alcohol biomarkers in clinical and forensic contexts, Dtsch Arztebl Int 115 (18) (2018) 309–315.

[2] W. Ulwelling, K. Smith, The PEth blood test in the security environment: what it is; why it is important; and interpretative guidelines, J. Forensic Sci. 63 (6) (2018) 1634–1640.

[3] V.L. Nguyen, P.S. Haber, D. Seth, Applications and challenges for the use of phosphatidylethanol testing in liver disease patients (mini review), Alcohol Clin. Exp. Res. 42 (2) (2018) 238–243.

[4] R.M. Nanau, M.G. Neuman, Biomolecules and biomarkers used in diagnosis of alcohol drinking and in monitoring therapeutic interventions, Biomolecules 5 (3) (2015) 1339–1385.

[5] C. Alling, L. Gustavsson, E. Anggard, An abnormal phospholipid in rat organs after ethanol treatment, FEBS Lett. 152 (1) (1983) 24–28.

[6] M.L. Hannuksela, M.K. Liisanantti, A.E. Nissinen, M.J. Savolainen, Biochemical markers of alcoholism, Clin. Chem. Lab. Med. 45 (8) (2007) 953–961.

[7] W. Weinmann, A. Schrock, F.M. Wurst, Commentary on the Paper of Walther L. et al.: phosphatidylethanol is Superior to CDT and GGT as an Alcohol Marker and Is a Reliable Estimate of Alcohol Consumption Level, Alcohol Clin. Exp. Res. 40 (2) (2016) 260–262.

[8] A. Schrock, A. Thierauf-Emberger, S. Schurch, W. Weinmann, Phosphatidylethanol (PEth) detected in blood for 3 to 12 days after single consumption of alcohol-a drinking study with 16 volunteers, Int. J. Leg. Med. 131 (1) (2017) 153–160.

[9] N. Kummer, A.S. Ingels, S.M. Wille, C. Hanak, P. Verbanck, W.E. Lambert, N. Samyn, C.P. Stove, Quantification of phosphatidylethanol 16:0/18:1, 18:1/18:1, and 16:0/16:0 in venous blood and venous and capillary dried blood spots from patients in alcohol withdrawal and control volunteers, Anal. Bioanal. Chem. 408 (3) (2016) 825–838.

[10] S. Hartmann, S. Aradottir, M. Graf, G. Wiesbeck, O. Lesch, K. Ramskogler, M. Wolfersdorf, C. Alling, F.M. Wurst, Phosphatidylethanol as a sensitive and specific biomarker: comparison with gamma-glutamyl transpeptidase, mean corpuscular volume and carbohydrate-deficient transferrin, Addiction Biol. 12 (1) (2007) 81–84.

[11] F.M. Wurst, N. Thon, M. Yegles, A. Schruck, U.W. Preuss, W. Weinmann, Ethanol metabolites: their role in the assessment of alcohol intake, Alcohol Clin. Exp. Res. 39 (11) (2015) 2060–2072.

[12] C. Heier, H. Xie, R. Zimmermann, Nonoxidative ethanol metabolism in humans-from biomarkers to bioactive lipids, IUBMB Life 68 (12) (2016) 916–923.

[13] O. Beck, N. Kenan Moden, S. Seferaj, G. Lenk, A. Helander, Study of measurement of the alcohol biomarker phosphatidylethanol (PEth) in dried blood spot (DBS) samples and application of a volumetric DBS device, Clin. Chim. Acta 479 (2018) 38–42.

[14] S. Velghe, L. Delahaye, C.P. Stove, Is the hematocrit still an issue in quantitative dried blood spot analysis? J. Pharmaceut. Biomed. Anal. 163 (2019) 188–196.

[15] L. Delahaye, H. Veenhof, B.C.P. Koch, J.C. Alffenaar, R. Linden, C. Stove, Alternative sampling devices to collect dried blood microsamples: state-of-the-art, Ther Drug Monit submitted, 2020.

[16] P.M. De Kesel, W.E. Lambert, C.P. Stove, Does volumetric absorptive microsampling eliminate the hematocrit bias for caffeine and paraxanthine in dried blood samples? A comparative study, Anal. Chim. Acta 881 (2015) 65–73.

[17] P. Denniff, N. Spooner, Volumetric absorptive microsampling: a dried sample collection technique for quantitative bioanalysis, Anal. Chem. 86 (16) (2014) 8489–8495.

[18] N. Verougstraete, B. Lapauw, S. Van Aken, J. Delanghe, C. Stove, V. Stove, Volumetric absorptive microsampling at home as an alternative tool for the monitoring of HbA1c in diabetes patients, Clin. Chem. Lab. Med. 55 (3) (2017) 462–469.

[19] K. Van Uytfanghe, L. Heughebaert, C. Stove, Self-sampling at home using volumetric absorptive microsampling: coupling analytical evaluation to volunteers' perception in the context of a large scale study, Clin Chem Lab Med (2020). In press.

[20] A. Helander, T. Hansson, National harmonization of the alcohol biomarker PEth, Lakartidningen 110 (39–40) (2013) 1747–1748.

[21] Equalis, Phosphatidyl Ethanol (295). https://www.equalis.se/en/products-services/eqa/phosphatidyl-ethanol/. (Accessed September 2020).

[22] W.L. Isaksson A, T. Hansson, A. Andersson, J. Stenton, A. Blomgren, High-throughput LC-MS/MS method for determination of the alcohol use biomarker phosphatidylethanol in clinical samples by use of a simple automated extraction procedure—preanalytical and analytical conditions, The Journal of Applied Laboratory Medicine 2 (6) (2019) 880–892.

[23] F.D.A. US Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Veterinary Medicine, Guidance for industry-Bioanalytical method validation, 2001.

[24] E.M.A.E. Committee, For Medicinal Products for Human Use (CHMP), Guideline on Validation of Bioanalytical Methods, 2009.

[25] S.M.R. Wille, F.T. Peters, V. Di Fazio, N. Samyn, Practical aspects concerning validation and quality control for forensic and clinical bioanalytical quantitative methods, Accred Qual. Assur. 16 (6) (2011) 279–292.

[26] S. Capiau, H. Veenhof, R.A. Koster, Y. Bergqvist, M. Boettcher, O. Halmingh, B. G. Keevil, B.C.P. Koch, R. Linden, C. Pistos, L.M. Stolk, D.J. Touw, C.P. Stove, J. C. Alffenaar, Official international association for therapeutic drug monitoring and clinical toxicology guideline: development and validation of dried blood spot-based methods for therapeutic drug monitoring, Ther. Drug Monit. 41 (4) (2019) 409–430.

[27] Nordtest, Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories, 2003. http://www.nordtest.info/index.php/technical-reports/item/handbook-for-calculation-of-measurement-uncertainty-in-environmental-laboratories-nt-tr-537-edition-3.html. (Accessed September 2020).

[28] G.f.T. uFC, GTFCH, Richtlinie der GTFCh zur Qualitätssicherung bei forensisch-toxicologischen Untersuchungen, 2009. https://www.gtfch.org/cms/images/stories/files/GTFCh_Richtlinie_For-Tox_Version-2.pdf. (Accessed September 2020).

[29] V.L. Nguyen, P. Paull, P.S. Haber, K. Chitty, D. Seth, Evaluation of a novel method for the analysis of alcohol biomarkers: ethyl glucuronide, ethyl sulfate and phosphatidylethanol, Alcohol 67 (2018) 7–13.

[30] M. Luginbuhl, S. Gaugler, W. Weinmann, Fully automated determination of phosphatidylethanol 16:0/18:1 and 16:0/18:2 in dried blood spots, J. Anal. Toxicol. 43 (6) (2019) 489–496.

[31] P.M. De Kesel, N. Sadones, S. Capiau, W.E. Lambert, C.P. Stove, Hemato-critical issues in quantitative analysis of dried blood spots: challenges and solutions, Bioanalysis 5 (16) (2013) 2023–2041.

[32] J. Jones, M. Jones, C. Plate, D. Lewis, The detection of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanol in human dried blood spots, Anal Methods-Uk 3 (5) (2011) 1101–1106.

[33] M. Luginbuhl, R.S.E. Young, F. Stoth, W. Weinmann, S.J. Blanksby, S. Gaugler, Variation in the relative isomer abundance of synthetic and biologically derived phosphatidylethanols and its consequences for reliable quantification, J. Anal. Toxicol. (2020), https://doi.org/10.1093/jat/bkaa034.

[34] S. Ullah, A. Helander, O. Beck, Identification and quantitation of phosphatidylethanols in oral fluid by liquid chromatography-tandem mass spectrometry, Clin. Chem. Lab. Med. 55 (9) (2017) 1332–1339.

[35] Merck, Understanding the label. https://www.sigmaaldrich.com/safety-center/understanding-the-label.html. (Accessed September 2020).

[36] H. Wang, Y. Zhang, X. Zhang, J. Li, Z. Lin, Z. Huang, J. Chang, Y. Zhang, J. Wang, C. Zhang, Y. Rao, An LC-MS/MS method for comparing the stability of ethanol's non-oxidative metabolites in dried blood spots during 90 days, Alcohol 83 (2020) 29–35.

[37] M. Luginbuhl, W. Weinmann, I. Butzke, P. Pfeifer, Monitoring of direct alcohol markers in alcohol use disorder patients during withdrawal treatment and successive rehabilitation, Drug Test. Anal. 11 (6) (2019) 859–869.

[38] S. Casati, A. Ravelli, I. Angeli, R. Durello, M. Minoli, M. Orioli, An automated sample preparation approach for routine liquid chromatography tandem-mass spectrometry measurement of the alcohol biomarkers phosphatidylethanol 16:0/18:1, 16:0/16:0 and 18:1/18:1, J. Chromatogr. A 1589 (2019) 1–9.

[39] B.C.H. van der Nagel, S. Wassenaar, S. Bahmany, B.C.P. Koch, Quantification of phosphatidylethanols in whole blood as a proxy for chronic alcohol consumption, using ultra performance convergence chromatography tandem mass spectrometry, Ther. Drug Monit. 40 (2) (2018) 268–275.

[40] I. Xie, Y. Xu, M. Anderson, M. Wang, L. Xue, S. Breidinger, D. Goykhman, E. J. Woolf, K.P. Bateman, Extractability-mediated stability bias and hematocrit impact: high extraction recovery is critical to feasibility of volumetric adsorptive microsampling (VAMS) in regulated bioanalysis, J. Pharmaceut. Biomed. Anal. 156 (2018) 58–66.

[41] L. Delahaye, E. Dhont, P. De Cock, P. De Paepe, C.P. Stove, Volumetric absorptive microsampling as an alternative sampling strategy for the determination of paracetamol in blood and cerebrospinal fluid, Anal. Bioanal. Chem. 411 (1) (2019) 181–191.

Supplementary Data

1.1 Chemicals, sample preparation and LC/MS Measurements in both labs

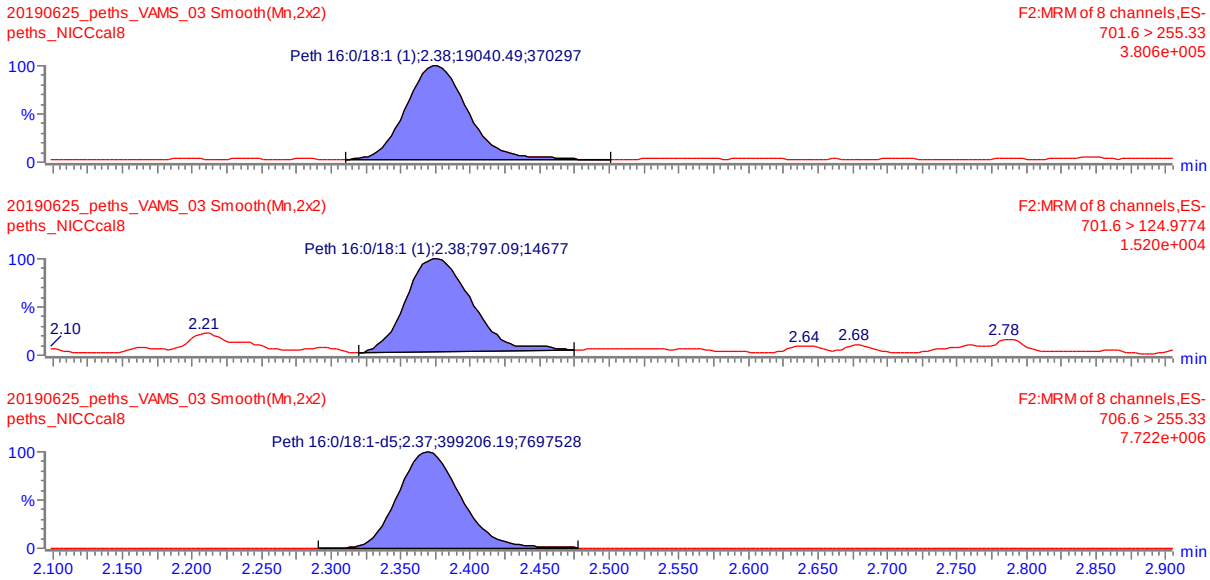
Table S1: Practical aspects of the sample preparation and LC/MS measurements as applied in the two different laboratories.

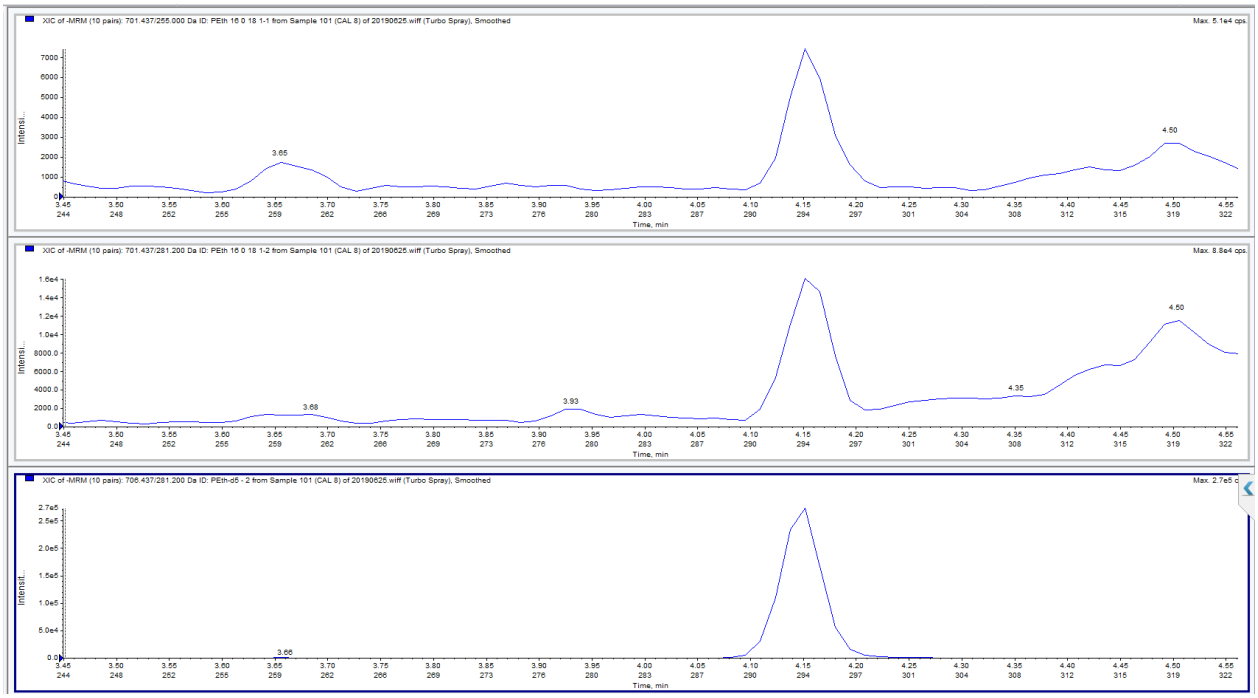
	UGent	NICC
Chemicals	Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanol (PEth 16:0/18:1) was obtained from Avanti Polar Lipids (Alabaster, Alabama, USA). As internal standard 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho(ethanol-d5)(ammonium salt) (PEth-D5) from Avanti Polar lipids (Alabaster, Alabama, USA) was used. Methanol (absolute, HPLC grade), isopropanol (HPLC grade), acetonitrile (HPLC grade) and tetrahydrofuran (ULC/MS grade) were purchased from BioSolve BV (Valkenswaard, the Netherlands). N-hexane (technical grade) was supplied by VWR (Leuven, Belgium). Ammonium acetate (ultra-pure, >98%) was obtained from Sigma Aldrich (Diegem, Belgium), formic acid (mass spectrometry grade) was from Honeywell (Bucharest, Romania). Ultrapure water was provided by a Synergy® Water Purification System (Merck Millipore, Overijse, Belgium).	Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanol (PEth 16:0/18:1) was obtained from Avanti Polar Lipids (Alabaster, Alabama, USA). As internal standard 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho(ethanol-d5)(ammonium salt) (PEth-D5) from Avanti Polar lipids (Alabaster, Alabama, USA) was used. Methanol (ULC-MS grade), isopropanol (ULC-MS grade), acetonitrile (ULC-MS grade) were purchased from BioSolve BV (Valkenswaard, the Netherlands). N-hexane (technical grade) was supplied by VWR (Leuven, Belgium). Ammonium acetate (ultra-pure, >98%) and formic acid (mass spectrometry grade) was obtained from Sigma Aldrich (Diegem, Belgium). Ultrapure water was provided by a Arium Comfort I® Water Purification System Sartorius, Göttingem, Germany).
Extraction of PEth 16:0/18:1 from the matrix	The absorptive tip of the VAMS sampler was disconnected from the plastic handler and transferred to a 2 mL Eppendorf tube. Then, 250 µL of a mixture of 2/8/0.2 (10 mM ammonium acetate 0.05% formic acid in water)/isopropanol/formic acid and 60 µL internal standard working solution (25ng/mL (0.034 µM) PEth-D5 in methanol) were added. Extraction was performed for 1 hour in a thermomixer comfort (Eppendorf®) at 23°C using 14000 rpm. Further liquid-liquid extraction (LLE) was done by adding 1 mL of n-hexane and mixing for 10 min at 23°C using 1400 rpm. The tubes were centrifuged (10 min, 14000 rpm (20160 g), 23°C) and then the clear supernatant was transferred to a total recovery vial (Waters, Zellik, Belgium).	
Evaporation of the supernatant	Under nitrogen using a Reacti-Vap evaporator (Pierce Biotechnology, Rockford, USA) for ~7 min at 50°C.	Using a rotational vacuum concentrator (RVC 2-33 IR, Martin Christ, Osterode am Harz, Germany) for approximately 12 minutes.
Reconstitution of the extracts	Dried samples were reconstituted in 50 µL of a 20/80 mixture of mobile phase A/mobile phase B.	Dried samples were reconstituted in 50 µL of a 25/75 mixture of mobile phase A/mobile phase B.
Injection volume	5 µL	5 µL (full loop), with overfill factor 2 times.

Mobile phases	Eluent A: 5 mM ammonium acetate solution in Millipore water; eluent B: 80/20 (v/v) mixture of methanol and isopropanol.	
Needle wash	Strong needle wash was a mixture of isopropanol/tetrahydrofuran (90/10, v/v).	Weak needle wash contained a mixture of eluent A and B (25/75, v/v). Strong needle wash was a mixture of methanol, acetonitrile, isopropanol, water and formic acid (25/25/25/25/0.1, v/v).
LC-MS/MS system	The analysis was performed using a Shimadzu Prominence setup (Shimadzu, Brussels, Belgium) including a CBM-20A system controller, two LC-20AD pumps, a DGU-20A5R degasser, a SIL-20ACHT autosampler (at 10°C) and a CTO-20AC column oven held at 55°C holding an XBridge BEH Phenyl 2.5 µm 4.6*150 mm XP column and an XBridge BEH Phenyl 2.5 µm XP VanGuard Cartridge, pumping a flow rate of 0.9 mL/min. The LC-system was coupled to a QTRAP 5500 instrument (SCIEX, Nieuwerkerk aan den Ijsel, The Netherlands), controlled by Sciex Analyst® 1.6.2 software.	An Acquity UPLC® system coupled to a Xevo TQS mass spectrometer (Waters, Manchester, UK) was used, using an Acquity UPLC® BEH Phenyl 1.7 µm x 2.1 x 100 mm column, with a flow rate of 0.35 mL/min. The LC-MS/MS was controlled by Masslynx® 4.2 software and the results were processed with Targetlynx®.
Ionization mode	Electrospray ionization, negative mode.	
Chromatography	The gradient initiated at 82% mobile phase B, which was linearly increased to 87.2% in 1.9 min and held for 1.1 min. From 3 to 5.4 min an isocratic phase of 99.2% eluent B was maintained as a washing step, followed by a 2.5 min re-equilibration at start conditions.	The gradient initiated at 75% mobile phase B, which was linearly increased to 95% in 3.5 min. From 3.5 to 4.5 min an isocratic phase of 95% eluent B was maintained as a washing step, which was followed by a 1.5 min re-equilibration at start conditions.
MS-detection	Multiple reaction monitoring mode was used and two ion transitions were monitored, a quantifier (Qt) and qualifier ion (Ql). For the internal standards only one ion transition was used. In addition, transitions for 2 other PEth species were monitored (i.e. 16:0/16:0 and 18:1/18:1- this is solely for qualitative purposes)	
Transitions	Quantitative: 701.4 → 255 (Qt) and 701.4 → 281.2 (Ql) for PEth 16:0/18:1 706.4 → 281.2 for PEth-d5. For validation of analytical results, the observed qualifier/quantifier ratio for samples should be within 0.945 to 1.755 of the ratio obtained for the calibrators (i.e. 1.35 times the ratio of the calibrator – with an allowable deviation of ±30%). Solely qualitative: 701.4 → 437.300 for PEth 16:0/18:1 706.4 → 255 for PEth-d5	Quantitative: 701.6 → 255.3 (Qt) and 701.6 → 125.0 (Ql) for PEth 16:0/18:1 706.6 → 255.3 for PEth-d5. Solely qualitative: 675.6 → 255.2 (Qt) and 675.6 → 125.0 (Ql) for PEth 16:0/16:0 727.6 → 281.2 (Qt), and 727.6 → 463.2 (Ql) for PEth 18:1/18:1

	675.4 → 255.1 and 675.4 → 437.2 for PEth 16:0/16:0 727.5 → 281.2 and 727.5 →463.2 for 18:1/18:1 184.3 → 184.3 for Phosphatidylcholines	
MS settings	The source temperature was set at 500 °C, the ion spray voltage at -4500 V. The nitrogen gas settings were as follows: curtain gas: 30 psi, ion source gas 1: 50 psi and ion source gas 2: 70 psi. Declustering potential was -5V, entrance potential -10V, collision energy -46V and collision cell exit potential -13V (Qt) or -19V (Ql).	The source temperature was 150 °C, capillary voltage 3 kV, desolvation temperature 650 °C, desolvation nitrogen flow 1000 L/h, cone nitrogen flow 50 L/h, collision Argon gas flow 0.15 mL/min, source offset 50V, nebulizer pressure 7 bar, entrance collision 1V, exit collision 0.5, gain 1.

Supplementary Figure S1 *Representative chromatograms for the PEth 16:0/18:1 analysis at NICC (upper panel) and UGent (lower panel). Both show a chromatogram for the lowest calibrator with a concentration of 10 ng/mL (0.014 µM)*





1.2 Details on the preparation of calibrators and QCs

To avoid issues with increased precision or poor accuracy related to pipetting blood (a difficult matrix to pipette), several measures are taken: (i) the use of fresh blood (cfr. stored or frozen blood may start to clot), (ii) avoiding pipetting volumes that are too small when handling blood (the volumes pipetted here are all 60 µl or higher) and (iii) every pipetting step influencing accuracy is controlled gravimetrically (for any matrix pipetted).

To prepare the stock solution from the reference standard obtained from Avanti Polar Lipids, the ampoule obtained from the manufacturer was allowed to reach room temperature (RT), opened and weighed (body and head). Then, PEth was quantitatively transferred by adding 5 times approximately 1 mL of methanol to the body and 5 times 0.2 mL of methanol to the head. Each time the walls of the body/head were carefully rinsed with methanol, after which the methanol was transferred to the same glass vial (of which the empty weight and final weight was also determined), using the same glass pipette. The final volume was topped off at 12.5 mL. The body and head of the original vial were then allowed to dry and the dried weight was determined. The difference in weight allowed us to calculate the exact amount of PEth in the original vial. Based on the difference in weight of the vial in which the solution of PEth was transferred, the exact amount of solvent can be calculated. This resulted in a master stock solution with a concentration of 8 mg/mL (11 mM). Two intermediate stock solutions of 250 µg/mL (345 µM) PEth 16:0/18:1 (one for the calibrators and one for the internal quality control samples (IQCs)) were diluted from the master stock solution using methanol. From the calibrator stock solution, eight intermediate solutions in water were prepared in polypropylene tubes, with concentrations of 0.25, 0.5, 1.25, 2.5, 6.25, 12.5, 25 and 50 µg/mL (0.345, 0.690, 1.72, 3.45, 8.62, 17.2, 3.45 and 6.90 µM). Ten µL of each of these solutions was then spiked to 240 µL whole blood in polypropylene tubes, yielding final calibrator concentrations of 10, 20, 50, 100, 250, 500, 1000 and 2000 ng/mL (0.014, 0.028, 0.069, 0.138, 0.345, 0.670, 1.38 and 2.76 µM). The lower and upper limit of quantification (LLoQ and ULoQ) were defined based on previous work (Kummer et al., 2016). For the IQC samples (4 levels – LLoQ, low, medium and high), a similar protocol was followed: 4 different intermediate solutions were made, and from these, 15 µL was added to variable volumes of whole blood, yielding final concentrations of respectively 10, 30, 500 and 1500 ng/mL (0.014, 0.041, 1.38 and 2.07 µM).

1.2.1 Direct comparison of the UGent and NICC stock solutions

The NICC stock solution was transferred to the UGent for direct comparison. To do so, 3 independent working solutions were generated from each of the two stock solutions. Each of these working solutions was further diluted with water to prepare 3 calibrators (2000 ng/mL (2.76 µM)). Hence, stock solutions were compared using a 2 *3 * 3 sampling protocol (2 stock solutions * 3 working solutions * 3 calibrators). From each of these 3 calibrators, 3 mixtures of calibrator and IS were prepared. The differences between the isotope ratios (area analyte/area IS) for the different solutions (normalized for their exact concentration) allow to assess the difference between the stock solutions. The allowable difference was set at 5% (1/3 of the allowable bias of the complete measurement procedure).

1.2.2 Details on the scouting experiment

The scouting experiment was performed to identify whether or not PEth results obtained in whole blood or after application to VAMS were comparable. To do so, a calibration curve and the IQC samples were measured together with 8 EQA samples. VAMS were prepared immediately after spiking the

blood. Application to VAMS was done in duplicate, the whole blood samples were run in singlicate. For the spiked QC samples, the average deviation in the results obtained by the two ways of processing the samples was 2% (ranging from -4% to 7%), while for the EQA samples, the difference was 39% (ranging from 28% to 58%). The experiment was repeated with leftovers from the initial experiments, stored at 4°C. VAMS were prepared 3 days later from the initially spiked blood, and the complete experiment was copied, only with 3 EQA samples at this time. On this occasion, the average deviation in the results obtained by the two ways of processing the samples was -5% (ranging from -16% to 3%), while for the EQA samples, the difference was -4.4% (ranging from -14.5% to 2.7%).

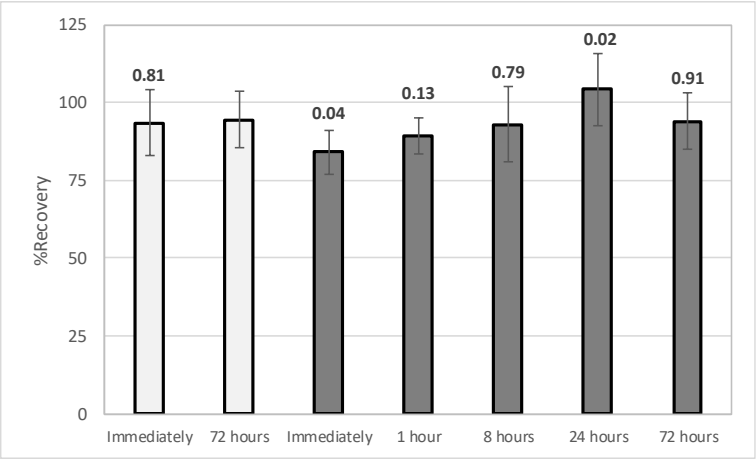
Hence, these scouting experiments revealed that discrepant results for authentic VAMS samples were obtained, depending on whether VAMS calibrators were prepared from freshly spiked blood vs. blood that was allowed to equilibrate with spiked PEth for some time. This was the impetus to systematically evaluate this variable.

1.2.3 Details on the experiment to determine the equilibration time needed

PEth concentrations were quantified in duplicate in 4 different external quality assessment (EQA) samples in both VAMS and whole blood, with calibrators and IQC samples prepared immediately after spiking whole blood and after equilibration at 4°C for 1, 8, 24 and 72 hours. Results are expressed as recovery, i.e. measured concentration/expected concentration *100. For the final method, calibrators were prepared by adding PEth to whole blood, followed by briefly vortexing and overnight equilibration at 4°C before sampling using the VAMS devices.

Supplementary Figure S2 shows the results for the EQA samples, measured with calibrators sampled immediately after spiking whole blood, or after storage for 72 hours at 4°C. Also shown are the results for VAMS-based calibrators, generated immediately after spiking whole blood, or after first equilibrating the blood at 4°C for 1, 8, 24 or 72 hours. Comparison of the obtained results with those obtained with liquid blood calibrators stored for 72 hours at 4°C, not applied to VAMS, revealed no significant differences for blood equilibrated for 1, 8 and 72 hours. For the final protocol the most time efficient option (i.e. overnight equilibration) was adopted. An impact of incubation time on the extractability was also observed by others (Nguyen et al., 2018b) and solved following our input (Luginbuhl et al., 2019a).

Supplementary Figure S2*: Influence of equilibration of calibrators on EQA results, expressed as recovery (n=8). Light grey: samples not applied to VAMS, dark grey: samples applied to VAMS. Data labels represent the P-values for a 2-sided paired student t-test (95% confidence), the error bars represent the SD on the recovery.*


125
100
75
50
25
0
%Recovery
0.81
0.04
0.13
0.79
0.02
0.91
Immediately
72 hours
Immediately
1 hour
8 hours
24 hours
72 hours

85

86

1.3 Details on the experimental set-up of the method validation

The selectivity was determined by investigating interfering peaks in 8 different blood samples from total alcohol abstainers. Peaks present in these blank samples should not exceed 20% of the peak area of the LLOQ and 5% of the IS.

Calibration curves were obtained by plotting the response (analyte : IS ratio) of 8 analyte concentrations against the spiked concentration levels. Eight curves were developed with data from eight different days. Homoscedasticity was investigated by performing an F-test at the lowest and the highest calibrator. To find the best calibration model, both weighted ($1/x$, $1/x^2$, $1/\sqrt{x}$, $1/y$, $1/y^2$ and $1/\sqrt{y}$) and unweighted linear and quadratic regressions were evaluated. In order to compare the different models with each other, the sum% relative error (%RE) was calculated and plotted against nominal concentrations. A selected model was confirmed as best model after a back-calculation was performed and the mean concentrations of the calibrators were within ± 15% of the nominal value or within ± 20% for the LLOQ.

The LoD, assessed at UGent, was estimated from the average signal:noise for the lowest calibrator (measured on 10 different days, in duplicate), by calculating what concentration would correspond with a S/N of 3. At the NICC, the LoD was not determined, as (quantitative) results are only reported from the LLoQ-concentration level onwards.

Carry-over was assessed by injecting a blank after the highest calibrator. This was repeated in every analytical run. Carry-over was considered acceptable if the obtained signal did not exceed 20% of that obtained for the LLoQ.

Matrix effect was examined using blood from seven different donors, with hct levels between 31.5 and 38.2 (31.5, 32.0, 32.8, 33.0, 35.0, 37.4, 38.2) at UGent and NICC. In addition, a sample with a low and with a high hct (respectively 24.3 and 48.0) was added. The blank blood samples were extracted. The supernatant was spiked with standard (at the level of QCL and QCH) and IS solutions before evaporation. The peak areas of these samples were compared with those obtained for similar spikes in neat extraction solvent. The peak area ratio analyte/IS of spiked extracts to those of the spiked solvents multiplied by 100 represents the IS-compensated matrix effect. The %RSD for the IS-compensated matrix effect should not exceed 15%.

The recovery of the entire procedure and the impact of the hct on the recovery was evaluated at low and high concentration (n=6), at three different prefixed hct levels (24.3, 41.3 and 48.0). The different hct levels were prepared by centrifuging blood with an hct of 41.3 and by adding or removing plasma. VAMS (10 µL) were made from blank blood, where the analytes were added post-extraction (in the supernatant). A second set of VAMS were prepared with blood spiked with the analyte. The recovery (%) was calculated by dividing the area of the analyte in the spiked blood VAMS to the area of the analyte spiked post-extraction multiplied by 100. To both sets, IS was added post extraction. The recovery was considered hct-independent when the relative IS-compensated recovery for both low and high hct was within 15% of the recovery for normal hct. The IS-compensated recovery is calculated as the peak area ratio analyte/IS of samples spiked with analyte before making VAMS to those spiked post-extraction.

In addition, to calculate the efficiency of the LLE procedure alone, 6 replicates were made with VAMS of spiked blood and addition of the IS before LLE. The efficiency is expressed as the peak area ratio analyte/IS of samples spiked with IS before LLE vs after LLE, multiplied by 100.

Autosampler stability and long term stability were evaluated at two concentrations (30 and 1500 ng/mL (0.041 and 2.07 µM)). Six samples at each concentration level were analyzed and kept in the autosampler for one and three days at 4°C before reinjecting them.

PEth was stable in extracts after 72 hours of storage in the autosampler at 4°C (differences ≤ 5.1%; **Supplementary Table S2**).

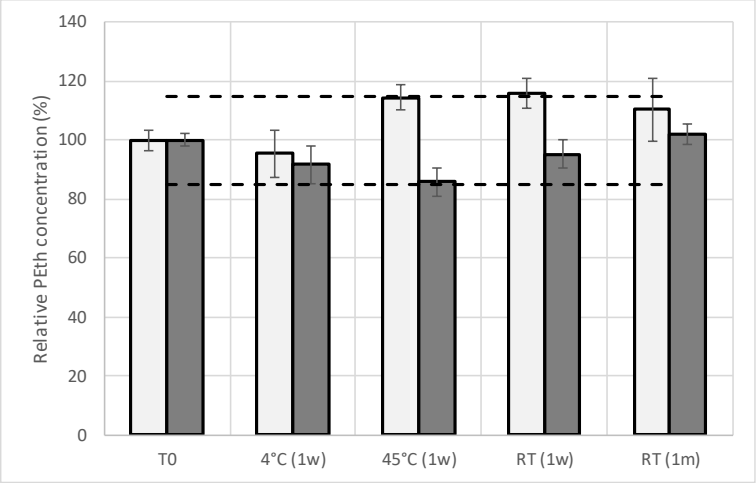
Supplementary Table S2. *Autosampler stability at 4°C.*

	%difference compared to T0	
	30 ng/mL (0.041 µM)	1500 ng/mL (2.07 µM)
Autosampler $_{24h}$	1.4	-0.1
Autosampler $_{72h}$	5.1	2.0

One week stability was investigated at 4°C, RT and 45°C to mimic possible situations during transport of the VAMS to the laboratory. One month stability was only established at RT because this mimics storage of the VAMS samples in the laboratory after arrival and before processing. In all situations the VAMS samples were stored in the accompanying plastic clamshell in a zip-closured plastic bag containing a 5 g MiniPax® absorbant packet from Sigma Aldrich®. The mean concentration of the QCs at the different time points should not exceed a variation of ± 20% from the nominal concentration. Additionally, long term stability can be assessed based on results obtained for the 4 EQA samples (concentrations respectively 735, 244, 465 and 146ng/mL (1.01, 0.337, 0.641 and 0.201 µM)), stored over a longer period of time (up to 200 days) at RT.

In VAMS samples, stability was demonstrated for at least one week at the three different evaluated temperatures (4°C, RT, and 45°C) and for one month at RT, as deviations did not significantly exceed ±15% (**Supplementary Figure S3**).

Supplementary Figure S3*: stability of PEth in VAMS under different storage conditions. Light grey: 30 ng/mL (0.041 µM), dark grey: 1500 ng/mL (2.07 µM). Concentrations are expressed relatively to T0 (RT: room temperature, w: week, m: month); The red lines represent the ±20% acceptance limits, error bars represent the 1-sided 95% confidence intervals.*



The trueness (expressed as bias) and precision were determined by analyzing the 4 QCs in duplicate for min 8 different days. The repeatability or within-run precision and the intermediate precisions are assessed using ANOVA single factor via excel. Trueness was established by dividing the difference between the mean of the measured concentration of the QCs and the nominal concentrations by the nominal concentration. One authentic QC sample with a concentration near the LLoQ was taken along for further evaluation of precision. In addition, 4 different EQA samples were also used. A z-score (i.e. (result - target)/(SD of all results)) less than 2 was set as the acceptance criterion.