Rachel S. Carling*, Zoe Barclay, Nathan Cantley, Nana Ghansah, Sarah L. Hogg, Alistair Horman, Stuart J. Moat, Simon Cowen, Chris Hopley, Chloe Deaves and Emily Whyte

Simple steps to achieve harmonisation and standardisation of dried blood spot phenylalanine measurements and facilitate consistent management of patients with phenylketonuria

https://doi.org/10.1515/cclm-2024-1367
Received November 22, 2024; accepted January 26, 2025; published online February 4, 2025

Abstract

Objectives: Management of phenylketonuria (PKU) relies upon life-long monitoring of phenylalanine (Phe) in dried blood spots (DBS), thus comparability of measurements is important. The lack of harmonisation and standardisation between laboratories, combined with the variable quality of patient-collected DBS specimens, are currently preventing this from being achieved. A traceable, matrix-matched Phe certified reference material, common methodology and means to ensure patient collected DBS specimens are of consistent quality would improve comparability between laboratories.

Methods: Baseline inter-laboratory (n=15) variation of DBS Phe was determined by triplicate measurement of four DBS materials, on three days. Laboratories prepared and analysed these samples using their routine method of analysis. A sub-set of laboratories (n=5) repeated the process using a common sample preparation and instrument methodology (LC-MS/MS), and three different calibration approaches. Samples prepared on dried blood spot microsampling cards (DBS-MCs) from whole blood, value assigned for Phe concentration by National Measurement Laboratories (NML), were then analysed using the harmonised methodology.

Results: Inter-laboratory co-efficient of variation (CV) differed with calibration approach; internal calibration 27.7 %; in-house aqueous calibration 4.7 %; centrally distributed aqueous calibration, 2.1 %. Inter-laboratory CV was reduced from 8.7 to 2.1 % by using common sample preparation and LC-MS/MS methodology. No significant difference was observed between consensus and assigned values for Phe in the four materials ($p>0.05$).

Conclusions: This study demonstrates a simple approach to harmonising and standardising DBS Phe measurements, traceable to value assigned materials. Combined with the introduction of DBS-MCs to ensure specimen quality, clinical laboratories can achieve comparability of patient results over time.

Keywords: harmonisation; standardisation; dried blood spots; phenylalanine; phenylketonuria; microsampling

***Corresponding author: Rachel S. Carling**, GKT School Medical Education, Kings College London, London WC2R 2LS, UK; and Biochemical Sciences, Synnovis, Guys & St Thomas' NHSFT, 4th Floor, North Wing, St Thomas' Hospital, Westminster Bridge Road, London, SE1 7EH, UK, E-mail: Rachel.Carling@kcl.ac.uk

Zoe Barclay, Biochemical Sciences, Synnovis, Guys & St Thomas' NHSFT, London, UK

Nathan Cantley, Department of Clinical Biochemistry, Severn Pathology, Southmead Hospital, North Bristol NHS Trust, Bristol, UK

Nana Ghansah, Neurometabolic Unit, National Hospital for Neurology and Neurosurgery, London, UK

Sarah L. Hogg, Biochemical Genetics Unit, Cambridge University Hospitals, Cambridge, UK

Alistair Horman, Chemical Pathology, Great Ormond Street Hospital, London, UK

Stuart J. Moat, Department of Medical Biochemistry, Immunology & Toxicology, University Hospital Wales, Cardiff, UK; and School of Medicine, Cardiff University, University Hospital Wales, Cardiff, UK

Simon Cowen, Chris Hopley, Chloe Deaves and Emily Whyte, LGC, Middlesex, UK

Introduction

Management of phenylketonuria (PKU, OMIM 261600) relies upon life-long monitoring of dried blood spot (DBS) phenylalanine (Phe) so comparability of results independent of time, place and measurement procedure is important, particularly when an individual patient's test results are compared with clinical decision points described in evidence-based clinical practice guidelines; The 2017 European guidelines for diagnosis and management of patients with PKU recommend consensus age-related blood Phe target treatment ranges to prevent adverse neurological outcomes and specify three clinical decision points (120, 360 and 600 µmol/L) [1], and sapropterin responsiveness is defined by a decrease in blood Phe ≥30 % [2]. It is evident

from recent work, based on DBS samples created in the laboratory under ideal conditions, that the current analytical performance of this test needs to be improved, with the inter-laboratory coefficient of variation (CV) of Phe reported to be 20.1 % [3]. Furthermore, when pre-analytical factors are considered, i.e., the quality of a patient collected DBS specimen, additional variability of approximately 15–20 % CV may be seen [4]. Thus, any measures to improve comparability of DBS Phe results must include consistent collection of a good quality DBS specimen from a heel/finger prick specimen, as well as a harmonised and/or standardised approach to measurement. This would facilitate comparability of patient results over time, enabling consistent management of patients across centres, and movement of patients between centres. Additionally, it would also provide confidence in the assessment of sapropterin responsiveness and ensure laboratory data from clinical research and trials was directly transferrable.

Data from the 2023 European Research Network for the evaluation and improvement of screening, Diagnosis and treatment of Inherited disorders of Metabolism (ERNDIM) Special Assays in Dried Blood Spots external quality assessment (EQA) scheme show that the mean inter-laboratory %CV for Phe (n=92) and Tyrosine (Tyr) (n=86) in DBS was 19.2 % and 15.1 % respectively [5]. However, mean intra-laboratory imprecision was 6.3 % and 6.7 % for Phe and Tyr respectively, demonstrating that accuracy is a significant contributor to inter-laboratory variation and highlighting a limitation of this EQA scheme; participant results are scored relative to the all-laboratory trimmed mean which does not provide any indication of 'trueness' or metrological traceability. At present, existing EQA schemes do not use matrix matched certified reference materials (CRM) with concentrations determined traceable to the international system of units (SI) and this poses a challenge for clinical laboratories. An SI traceable, matrix matched Phe CRM would standardise an EQA scheme and enable participating laboratories to compare their results against the true value.

Another contributing factor is that methods for the measurement of Phe in DBS are typically legacy methods, remaining unchanged since they were first introduced for newborn screening over 30 years ago [6], yet the analytical performance requirements of a screening method differ from those used for longitudinal measurements. The situation is complicated further because most centres rely on laboratory developed tests (LDTs); In the UK, 15/16 laboratories use flow injection analysis tandem mass spectrometry (FIA-MS/MS) which lacks the selectivity of liquid chromatography tandem mass spectrometry (LC-MS/MS) and furthermore, 8/16 laboratories quantify results without the use of an external calibrant, hence the assumption about the volume of blood present in the sub-punch of the DBS specimen analysed, directly impacts the calculated concentration of Phe & Tyr. It is common practice in the UK to assume a 3.2 mm sub-punch is equivalent to 3.1 µL of blood and this has been reported to introduce a negative bias of approximately 30 % to DBS Phe measurements [7].

There were three aims to this study. Firstly, to ascertain the current inter-laboratory variation of DBS Phe results in the UK. Secondly, to determine if harmonisation of results could be achieved by implementing a common sample preparation protocol, methodology and calibration approach and thirdly, to assess standardisation by comparison of consensus Phe results with four DBS materials prepared from value assigned whole blood. Underpinning this was the requirement to demonstrate that any recommended approach was compatible with the use of microsampling devices, thus ensuring that improvements in analytical performance would not be negated by poor-quality patient-collected DBS specimens. The four value assigned DBS materials were prepared on Capitainer®B fixed-volume dried bloodspot microsampling cards (DBS-MCs). Tyr has typically been measured simultaneously with Phe, to demonstrate that patients with PKU are not Tyr deficient due to their synthetic diet, hence is has been included in this study although it is acknowledged that with the current dietary management principles, this analyte is of less importance.

Materials and methods

Full details of the materials and methods used are provided in Supplementary Material S1, including those of the value assigned materials.

Study samples

Pooled blood was obtained from a healthy volunteer, in line with local governance arrangements, and split into four equal portions. Baseline Phe and Tyr concentrations were established and subsequently three of the four pools were enriched to approximate concentrations of 360, 600, and 1,250 µmol/L Phe and Tyr. Traditional (non-volumetric) DBS samples were prepared from each pool by pipetting 50 µL aliquots of blood onto Whatman 226 filter paper. Fixed volume DBS were prepared from each pool by accurately

pipetting 10 µL aliquots of blood on to perforated Whatman 222 filter paper. All DBS specimens were allowed to dry at ambient temperature for at least 4 h and then stored frozen until use in foil bags with desiccant. A third set of specimens were prepared from the value assigned materials on 10 µL fixed-volume DBS-MCs as described in Supplementary Material S1.

Preparation of value assigned materials

Value assignment of the pooled blood materials was performed using a higher order reference method procedure (RMP), which used double exact matching isotope dilution mass spectrometry methodology and gas chromatography tandem mass spectrometry (GC-MS/MS) for measurements. Measurement of derivatised extracts by GC-MS/MS provided optimum ratio precision, to achieve the desired expanded uncertainty (<3 %) associated with Phe and Tyr values in the materials. The assigned value is the mass fraction of Phe and Tyr determined using gravimetric preparation. The assigned value is considered traceable to the SI through use of pure certified reference material for instrument calibration (National Measurement Institute of Japan CRM 6014-a and 6019-a). The quoted uncertainty is the half-width of the expanded uncertainty interval calculated using a coverage factor, *k*, of 2.0, which gives a level of confidence of approximately 95 %. Details of double exact matching isotope dilution mass spectrometry methodology, sample preparation and the GC-MS/MS method can be found in Supplementary Material S1.

The mean results in µg/g with combined expanded uncertainty were converted to µmol/L. The density of the blood was accurately determined to maintain traceability of the converted concentration to the SI base unit, the kilogram. The weight of fixed volume aliquots of ultrapure water, the value assigned materials, and an internal quality control (IQC) solution was determined. The volume of each aliquot was calculated using the average weight of the water aliquots and the density of water considering temperature and air pressure. The density of the value assigned materials and the IQC solution was obtained by relating the average weight of the aliquots of each material or solution with the previously calculated volume dispensed. The uncertainties of both the blood densities and the molecular masses were calculated and combined with the uncertainty of the value assignment. Details of the method for accurate and traceable unit conversion are included in Supplementary Material S1.

Participants

All laboratories in the UK Metabolic Biochemistry Network (n=16) were invited to participate in this study. Fifteen laboratories agreed to participate, and 15/15 laboratories returned results (see Supplementary Table S1).

Study design

Four traditional (non-volumetric) DBS, four fixed volume (10 µL) DBS and four value assigned materials prepared on DBS-MCs were shipped to participating laboratories. Specimens were transported in sealed foil bags with desiccant, at ambient temperature, and stored frozen upon receipt. Each laboratory provided details of their sample preparation, methodology and calibration (see Supplementary Table S1, Supplementary Material S1). Participants (n=15) analysed the four traditional DBS samples in triplicate on each of three different days, using their routine method of analysis (LDT). A sub-set of participants (n=5) then analysed the four fixed volume DBS samples, in triplicate on each of three different days, using the common sample preparation and LC-MS/MS protocols provided (see Supplementary Material S1) and three different calibration approaches. The harmonised methodology was used to analyse the four value assigned materials prepared on DBS-MCs in triplicate on each of three different days.

The three calibration approaches evaluated were as follows: Internal calibration (IC) refers to single point isotope dilution calibration in which a fixed amount of a stable isotope label (SIL) is added to each sample. The concentration of the analyte of interest is determined using the ratio of the peak area response of the endogenous analyte to that of the stable isotope label. The SILs were sourced and prepared independently by each laboratory. Common multi-point aqueous calibration (CMAC) refers to an aqueous, multi-point (six plus blank) calibration prepared centrally from a certified reference material (*Trace*CERT, Sigma-Aldrich) and distributed to each participating laboratory. A fixed amount of SIL is added to each calibrator (and sample). The concentration of the analyte of interest is determined using a calibration curve whereby the ratio of the peak area response of the endogenous analyte to that of the SIL is plotted against the assigned value of the calibrator. The SILs were sourced and prepared independently by each laboratory. In-house multi-point aqueous calibration (IMAC) refers to an aqueous multi-point (six plus blank) calibration prepared by each laboratory from independently sourced

materials to specified concentrations. A fixed amount of SIL is added to each calibrator (and sample). The concentration of the analyte of interest is determined as for the CMAC.

Statistical analysis

Datasets were analysed with mixed effects models using the R statistical programming environment [8] with the lme4 package [9]. Laboratory and day nested within laboratory were treated as random effects, while calibration strategy (IC, IMAC or CMAC) was treated as a fixed effect. Where the distribution of the results permitted (that is, where the distribution was sufficiently close to lognormal with a constant CV) and a better fitting model was obtained, data from all four samples were aggregated, a log transform applied, and a single model fitted with sample as an additional fixed effect. Otherwise, the data for each sample were analysed separately without applying the log transform. The data for the first two stages were analysed using the latter approach, the former approach was used for the final stage.

The significance of the observed differences in the between-laboratory variability for the different calibration strategies was tested by stratifying the random effects by calibration strategy (modelling a different variance for each), fitting the model then removing the stratification, refitting and testing the two models with a likelihood ratio test. Model diagnostics such as residuals were also examined to determine which model provided the better fit.

Results

Using each centres routine method of analysis and calibration (LDTs), inter-laboratory %CV (n=15) for Phe at nominal concentrations of 60, 300, 600 and 1,250 µmol/L was 11.6 %, 8.7%, 6.6% and 7.0 %, respectively. Inter-laboratory %CV for Tyr at nominal concentrations of 62, 300, 600 and 1,250 µmol/L was 13.7 %, 10.2%, 9.9% and 7.7 %, respectively. The root mean square averages for Phe and Tyr were 8.7 and 10.6 %, respectively (see Figure 1 and Supplementary Figures).

When DBS samples were of fixed volume and sample preparation and methodology were harmonised, mean inter-laboratory CV (n=5) was reduced significantly ($p<0.01$) compared with the in-house approach (2.1 vs. 8.5 %). Calibration approach was shown to have a significant impact on variability. Mean inter-laboratory %CV (n=5) for Phe when results were quantified using IC, IMAC and CMAC were 27.7 %, 4.7 % and 2.1, respectively. Mean inter-laboratory % CV (n=5) for Tyr when results were quantified using IC, IMAC and CMAC were 27.7 %, 5.8% and 2.9 %, respectively. IC was appreciably more variable than IMAC and CMAC. See Figure 2 and Supplementary Figures. The statistical model provided the better fit when the random effects were stratified by calibration strategy suggesting the differences in the between-laboratory variability were significant.

The concentration of Phe in the DBS-MCs, LGC 001 to LGC 004, was traceably value assigned as 120 ± 2.1, 339 ± 5.9, 636 ± 9.1 and 1,251 ± 21 µmol/L (95 % CI, k=2), respectively. The concentration of Tyr in these materials was traceably value assigned as 103 ± 2.1, 318 ± 5.0, 620 ± 8.9 and 1,226 ± 20 µmol/L (95 % CI, k=2), respectively. Using the harmonised methodology and the CMAC calibration, the consensus estimates for Phe in the four value assigned materials were 123.8 ± 5.4, 339 ± 14, 647 ± 13 and 1,262 ± 37 µmol/L (95 % CI, k=2), respectively. Consensus estimates for Tyr were 110 ± 15, 322.3 ± 7.9, 633 ± 14 and 1,248 ± 44 µmol/L (95 % CI, k=2), respectively (see Table 1, Figure 3 and Supplementary Figures). The consensus estimates agreed with the assigned values for both analytes in all materials ($p>0.05$), demonstrating metrological traceability to a RMP.

Discussion

The current inter-laboratory variation seen for DBS Phe and Tyr in the UK is of a similar magnitude to that described elsewhere [10] and was the largest contributor to variability for both analytes in each DBS material analysed; intra-laboratory variability was not significantly different amongst the 15 laboratories. The variability associated with different calibration approaches was significant, with IC being associated with inter-laboratory variability of 28 %. Whilst IC may be an acceptable approach for a screening method, this evidence supports the requirement for a different approach to longitudinal monitoring. Variability of this magnitude could have a significant impact on patient management resulting in inappropriate dietary change and potentially adverse patient outcomes.

With more than 50 % of UK laboratories currently utilising IC, there may be merit in producing best practice guidelines to help implement this change. Although the difference between the CMAC and IMAC approaches was significant, with the CMAC producing the most consistent inter-laboratory results, the magnitude of the difference was marginal in the context of patient results and either calibration approach would be clinically acceptable.

Ensuring comparable results for patients with PKU is important; patients are often seen in a variety of health care settings in which different laboratories may use different

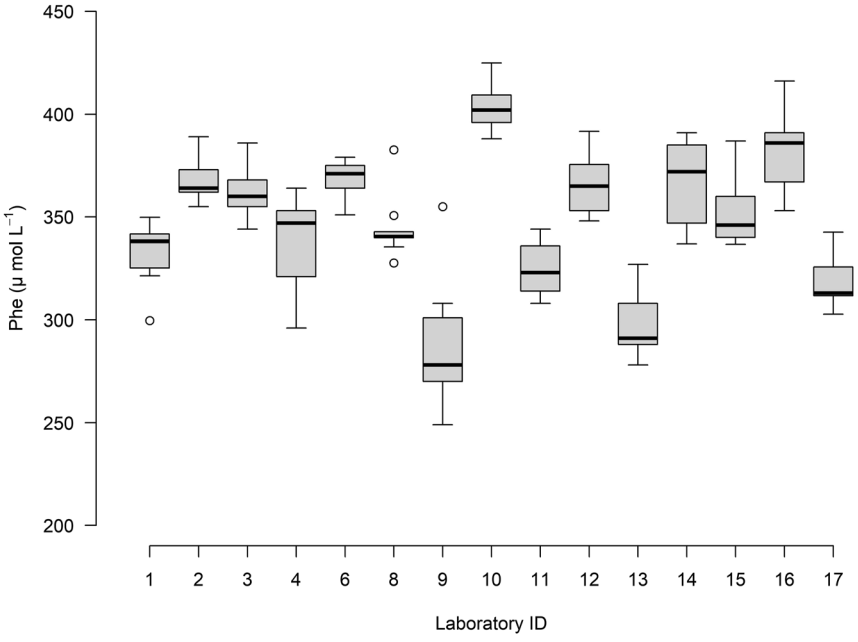

Figure 1: Box-plots showing intra- and inter-laboratory variation of phenylalanine in a traditional DBS sample (360 μmol/L). Analysis was in triplicate over three days. Each laboratory used their routine method of analysis. The box represents the interquartile range, the median is denoted by the solid horizontal line and the whiskers extend to the remaining data with open circles representing outlier data points.

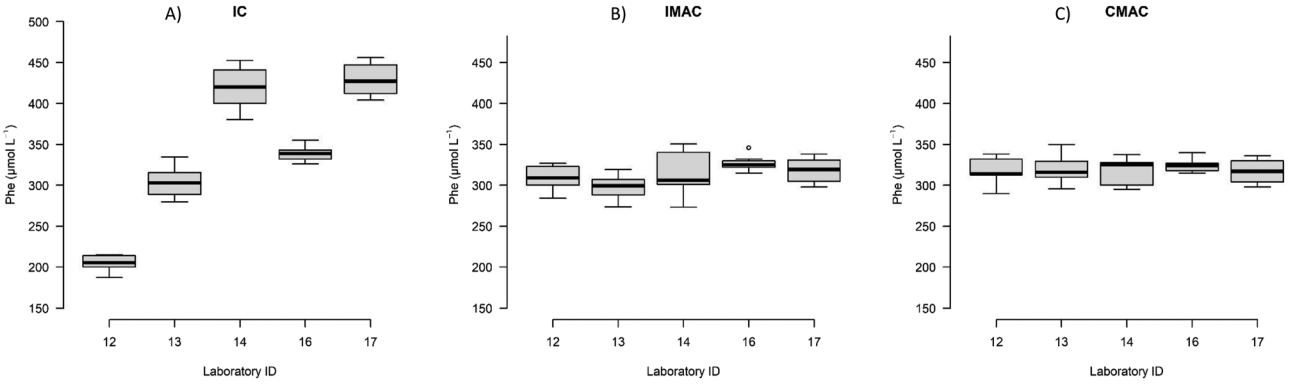

Figure 2: Impact of the three calibration approaches on phenylalanine results from the sub-set of five laboratories with result quantified by (A) internal calibration, (B) in-house multi-point aqueous calibration and (C) common multi-point aqueous calibration. Analysis was in triplicate on three different days, with each laboratory using the common sample preparation protocol and LC-MS/MS method. Box plots as defined in Figure 1.

Table 1: Comparison of the consensus estimates of phenylalanine and tyrosine in the dried blood spot microsampling cards measured using the harmonised methodology with the assigned values.

Analyte	Sample	Consensus value, μmol/L	Expanded uncertainty, μmol/L	Assigned value, μmol/L	Expanded uncertainty, μmol/L	t-Value	p-Value
Phe	LGC 001	123.8	5.4	120	2.1	1.332	0.183
	LGC 002	339	14	339	5.9	−0.082	0.935
	LGC 003	647	13	636	9.1	1.442	0.149
	LGC 004	1,262	37	1,251	21	0.672	0.502
Tyr	LGC 001	110	15	103	2.1	1.408	0.159
	LGC 002	322.3	7.9	318	5.0	0.964	0.335
	LGC 003	633	14	620	8.9	1.712	0.087
	LGC 004	1,248	44	1,226	20	1.167	0.243

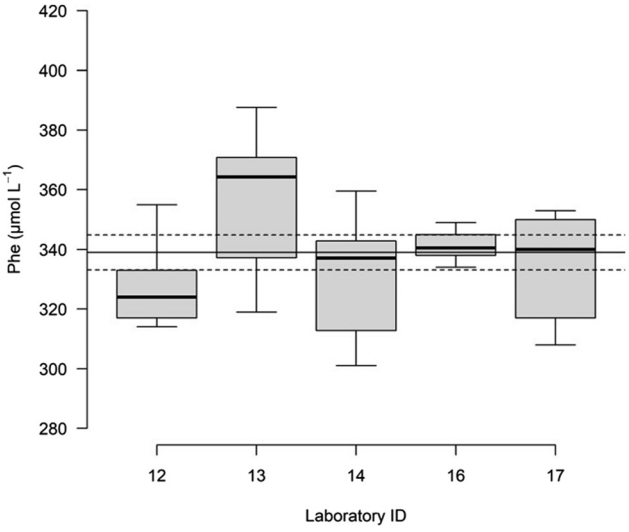

Figure 3: Box plots comparing consensus estimates with assigned value of phenylalanine in blood collected on a Capitainer®B dried blood spot microsampling card and measured by five laboratories using harmonised methodology and quantifying against common multi-point aqueous calibration. The dashed horizontal line running the length of the axis is the assigned value of phenylalanine (339 ± 5.9 µmol/L, 95 % CI, k=2).

measurement procedures; measurement procedures used by a given laboratory may change whilst patients are being monitored; when comparing a patient's Phe results with clinical decision points described in evidence based management guidelines, results that are not comparable across measurement procedures can lead to inconsistent dietary management/advice. The results reported by 15 laboratories for a single DBS material with a Phe concentration at a clinical decision point (360 µmol/L) illustrate this point with results ranging from 249 to 425 µmol/L. Analysis of the four value assigned materials on the DBS-MCs using the harmonised methodology enabled the overall accuracy of the method to be assessed against a metrologically traceable matrix matched material; standardisation of the Phe and Tyr results was demonstrated across the entire range of concentrations measured. Therefore, a matrix CRM with proven commutability would provide clinical laboratories with a route to achieving standardisation of Phe measurements.

Whilst minimizing uncertainty of the measurement procedure is useful, it will only have merit if pre-analytical variables such as haematocrit and volume of blood applied to the filter paper card are controlled as these can introduce variability of a far greater magnitude (up to 30 %) [4]. Obtaining good quality, accurate and precise DBS specimens is essential if improvements in analytical performance are to translate into improved patient management. Given that the majority of patients with PKU are self-monitored from home, via capillary blood collected from a finger prick onto traditional filter paper card, this has proved challenging. However, the introduction of fixed volume microsampling devices in recent years has overcome this issue, providing a simple specimen collection technique that has been shown to consistently produce a good quality, accurate and precise DBS sample that is acceptable to patients [3, 7].

Adopting a harmonised and standardised approach would be advantageous for medical laboratories. Analytical methods, particularly LDTs, are under increasing scrutiny. Regulatory requirements mean laboratories must be able to demonstrate the clinical utility of a given test and provide assurances as to its traceability, robustness and uncertainty of measurement. The benefits of calibrating against matrix matched materials are well documented; commutability, accuracy, reliability, recovery [11] and US FDA Bioanalytical Method Validation Guidance for Industry [12] and EMA Guideline on Bioanalytical Method Validation and Study Sample Analysis [13] both state that calibration standards should be prepared in the same biological matrix as the samples in the intended study. This approach would also align with current expert opinion: international entities such as the International Federation of Clinical Chemistry & Laboratory Medicine, the WHO, the European Joint Research Centre Institute of Reference Materials and Methods and the Joint Committee for Traceability in Laboratory Medicine are organising a range of activities and efforts directed toward harmonisation and standardisation of clinical laboratory testing, emphasising the importance of these concepts [11, 14–16].

The value assigned materials were produced solely for this study. A commercially available supply of a matrix matched certified reference material would require many years development, extensive commutability studies and ongoing stability assessments. Furthermore, the costs associated with producing and distributing a whole blood material are significant in comparison to a calibration solution, making this a less viable commercial proposition. In the absence of such a material, whilst it will not be possible to achieve metrological traceability, the results of this study provide evidence of a route to harmonisation. Harmonisation is acknowledged to be an acceptable alternate approach when standardisation cannot be established by conventional means [11] and would ensure patient results were comparable. This study also demonstrates the accuracy of measurement, within a sub-set of clinical laboratories, in comparison to a RMP as there was no significant difference in the consensus estimates and assigned values of the four materials when the harmonised methodology was used. Although the CMAC was prepared from a commercially available aqueous CRM and is therefore not considered commutable, accurate results were obtained by this

approach which will provide a simple, inexpensive, readily available solution for medical laboratories to standardise Phe measurements and improve comparability of patient results. This approach has practical merit too. Preparation of matrix matched calibrators for endogenous compounds like Phe, which are already present in the sample matrix, is not entirely straightforward and the DBS matrix itself poses further complications, for example, influence of haematocrit, volume of blood applied to filter paper and the stability and storage requirements of such materials [7].

Conclusions

This study demonstrates a simple approach by which laboratories can achieve standardisation of DBS Phe and Tyr results by adopting an LC-MS/MS method that has been validated against a higher order RMP with a traceable material. It also provides evidence to support the recommendation that IC is not an acceptable approach for the longitudinal monitoring of patients with PKU. Combined with the introduction of DBS-MCs to ensure the consistent collection of good quality DBS specimens from patients, the uncertainty associated with DBS Phe & Tyr results can be minimised. This will improve the long-term management of patients with PKU, facilitate movement of patients between centres and ensure the quality of data collected as part of on-going and future outcome studies for PKU. Best practice guidelines will be key to ensuring laboratories adopt these recommendations.

There would be merit in investigating whether a similar approach to that described here might be applicable to other DBS tests routinely used to monitor dietary therapy in inherited metabolic disease, for example, branched chain amino acids in maple syrup urine disease, lysine in Glutaric Aciduria type 1 and total homocysteine/methionine in classical homocystinuria.

Acknowledgments: Many thanks to the UK Metabolic Biochemistry Network for analysis of the baseline samples. Ann Bowron, Newcastle; Catherine Collingwood, Liverpool; Robert Barski, Leeds; Philippa Goddard, Birmingham; Patricia Fitzsimmons, Dublin; Hoiyee Wu, Manchester; Katherine Wright, Sheffield; Jane McNeilly, Glasgow; Gillian Hamilton, Belfast.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interest: The authors state no conflict of interest.

Research funding: This project was funded by NHS England Specialised Commissioning. Emily Whyte, Chloe Deaves and Chris Hopley also received funding from UK government Department for Science, Innovation & Technology (DSIT).

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. van Spronsen FJ, van Wegberg AM, Ahring K, Belanger-Quintana A, Blau N, Bosch AM, et al. Key European guidelines for the diagnosis and management of patients with phenylketonuria. Lancet Diabetes Endocrinol 2017;5:743–56.
2. Muntau AC, Adams DJ, Belanger-Quintana A, Bushueva TV, Cerone R, Chien YH, et al. International best practice for the evaluation of responsiveness to sapropterin dihydrochloride in patients with phenylketonuria. Mol Gen Metab 2019;127:1–11.
3. Carling RS, Barclay Z, Cantley N, Emmett EC, Hogg SL, Finezilber Y, et al. Investigation of the relationship between phenylalanine in venous plasma and capillary blood using volumetric blood collection devices. JIMD Reports 2023;64:468–76.
4. George RS, Moat SJ. Effect of dried blood spot quality on newborn screening analyte concentrations and recommendations for minimum acceptance criteria for sample analysis. Clin Chem 2016;62:466–75.
5. ERNDIM Special assays in dried blood spots annual report 2023. https://www.erndim.org/meetings-reports-cat/2023-eqa-scheme-annual-reports/ [Accessed Sept 2024].
6. Carling RS, Whyte E, John C, Garstone R, Goddard P, Greenfield T, et al. Improving harmonisation and standardisation of expanded newborn screening results by optimization of the legacy flow injection analysis tandem mass spectrometry methods and application of a standardized calibration approach. Clin Chem 2022;68:1075–83.
7. Carling RS, Emmett EC, Moat SJ. Evaluation of volumetric blood collection devices for the measurement of phenylalanine and tyrosine to monitor patients with phenylketonuria. Clin Chim Acta 2022;535:157–66.
8. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. R Core Team (2022). Available from: https://www.R-project.org/.
9. Bates D, Machler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Software 2015;67:1–48.
10. Coene K, Timmer C, Goorden S, Ten Hoedt A, Kluijtmans L, Janssen M, et al. Monitoring phenylalanine concentrations in the follow-up of phenylketonuria patients: an inventory of pre-analytical and analytical variation. J Inherit Metab Dis 2021;58:70–9.
11. Vesper HW, Myers GL, Miller WG. Current practices and challenges in the standardisation and harmonisation of clinical laboratory tests. Am J Clin Nutr 2016;104:907S–12S.
12. Food and Drug Administration. Guidance for industry: Bioanalytical method validation. Rockville, MD: US Department of Health and Human Services, FDA, Center for Drug Evaluation and Research and Center for

Veterinary Medicine; 2018. Available from: https://www.fda.gov/media/70858/download.
13. ICH Guideline. M10 on bioanalytical method validation and study sample analysis. Amsterdam, Netherlands: European Medicines Agency; 2023.
14. Panteghini M, Camara JE, Delatour V, Van Uytfanghe K, Vesper HW, Zhang T. Feasibility of metrological traceability implementation using the Joint committee on traceability in laboratory medicine database entries including the fulfillment of "Fit-for-Purpose" maximum allowable measurement uncertainty. Clin Chem 2024;70:1321–33.
15. Seger C, Kessler A, Taibon J. Establishing metrological traceability for small molecule measurands in laboratory medicine. Clin Chem Lab Med 2023;61:1890–901.
16. Greaves R, Mackay L. The development of reference measurement procedures to establish metrological traceability. Clin Chem Lab Med 2023;61:1887–9.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2024-1367).