Journal of Mass Spectrometry and Advances in the Clinical Lab 20 (2021) 42–47

Contents lists available at ScienceDirect

Journal of Mass Spectrometry and Advances in the Clinical Lab

journal homepage: www.sciencedirect.com/journal/journal-of-mass-spectrometry-and-advances-in-the-clinical-lab

A data-driven approach for the detection of internal standard outliers in targeted LC-MS/MS assays

E.H. Wilkes *, M.J. Whitlock, E.L. Williams

Department of Clinical Biochemistry, North West London Pathology, Imperial College Healthcare NHS Trust, Charing Cross Hospital, Fulham Palace Road, London W6 8RF, United Kingdom

ARTICLE INFO

Keywords:
Internal standard
Quality control
Mixed-effects models
LC-MS/MS

ABSTRACT

Heavy-labelled internal standard (IS) compounds are commonly used in liquid chromatography-tandem mass spectrometry (LC-MS/MS) assays to control for stochastic and systematic variation. Identifying samples that suffer from unwanted variation is critically important in order to avoid factitiously inaccurate results. Current approaches for outlier detection typically employ arbitrary thresholds and ignore systematic drift. To improve this, we applied robust linear mixed-effects models (LMMs) to capture the within- and between-run variability in IS signal and generate data-driven acceptance ranges for routine use.

Data from in-house LC-MS/MS assays for 25-hydroxyvitamin D_3 and D_2 and prednisolone were retrospectively collected. The variation in the percentage deviation of the internal standard area from the mean of the calibrators was modelled through the use of robust LMMs. The fitted LMMs revealed significant positive drift in IS signal over the analytical runs for vitamin D, with slope coefficients of 0.118 (95% CI: 0.098, 0.138) and 0.192 (0.168, 0.215) for D_3 and D_2, respectively. In contrast, the models for prednisolone demonstrated a significant negative drift in IS signal, with a slope coefficient of −0.164 (−0.297, −0.036). Non-parametric, cluster bootstrap resampling enabled us to define acceptance ranges for use in future assays.

Here, we have described a computational approach to extensively characterise the variation in IS signal in routinely-performed LC-MS/MS assays. This approach facilitates a robust quality assessment of IS outliers in routine practice and thus has the potential to improve patient safety. Importantly, this approach is applicable to other MS assays where linear variation in IS signal is observed.

Introduction

Deuterated or ^{13}C-labelled internal standards (IS) are commonly used in targeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) assays to control for stochastic and/or systematic variation in analyte extraction and analysis. This is typically achieved by adding a fixed amount of IS to each sample prior to further processing. Due to their structural similarity to the compound of interest, heavy element-labelled internal standards behave effectively identically during extraction and mass spectrometric analysis. As a result of the selectivity of mass spectrometers, however, the analyte and IS can be distinguished by their unique mass-to-charge ratios. The analyte concentration in each sample is then interpolated by calculating the ratio between the analyte's peak area/height and the peak area/height of the IS (i.e., the response ratio). Despite the success of this approach, it is imperative to be able to detect outliers in IS signal in order to avoid falsely suppressed or enhanced response ratios that may lead to inaccurate estimates of analyte concentration. Such anomalies in IS signal can result from ion suppression or enhancement (caused by a co-eluting interferent in the sample matrix) or from analytical errors (e.g., inaccurate or imprecise addition of the internal standard) [1,2]. In order to detect such outliers, IS signals can be expressed as a ratio to the mean IS signal of the calibrators or other, stable reference sample, in which it is assumed that no suppression or enhancement effects occur (or where these are identical to those observed in patient samples). These values, termed "relative deviations", can then be assessed against acceptance limits in order to identify outliers [1]. Once identified, the extraction of these samples can be repeated or further investigations performed (e.g., dilution series) [3]. There is, however, a dearth in the clinical mass spectrometry literature regarding the definition of these acceptance limits and the typically suggested fixed limits (e.g. −50 to +50% [1]) are problematic as: (i) they are largely arbitrary and thus may not detect more subtle

* Corresponding author.
E-mail address: edmund.wilkes@nhs.net (E.H. Wilkes).

https://doi.org/10.1016/j.jmsacl.2021.06.001
Received 30 November 2020; Received in revised form 2 June 2021; Accepted 3 June 2021
Available online 8 June 2021

variations in IS signal; (ii) they do not account for potential drift in the measurement system, leading to varying detection power across the course of a run; and (iii) such limits do not appropriately account for the within- and between-variability in IS signal observed in reality. Literature from other fields, such as bioanalysis, has detailed approaches for the detection and assessment of IS outliers [4]. These guidelines define a rule-of-thumb approach that makes use of intra-assay low and high IS signal references (Ref_{low} and Ref_{high}), against which sample-specific deviations can be judged. Although these guidelines incorporate rules for the detection of systematic deviations, the proposed approach still relies on arbitrary limits (e.g., <10% or >2x the signal of the Ref_{low}) and the definition of local minima and maxima will be sensitive to outliers in IS signal.

Linear mixed-effects models (LMMs) – so-called because they combine both fixed/constant and random/varying effects – are useful tools for the estimation of experimental factors that may vary between groups within a hierarchy (the varying effects), while also estimating the effects that are identical for all groups within a hierarchy (the constant effects) [5,6]. In a typical experimental design, LMMs can therefore be used to assess the relationship between an outcome variable and its predictor variables (e.g., run position or time) and appropriately model the variance at each level of the experiment (e.g., within- and between-runs). Despite this, traditional LMMs are sensitive to the presence of outliers or other forms of contamination in the data on which they are modelled, and this can lead to inflated varying effect estimates [7]. Robust LMMs overcome this limitation by making no assumptions about the data's grouping structure and by weighting the varying effects and residuals to reduce the influence of outlying observations on the model's parameter estimates [7]. Here, we present the use of robust LMMs to model the variability in IS signal in routinely performed LC-MS/MS assays for the measurement of 25-hydroxyvitamin D_3/D_2 and prednisolone. Once fitted, these models were utilised to obtain robust, empirical acceptance ranges that properly account for system drift and within- and between-run variation in IS signal. Our approach, therefore, provides a robust, statistical tool for detecting IS outliers that we hope can be easily adopted by the clinical mass spectrometry community.

Materials and methods

Source data

Data from analytical runs from two in-house assays for the measurement of: (i) 25-hydroxyvitamin D_3 and D_2 (herein referred to as vitamin D), and (ii) prednisolone, were retrospectively collected (between November 2018 – September 2019 for vitamin D and August 2018 – September 2019 for prednisolone). Each run consisted of a mix of blanks, calibrators, QC material, and patient samples (Tables S1 and S2). The calibrators for the vitamin D assay were purchased from Chrom-Systems and consisted of pooled serum (product number: 38033). The calibrators for the prednisolone assay were prepared in-house in a solution of 0.1% bovine serum albumin dissolved in phosphate-buffered saline. The collected data consisted of runs that had been processed *via* the relevant standard operating procedure performed in routine clinical practice (i.e., examination of peak quality, peak integration, calculation of responses, calculation of analyte concentrations, etc. in Waters TargetLynx software) by appropriately trained and competency-assessed members of laboratory staff. The vitamin D samples were prepared in an automated fashion through the use of Freedom EVO 100 (Tecan, Reading, UK) and Biotage Extrahera (Biotage, Uppsala, Sweden) robotic platforms. In brief, this involved the automated addition of internal standard to each sample (150 μL), mixing for 5 min, and a 10 min equilibration at room temperature. The samples then underwent supported liquid extraction (SLE) with hexane on the Extrahera robot. The prednisolone samples were prepared through the manual pipetting of samples followed by manual addition of internal standard (60 μL) and a simple protein precipitation, as previously described [8]. The internal standards used for the vitamin D assay were: (i) 26,26,26,27,27,27-hexadeutero-25-hydroxyvitamin D_3 and (ii) 6,19,19-trideutero-25-hydroxyvitamin D_2 (QMX Laboratories Ltd, Thaxted, UK). The internal standard used for the prednisolone assay was 2,4,6,6,21,21-hexadeutero-prednisolone (Merck Millipore, Watford, UK). System suitability tests were performed prior to the start of each run in-line with ISO 15189:2012 standards against which our laboratory is accredited by the UK Accreditation Service (medical laboratory 8673). These involved the injection of purified analyte, dissolved in solvent, to ensure acceptable chromatography and adequate signal. The runs analysed here were each injected following a successful system suitability test.

Data analysis and statistical modelling

The collected data were manipulated and analysed within the R statistical computing environment (v3.6.1) using a combination of the *doParallel*, *dplyr*, *foreach*, *ggplot2*, *iterators*, *readxl*, *reshape2*, and *stringr* packages [9–17]. Robust linear mixed-effects models were fitted to the data through the use of the *robustlmm* package using the default settings [7]. Robust linear mixed-effects models were fitted to each analyte measured within the given assay (e.g., 25-hydroxyvitamin D_3 and D_2 for the vitamin D assay). The general equation describing the fitted models is shown in Equation 1, and summarised in Fig. 1, where y_{ij} is the percentage deviation of internal standard area from the mean of the calibrators for the j^{th} sample in the i^{th} run; β_0 is the constant effect of the average percentage deviation across all of the included runs; $\beta_1$is the constant effect of run position across all of the included runs; u_i is the varying effect for the i^{th} run (i.e., the random/varying intercept); x_{ij} is the run position at which the j^{th} sample within the i^{th} run was performed; and ε_{ij} is the residual error (i.e., sample-specific deviation within the run) for the j^{th} sample in the i^{th} run. These models were then utilised to obtain estimates of the sample-specific variation (σ^2_ε, the variance of ε_{ij}) and between-run variation of the intercepts (σ^2_u, the variance of u_i).

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \varepsilon_{ij} \quad (15)$$

Confidence intervals for the constant and varying effects and prediction intervals for the overall models were estimated through the use

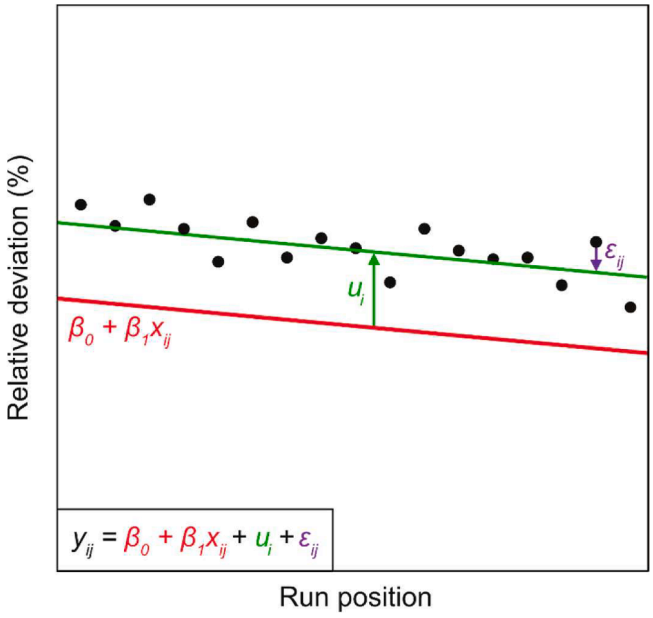

Fig. 1. A simplified schematic of the fitted linear mixed-effects models. Black points represent individual data-points for run *i*. The red, solid line indicates the overall trend (constant effects) across all *n* runs in an assay that presents with negative IS signal drift. The green, solid line indicates an example varying intercept for run *i*.

of non-parametric, cluster bootstrap resampling [18,19]. This was performed as follows:

1. Let the original dataset take the form of (y_{ij}, x_{ij}, z_i): where z_i represents the identifier for run i and y_{ij} represents the relative deviation for sample j within run i at position x_{ij}
2. Sample with replacement from the analytical runs $\{z_i\}$ to create the bootstrap sample $\{z_i^*\}$, where each z_i^* contains its original (y_{ij}, x_{ij}) values
3. Fit a robust linear mixed-effects model to the bootstrapped sample
4. Estimate all the parameters from the bootstrapped model ($\widehat{\beta}_0^*$, $\widehat{\beta}_1^*$, $\widehat{\sigma}_u^*$, and $\widehat{\sigma}_\varepsilon^*$)
5. Sample globally with replacement from the bootstrapped model's residuals $\{\varepsilon_{ij}^*\}$ to create the sample $\{\in_{ij}^*\}$
6. Predict new relative deviation values from the bootstrapped model: $y_{ij}^* = \widehat{\beta}_0^* + \widehat{\beta}_1^* x_{ij} + \widehat{u}_i^* + \in_{ij}^*$, where the addition of the term $\in_{ij}^*$ accounts for within-run variation
7. Repeat steps 2–6, B times (where $B = 999$)

The confidence intervals presented herein represent the 2.5th and 97.5th percentiles of the parameter estimates obtained from each bootstrapped model and the prediction intervals represent the 2.5th and 97.5th percentiles of the predicted values derived from each bootstrapped model (y_{ij}^*). The code used for the analyses described here can be found on GitHub (github.com/ed-wilkes/general-modelling).

Results

Data overview

We retrospectively collected data from our in-house vitamin D and prednisolone assays. A total of 105 and 55 analytical runs were collected for the vitamin D and prednisolone assays, respectively. Each run consisted of a consistent, pre-determined layout of calibrators, quality control (QC) material, and patient samples as dictated by the assays' standard operating procedures and as outlined in Tables S1 and S2. The vitamin D and prednisolone assays were prepared as described in the Materials and Methods (Section 2.1).

Modelling internal standard variability in a routine assay for the measurement of prednisolone

Each prednisolone run contained a variable number of patient samples, depending on the workload present within the laboratory at the time of assay preparation. Each assay did, however, contain a standardised number of calibrator and QC samples at the beginning of each run (Table S1). We first sought to determine the pattern of internal standard areas measured within each patient sample as a function of each samples' position in the run. In order to account for between-run variation in raw signal intensity, we expressed the IS areas as a percentage of the mean internal standard area of the calibrators within each run (herein referred to as "relative deviation", Fig. 2A). This demonstrated that the relative deviations for each sample varied linearly across the analytical run. It is important to note that this observed drift in IS signal did not impact the responses of QC materials of known

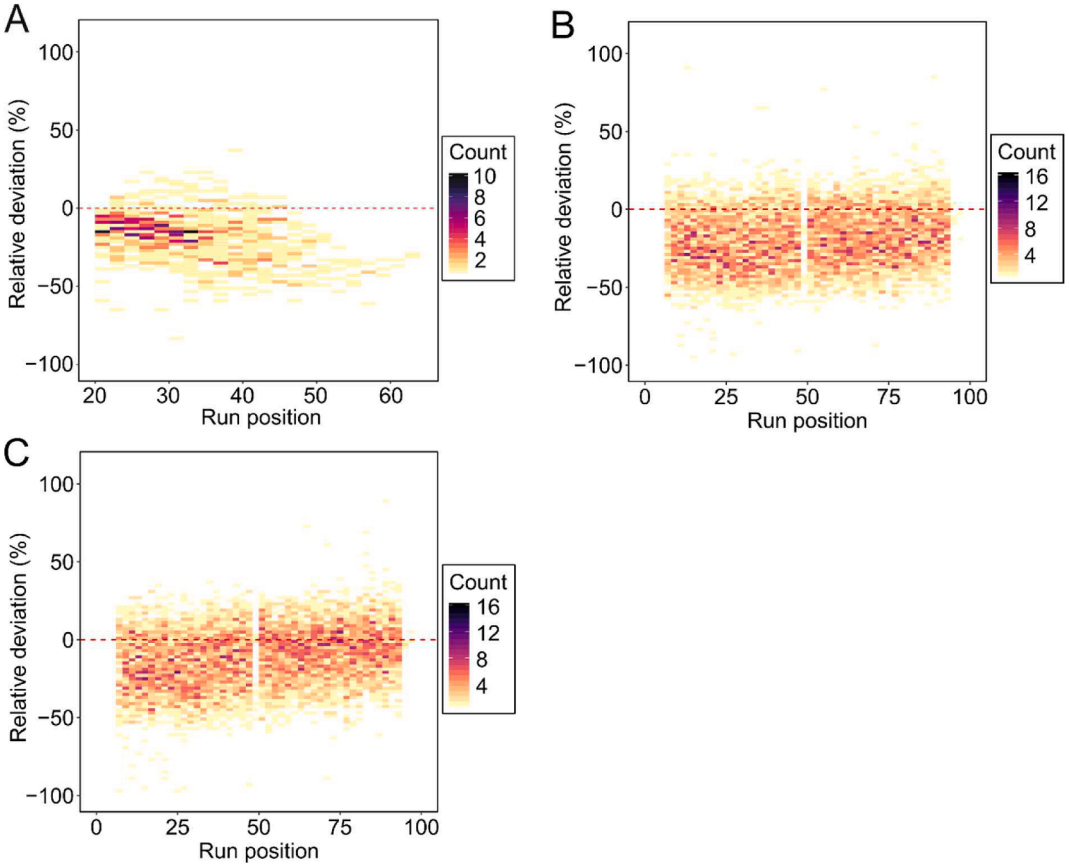

Fig. 2. Internal standard relative deviation changes linearly as a function of run position. Data points are binned and coloured as indicated in the associated scales. Dashed, red lines indicate zero relative deviation. (A) Data for the prednisolone assay. The x-axis is restricted as samples at the beginning of the run were standardised to consist of calibrators and QC material for each assay (Table S1). (B-C) Data for the 25-hydroxyvitamin D3 and D2 assays, respectively. Gaps at the beginning, middle, and end of the run represent the standardised positioning of calibrators and QC material (Table S2).

concentration injected at the beginning, middle, and end of the runs (Fig. S1), as the calculated concentrations showed no significant drift over the course of the run. To overcome the observed drift in IS signal, we hypothesised that an LMM would allow us to properly account for the effect of run position and incorporate the variance of IS signal within- and between-runs (Equation 1). Typical LMMs, however, assume that the residuals and random effects are normally distributed and do not contain outlying observations. The latter of these assumptions clearly does not hold in this experimental context, as the runs present within the data are likely to contain the outliers that we are seeking to identify. We did assume, however, that the majority of patient samples were unaffected by IS issues. We, therefore, fitted a robust LMM to the data in order to appropriately account for this contamination and used non-parametric cluster bootstrap resampling to estimate the confidence intervals of the fitted model's parameters [16,17] (see Materials and Methods for more information, Section 2.2). A simplified schematic of the fitted model is shown in Fig. 1 and the estimated parameters for this model are shown in Table 1. These data demonstrated that a statistically significant negative drift in IS relative deviation was apparent (Table 1, β_1 coefficient and associated confidence intervals). These models also demonstrated that between-run variability in IS signal contributed significantly more to the overall observed variance than within-run variability (Table 1, σ_u vs σ_ε). Diagnostic plots for the fitted model are shown in Fig. S2. These validated the rationale for using robust LMMs, as neither the residuals (Fig. S2B) nor the random effects (Fig. S2C) were normally distributed and both contained clear outlying observations. Distributions of the bootstrapped parameters for the model are shown in Fig. S3.

Modelling internal standard variability in a routine assay for the measurement of 25-hydroxyvitamin D

Each vitamin D run contained a maximum of 85 patient samples, with a fixed number of calibrator and QC samples. As with the prednisolone assay data shown in Fig. 2A, we first sought to determine the pattern of internal standard relative deviation measured within each patient sample as a function of each sample's position in the run (Fig. 2B-C). This demonstrated that the relative deviations for each sample also varied linearly across the analytical run; however, unlike prednisolone, the relative deviation trended upwards. To appropriately model this effect, we fitted robust LMMs to the data (Equation 1 and Fig. 1). The fitted parameters for each model are shown in Table 1. These data demonstrated that a statistically significant positive drift in IS relative deviation was apparent for both compounds' internal standards (Table 1, β_1 coefficients and associated confidence intervals). Diagnostic plots for the fitted models for both compounds are shown in Fig. S4. Distributions of the bootstrapped parameters for the models are shown in Figs. S5 and S6 for 25-hydroxyvitamin D_3 and D_2, respectively. In contrast to the models fitted to the prednisolone data (Table 1), the between- and within-run variation in IS signal contributed approximately equally to the overall observed variance (Table 1, σ_u and σ_ε).

Table 1
Estimated parameters from the fitted models for prednisolone, 25-hydroxyvitamin D3, and 25-hydroxyvitamin D2. Confidence intervals (95%) derived from non-parametric cluster bootstrapping are shown in parentheses. β_0, overall intercept; β_1, overall slope; σ_u, standard deviation of the random intercepts (between-run variation); σ_ε, standard deviation of the residuals (within-run variation).

Parameter	Prednisolone model	D_3 model	D_2 model
Intercept (β_0)	−11.6 (−15.6, −7.4)	−26.3 (−28.5, −24.2)	−21.4 (−23.5, −19.3)
Slope (β_1)	−0.164 (−0.297, −0.036)	0.118 (0.098, 0.138)	0.192 (0.168, 0.215)
Between-run variation (σ_u)	10.1 (7.3, 13.0)	12.9 (11.4, 14.3)	11.3 (10.0, 12.5)
Within-run, residual variation (σ_ε)	6.10 (5.49, 6.69)	12.0 (11.4, 12.5)	13.0 (12.4, 13.5)

Using the fitted models in routine practice for detecting internal standard outliers

A visual representation of the models fitted to the vitamin D and prednisolone data is shown in Fig. 3. These visualisations served to demonstrate the overall positive and negative drifts in IS signal captured by the models for the vitamin D and prednisolone assays, respectively (Fig. 3, solid red lines). In order to derive acceptance ranges from each model for assessing future assay runs in routine practice, we reasoned that the relative deviation for a typical sample within a typical run should fall within the 95% prediction interval of the relevant fitted model. As such, we calculated the prediction intervals for each model by means of non-parametric, cluster bootstrap resampling. The resulting derived acceptance ranges for each model are shown as dashed purple lines in Fig. 3A-C. As expected, using these derived acceptance ranges retrospectively on the original data resulted in approximately 5% of the samples being deemed as outliers (5.7%, 4.2%, and 4.3% for the prednisolone, 25-hydroxyvitamin D_3, and 25-hydroxyvitamin D_2 assays, respectively). Taken together, these data serve to demonstrate how this modelling approach is applicable to different assays with alternate patterns of IS signal drift, with the vitamin D and prednisolone assays demonstrating different directions of systematic signal drift. In addition, these data demonstrate how these models can be used to derive acceptance ranges that appropriately and robustly account for IS signal drift that can be used for future assay runs.

Discussion and conclusions

Heavy element-labelled internal standards are useful tools for the correction of stochastic variation in sample extraction processes and mass spectrometric analysis. It is critically important, however, to be able to identify outliers in IS signal that may represent analytical errors (e.g., double-spiking or missed addition of internal standard) or ion suppression/enhancement [1,2]. Despite this, defined acceptance criteria for the detection of internal standard outliers in clinical mass spectrometry assays are lacking and arbitrary criteria that disregard a sample's position within a run do not account for potential system drift during an assay [1]. We, therefore, sought to develop empirical and robust acceptance criteria for internal standard signal in two of our routinely performed targeted LC-MS/MS assays.

We retrospectively collected data from two of our routinely performed assays; one for the measurement of prednisolone and other for the simultaneous measurement of 25-hydroxyvitamin D_3 and D_2. Initial visualisation of the data revealed a linear relationship between the position of the samples with the assay runs and the relative deviation of the IS signal (Fig. 2A-C). We hypothesise that the drift observed was the result of a build-up of mobile phase/matrix constituents on the source and/or changes in room temperature over the course of the runs, as most of the runs are processed overnight and, thus, there is less traffic within the laboratory. Given the drift present in the data, the use of fixed, arbitrary thresholds would lead to false positive and negative identifications of IS outliers depending on the direction of drift in IS signal and the sample's position in the run. To solve this problem, we fitted robust LMMs to the data in order to appropriately model IS signal drift and the between- and within-run variation in IS relative deviation (Table 1).

Throughout the analyses presented here, we have made the assumption that the majority of patient samples present in our data are not outliers and have not been subject to sampling errors or ion suppression/enhancement. It is likely, however, that the data are contaminated with samples affected by these issues. By using robust LMMs, however, we have reduced the impact of these on the estimates of the

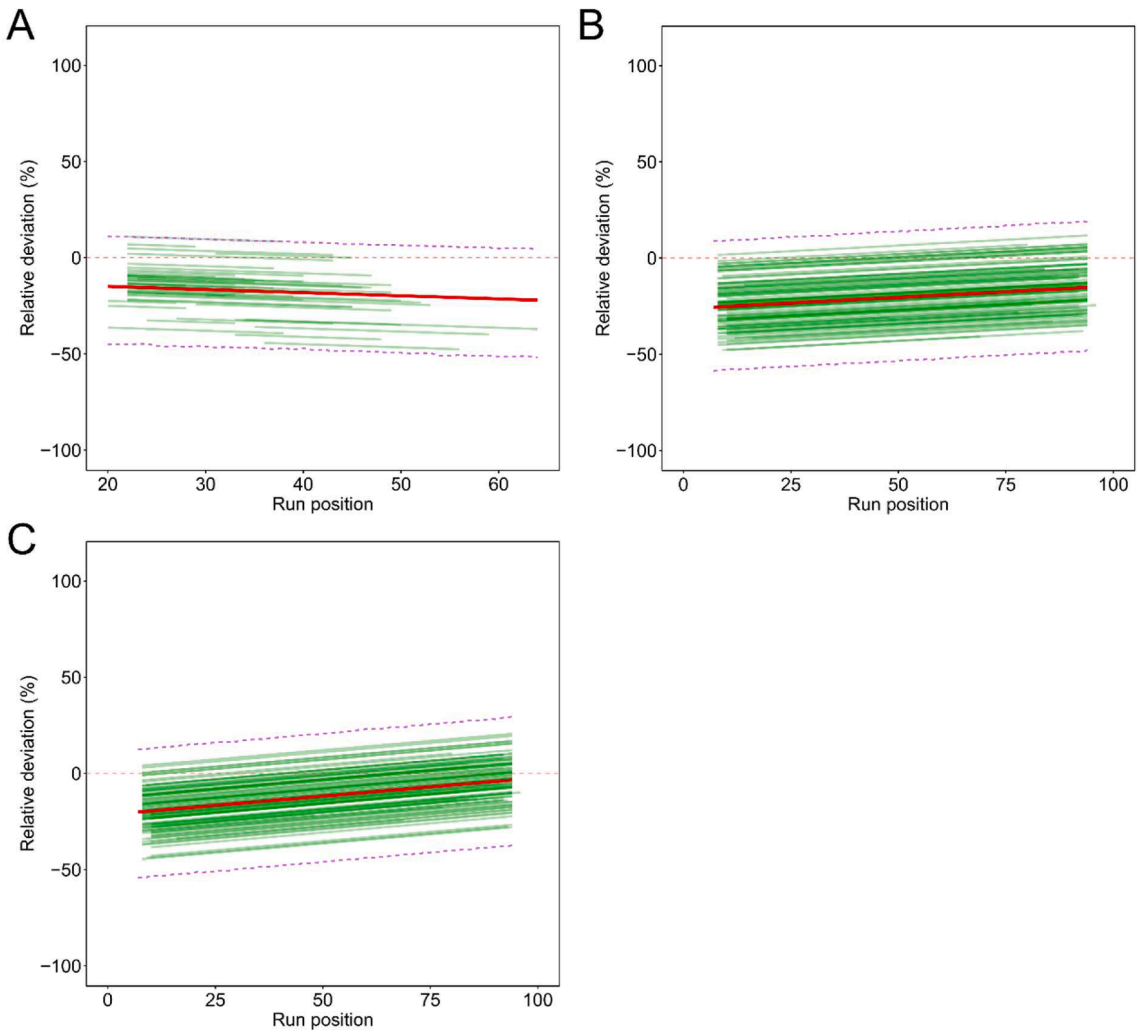

Fig. 3. Fitted models and derived acceptance ranges for the vitamin D and prednisolone assays. (A) The model fitted for prednisolone. (B) The model fitted for 25-hydroxyvitamin D_3. (C) The model fitted for 25-hydroxyvitamin D_2. Dashed, red lines indicate zero relative deviation. The red lines indicate the constant effects fitted to the runs ($\beta_0 + \beta_1 x_{ij}$). Green lines indicate the varying intercepts fitted to each individual run (u_i). Dashed, purple lines indicate the derived prediction intervals.

models' parameters and derived acceptance ranges. An additional limitation of the analyses presented here is that, as we have modelled the relative deviation values of each IS, the developed models are dependent on the matrix of the calibrators. Should this change significantly (e.g., changes in lot number or supplier changes), the relationship between run position and relative deviation should be reassessed and the models re-fitted. This should not, however, represent a significant obstacle to laboratories that already perform verifications of new reagent or QC material lots on a regular basis; as is required for laboratory accreditation to international standards (e.g., ISO 15189:2012).

The data presented here demonstrate that robust LMMs can be utilised to develop empirical, robust acceptance ranges for internal standard signals using data that are abundantly available to laboratories that routinely perform these types of analyses. The code used to perform these analyses is publicly available on GitHub (github.com/ed-wilkes/general-modelling) and can be run within the open source and freely available R statistical computing environment. The code can also be used to fit more simple varying (random) effects models (ignoring the effect of run position) to develop acceptance ranges for assays where system drift is absent (i.e., where β_1 is not significant). Indeed, this modelling approach could be used during assay development to determine if a significant drift is present in order to trigger further investigations as to its cause. Taken together, our analyses provide a freely available method by which any laboratory performing targeted LC-MS/MS analysis can more robustly identify internal standard outliers.

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.

Acknowledgements

We would like to acknowledge both past and present members of our laboratory for the preparation and processing of the assays presented here. In addition, we wish to acknowledge the patients whose samples are represented in these data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmsacl.2021.06.001.

References

[1] M. Vogeser, C. Seger, Pitfalls associated with the use of liquid chromatography-tandem mass spectrometry in the clinical laboratory, Clin. Chem. 56 (8) (2010) 1234–1244.

[2] S. Wang, M. Cyronak, E. Yang, Does a stable isotopically labelled internal standard always correct analyte response? A matrix effect study on a LC/MS/MS method for the determination of carvedilol enantiomers in human plasma, J. Pharm. Biomed. Anal. 43 (2) (2007) 701–707.

[3] M.H. Buonarati, D. Schoener, Investigations beyond standard operating procedure on internal standard response, Bioanalysis. 11 (18) (2019) 1669–1678.

[4] S. White, N. Adcock, W. Elbast, J. Hiller, M. Knutsson, B. Lausecker, J. Long, S. McDougall, F. Salavert, T. Sangster, T. Verhaeghe, P.V. Amsterdam, European Bioanalysis Forum: recommendation for dealing with internal standard variability, Bioanalysis. 6 (20) (2014) 2767–2774.

[5] D. Bates, M. Maechler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4, J. Stat. Softw. 67 (1) (2015) 1–48.

[6] A. Gelman, Analysis of variance – why it is more important than ever, Ann. Stat. 33 (1) (2005) 1–53.

[7] M. Koller, robustlmm: An R package for robust estimation of linear mixed-effects models, J. Stat. Softw. 75 (6) (2016) 1–24.

[8] E.L. Williams, S. Choudhury, T. Tan, K. Meeran, Prednisolone replacement therapy mimics the circadian rhythm more closely than other glucocorticoids, J. Appl. Lab Med. 1 (2) (2016) 152–161.

[9] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. https://www.R-project.org.

[10] Microsoft Corporation, Weston S. doParallel: Foreach parallel adaptor for the 'parallel' package. R package version 1.0.16. 2020. https://CRAN.R-project.org/package=doParallel.

[11] H. Wickham, R. François, L. Henry, K. Müller. dplyr: A grammar of data manipulation. R package version 1.0.0. 2020. https://CRAN.R-project.org/package=dplyr.

[12] Microsoft Corporation, Weston S. foreach: Provides foreach looping construct. R package version 1.5.1. 2020. https://CRAN.R-project.org/package=foreach.

[13] H. Wickham. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York. 2016. https://ggplot2.tidyverse.org.

[14] Revolution Analytics, Weston S. iterators: Provides iterator construct. R package version 1.0.13. 2020. https://CRAN.R-project.org/package=iterators.

[15] H. Wickham, J. Bryan, readxl: Read excel files. R package version 1.3.1. 2020. https://CRAN.R-porject.org/package=readxl.

[16] H. Wickham, Reshaping data with the reshape package, J. Stat. Softw. 21 (12) (2007) 1–20.

[17] H. Wickham, stringr: Simple, consistent wrappers for common string operations.. R package version 1.4.0. 2020. https://CRAN.R-project.org/package=stringr.

[18] S. Ren, H. Lai, W. Tong, M. Aminzadeh, X. Hou, S. Lai, Nonparametric bootstrapping for hierarchical data, J. Appl. Stat. 37 (9) (2010) 1487–1498.

[19] H.-T. Thai, F. Mentré, N.H.G. Holford, C. Veyrat-Follet, E. Comets, A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models, Pharm Stat. 12 (3) (2013) 129–140.