mzQuality: an open-source software tool for quality monitoring and reporting of targeted mass spectrometry measurements

Peet, M.P. van der; Maas, P.; Wegrzyn, A.; Lamont, L.; Fleming, R.; Bordes, C.; ... ; Kindt, A.S.D.

ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Article

mzQuality: An Open-Source Software Tool for Quality Monitoring and Reporting of Targeted Mass Spectrometry Measurements

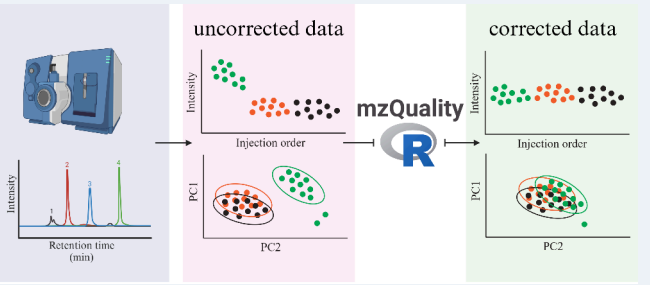
Marielle van der Peet, Pascal Maas, Agnieszka Wegrzyn, Lieke Lamont, Ronan Fleming, Constance Bordes, Stéphanie Debette, Amy Harms, Thomas Hankemeier, and Alida Kindt*

Cite This: J. Am. Soc. Mass Spectrom. 2025, 36, 1669−1676

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Analyzing metabolites using mass spectrometry provides valuable insight into an individual's health or disease status. However, various sources of experimental variation can be introduced during sample handling, preparation, and measurement, which can negatively affect the data. Quality assurance and quality control practices are essential to ensuring accurate and reproducible metabolomics data. These practices include measuring reference samples to monitor instrument stability, blank samples to evaluate the background signal, and strategies to correct for changes in instrumental performance. In this context, we introduce mzQuality, a user-friendly, open-source R-Shiny app designed to assess and correct technical variations in mass spectrometry-based metabolomics data. It processes peak-integrated data independently of vendor software and provides essential quality control features, including batch correction, outlier detection, and background signal assessment, and it visualizes trends in signal or retention time. We demonstrate its functionality using a data set of 419 samples measured across six batches, including quality control samples. mzQuality visualizes data through sample plots, PCA plots, and violin plots, which illustrate its ability to reduce the effect of experiment variation. Compound quality is further assessed by evaluating the relative standard deviation of quality control samples and the background signal from blank samples. Based on these quality metrics, compounds are classified into confidence levels. mzQuality provides an accessible solution to improve the data quality without requiring prior programming skills. Its customizable settings integrate seamlessly into research workflows, enhancing the accuracy and reproducibility of the metabolomics data. Additionally, with an R-compatible output, the data are ready for statistical analysis and biological interpretation.



KEYWORDS: *Metabolomics, Mass spectrometry, Data quality, R package, R-shiny app*

1. INTRODUCTION

Metabolites are influenced by both endogenous factors and an individual's immediate environment, such as diet, living arrangements, or other, more general lifestyle factors. All of these metabolites are reflected in the metabolome.[1] Analyzing the metabolome, also known as metabolomics, can provide valuable information on an individual's health status and help elucidate disease-related metabolic phenotypes. This can be of clinical interest when investigating disease progression, possible treatment outcomes, or even biomarkers for early detection and early intervention.[2]

Targeted Mass Spectrometry (MS)-based analytical techniques are often used to detect and quantify hundreds of metabolites simultaneously. Mass spectrometers are generally combined with separation techniques, such as liquid chromatography (LC), due to their high sensitivity and selectivity.[3] However, in practice, it can be challenging to acquire high-quality MS-based metabolomics data due to various sources of experimental variation. Sources of variation can be introduced during sample handling, sample preparation, and/or varying experimental conditions, such as differences in chromatographic mobile phases, column temperature, column pressure variation, instrumentation, instrumental drift, and even regular maintenance.[4,5] It can be challenging to keep these conditions stable in larger metabolomics studies where samples are divided, prepared, and measured over multiple subsets, termed batches, that are analyzed individually.[6,7]

All of these experimental variations can collectively contribute to less reliable and reproducible results, which makes it harder to interpret the data properly and draw meaningful conclusions from the findings. To minimize variation as much as possible, it is important to have quality assurance (QA) guidelines and corresponding quality control (QC) practices to evaluate and

Received: March 12, 2025
Revised: July 11, 2025
Accepted: July 16, 2025
Published: July 25, 2025

correct data variability.[8,9] These guidelines should include protocols for the study design, sample collection and storage, sample preparation, and sample measurements.[9] An important part of QC practices is the use of QC samples, which can consist of a long-term reference material (LQC) or short-term reference material (SQC). LQCs are used over extended periods and function as a reference over time to monitor the data quality throughout different studies. SQCs, also named intrastudy QC samples, often consist of pooled study samples that contain an average amount of metabolites representative of the study. Additional QC practices should include a systems suitability test (SST) to assess analytical stability, measurements of blanks to evaluate background signal, and the use of internal standards (ISs) for normalization strategies.[10]

Even when variation is minimized as much as possible, it is still important to assess the data thoroughly. Several open-source data processing tools, such as OpenMS[11] and MZmine,[12] offer features like baseline noise filtering, retention time alignment, and a few quality metrics. However, these tools do not address the variations in sample measurements over extended periods. On the other hand, the msQuality package by Naake et al. focuses more on quality metrics but does not automatically correct the identified issues.[13]

As the field of metabolomics continues to evolve, the integration of innovative software solutions becomes increasingly important to unlock the full potential of the field and contribute to more reliable findings. In this article, we introduce mzQuality, a new open-source metabolomics data quality assessment tool. Unlike other tools that process peaks, mzQuality distinguishes itself by providing a user-friendly interface for evaluating and correcting technical variations in the data without requiring extensive programming skills. Recognizing the importance of experimental design and procedural quality in metabolomic studies, we included guidelines as to how to implement mzQuality into a metabolomics workflow and optimize the software's application.[9] mzQuality performs several important quality control steps, such as identifying sample outliers, calculating background signal, performing batch correction, detecting trends in total signal intensity for either metabolite or IS, and calculating metrics that highlight variations in metabolite/IS ratios. In addition, compounds with a low signal-to-noise ratio are flagged for removal by the user. mzQuality uses repeat measurements of quality control samples to flag compounds that do not reach predefined thresholds as low quality. We defined the threshold at a relative standard deviation of QC > 30%, which follows the guidelines of the metabolomics quality assurance and quality control consortium,[14] and the background signal cannot be higher than 40%.[15] Furthermore, mzQuality allows users to adjust settings according to their preferences and analytical methods. How these assessment tasks are performed is showcased here using a large example data set included in GitHub: https://github.com/hankemeierlab/mzQuality, consisting of 419 study samples analyzed over six batches. Through this paper, we showcase how mzQuality identifies and corrects variations within the data, illustrating its functionality and effectiveness.

2. METHODS

2.1. Software Information. mzQuality is distributed as an R package for which version 4.0 or later is required. It is available through GitHub: https://github.com/hankemeierlab/mzQuality. For additional information on installation and usage, see the manual on GitHub.

2.2. Example Data Set. An example data set obtained in our laboratory showcases the utility and effects of mzQuality. This data set comprises 6 random yet consecutively measured batches containing 419 EDTA plasma study samples originally part of the CoSTREAM consortium (https://www.eibir.org/projects/costream/). Also included in the batches were different types of QC samples, blank samples, and calibration curves. The samples were prepared and measured according to the method described by Yang. et al.[16] Briefly, a targeted LC-MS method is operated in polarity switching mode where all metabolites and the corresponding stable isotope labeled IS are analyzed in dynamic Multiple Reaction Monitoring (dMRM) mode. The data were integrated using SCIEX OS (version 2.1.6.59781) and exported using the recommended data format.

3. RESULTS AND DISCUSSION

Here, we discuss several steps needed for or performed by mzQuality: (1) the recommended batch design including the different sample types, (2) the data input format, (3) the calculations performed, (4) the generated output plots used for assessment, and (5) the report format. The example data set was used to showcase these steps.

3.1. Batch Design. Different sample groups or phenotypes should be randomly distributed across batches, ensuring a balanced representation of different phenotypes within the study. If phenotypes are not balanced well, batch-related differences might introduce errors in downstream data analysis.[17]

The recommended order of sample injections begins with a system suitability test (SST), highlighted in orange, followed by one or more blank injections, as shown in Figure 1. SST samples are used to ensure the analytical system functions within predefined criteria to detect any potential contamination issues.[8] Blank samples, consisting of water but prepared without the addition of IS, are injected next. Following this, procedure blank samples (PROCs), which are prepared with IS are injected; these are highlighted in light blue. It is important to inject blank samples at both the beginning and the end of each batch, with a minimum of two injections. These samples can be used to assess carryover and background signals in the measured data.

Although a calibration curve is not mandatory for the tool's functionality, it can be incorporated into the batch design. Calibration curves can be made by adding an increasing concentration of an analytical standard mix to water (ACAL) or a biological matrix (CAL). While mzQuality does not use the calibration lines for corrections, it can calculate absolute concentrations and inform about the dynamic range in which samples are measured.

It is important to consider whether to use LQC, SQC, or a combination of both in the batch design, as these samples are used for batch correction and assessment of technical variation, as assessed by the RSDqc per compound. If SQCs were generated from pooling study samples, then they provide an average signal over the study samples, therefore allowing a direct comparison with the samples investigated with the biological samples. For example, if any technical drift were to occur, the impact on SQC samples would be most representative of the actual response in the study samples. This makes batch corrections based on SQC samples the most reliable.[8]

LQCs provide information about stability over time, allowing us to monitor instrument stability and sensitivity beyond the

https://doi.org/10.1021/jasms.5c00073
J. Am. Soc. Mass Spectrom. 2025, 36, 1669−1676

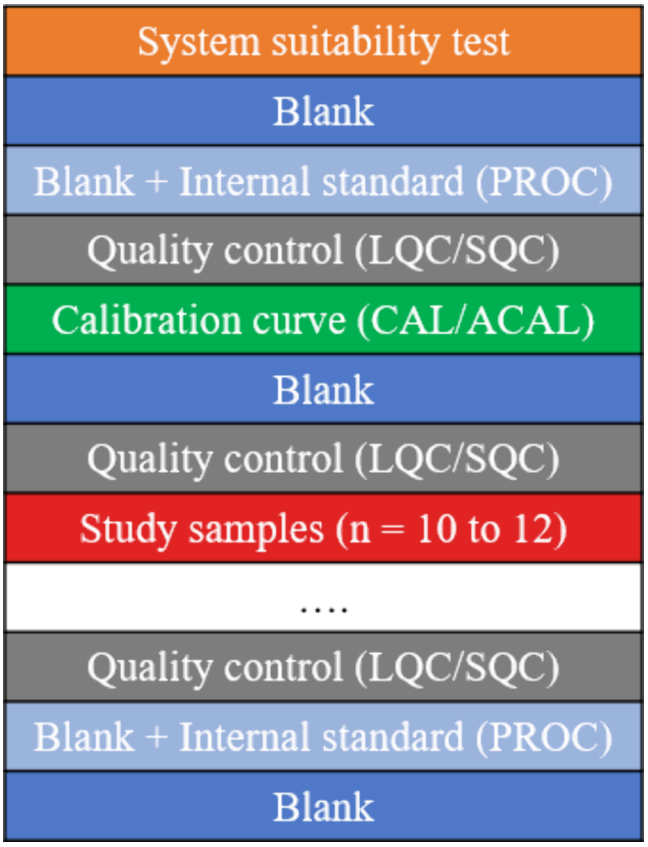


Figure 1. Batch design. The figure shows a summarized batch design starting with a system suitability test. Quality control samples must be dispersed throughout the whole batch and at least every 12 injections. The quality control and study sample block can be repeated.

duration of one study and correct for batch effects when a project was measured over a long period. When a compound is not detected in LQC or SQC samples, the compound will be assumed to not be present and will not be reported by mzQuality.

The recommendation is to have a minimum of four QC samples per batch and, depending on the batch size, for them to be evenly spaced throughout the batch. We recommend one injection for every 15 study samples (highlighted in red). The injections of QC samples must be evenly distributed throughout the whole batch to check for instrument and extraction variance.

Replicates of study samples are treated independently, meaning that no average of these samples is calculated or utilized for any analysis or correction. However, Rosner's test is used on the IS of study samples to identify and flag potential misinjections and, thus, outliers. The complete batch design for the example data set can be found in Table S1.

3.2. Data Input and Data Validation. Chromatographic peak areas can be integrated by using open-source or vendor software. The peak areas of compounds and IS, if applicable, need to be exported for use in mzQuality. We strongly recommend using a tab-separated format while exporting; an example can be found in GitHub: https://github.com/hankemeierlab/mzQuality. Additionally, it is important that numbers and symbols are compatible with R. The input file should include the following columns with the specified headers (Table 1): aliquot, type, batch, compound, area, and datetime of the sample measurement. For quantitative or semiquantitative analysis, the columns compound_is and area_is must also be included. Additional columns can be added for visualization, such as replicate number or retention time; however, this information is not used for corrections.

Table 1. Columns Included in the Input File

Column name	Info included
aliquot	Unique sample names
type	Sample type; SQC, LQC, BLANK, PROC, SAMPLE, CAL..., ACAL..., SST
batch	Batch number
compound	Compound name
area	The integrated peak area of the compound. These cells must have a numerical value or NA.
compound_is	Internal standard name
area_is	The integrated peak area of the internal standard. These cells must have a numerical value or NA.
datetime	The time the samples have been measured. The format of the date/time column should be year-month-day and hours:minutes:seconds, separated by a space.

mzQuality performs a validation on the input file to ensure all necessary information is included and to detect any formatting errors. It checks for the presence of the required columns and verifies whether a batch is assigned to all aliquots. If no batches are defined, all samples are assigned to batch 1.

During the validation process, mzQuality checks if the column "type" includes the sample type QC. Additionally, it verifies if there is an observation for every aliquot and every compound and inserts "NA" for any missing data. Any NA value is treated as an absent data point and will not be used for the calculations. Finally, the injection order is checked based on the datetime column. If any of the necessary columns are absent, mzQuality will not be able to run.

3.3. Data Processing. Once the data are imported and validated, mzQuality initiates the first steps of data evaluation to ensure data integrity.

3.3.1. IS Normalization. The first step involves IS normalization, where for all samples ratios are calculated by dividing the compound area (Peak area_compound) by the corresponding IS area (Peak area_IS) as shown in eq 1.

$$\text{Ratio} = \frac{\text{Peak area\_compound}}{\text{Peak area\_IS}} \quad (1)$$

ISs can be predefined in the input file and should be selected based on class behavior and retention times. mzQuality also provides suggestions for alternative ISs, but the user must check the compatibility of the IS with the compound. These suggestions are based on calculations that evaluate all compound and IS combinations to determine which combination gives the best RSDqc after batch correction.

If no IS is available, mzQuality assumes an IS area of 1, meaning that the area ratio, on which most of the calculations are based, is the same as the area.

3.3.2. Sample Outlier Determination. The next steps involve the identification of QC and study sample outliers using Rosner's test. For QC samples, outliers are identified based on the median area/area_IS ratios, as these samples are assumed to be identical, while for study samples, only the IS areas are used since they may vary biologically. Potential outliers are flagged by mzQuality and are automatically excluded from further calculations.[18] For study samples, the user needs to determine whether an outlier is biological or technical and whether it should be included in further analysis.

https://doi.org/10.1021/jasms.5c00073
J. Am. Soc. Mass Spectrom. 2025, 36, 1669−1676

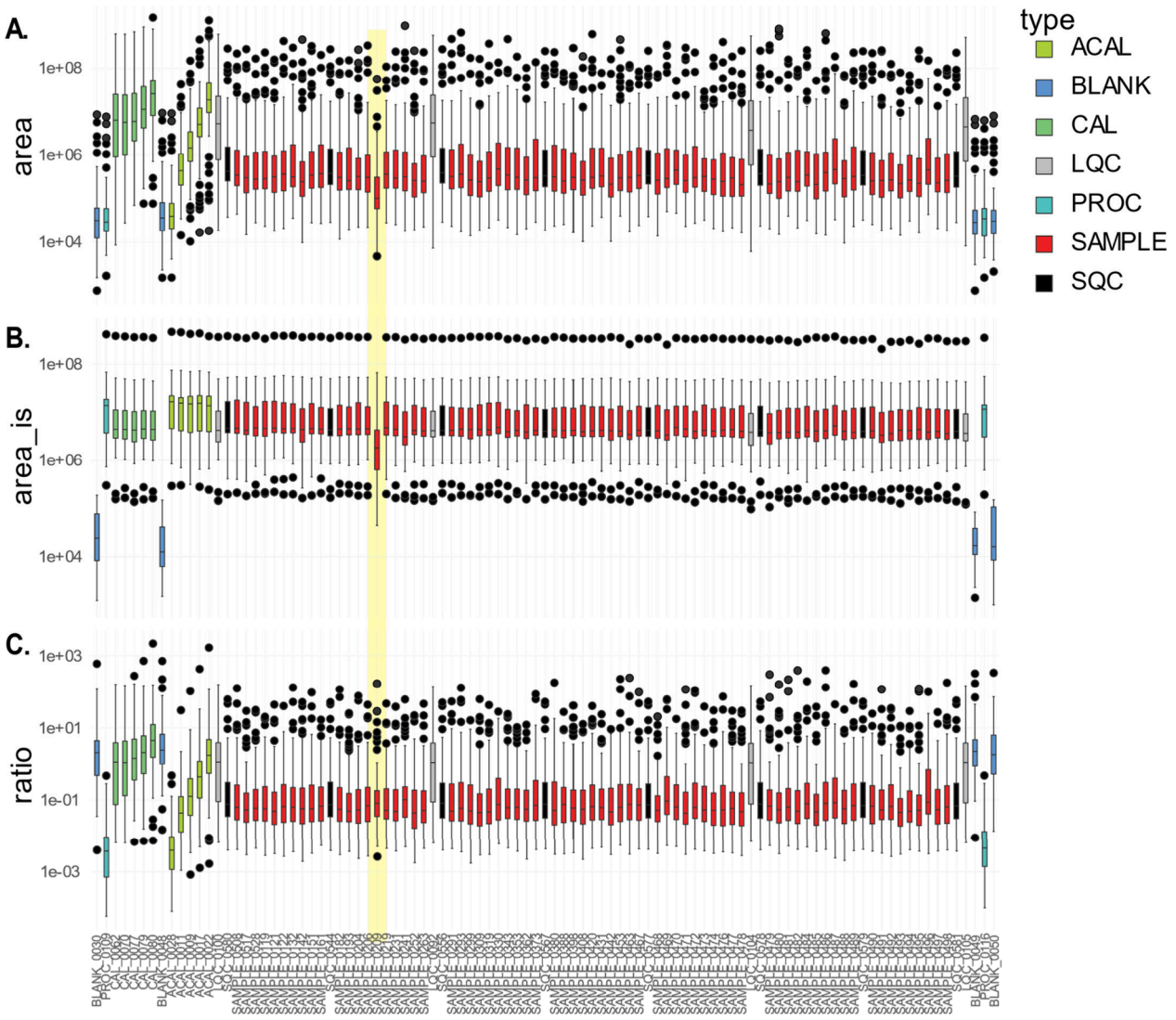


Figure 2. The aliquot plot shows successful internal standard correction of batch one. Every box-whisker plot is color-coded per sample type: calibration mix in water (ACAL; light green), BLANK (dark blue), calibration mix in matrix (CAL; dark green), long-term quality control (LQC; gray), procedure blank (PROC; light blue), SAMPLE (red), and short-term quality control (SQC; black). Black dots represent outlier compounds within the samples. Panel A shows the median of all metabolite areas per sample; panel B shows the median of all the internal standards per sample, and panel C shows the median ratios (area/area_IS) per sample. Sample 0209 is highlighted in yellow due to its relatively low area compared to the other samples. Panel C shows that the variation in sample 0209 is corrected by the internal standard use.

3.3.3. Batch Correction. To correct for technical variation between individual batches by, e.g., sample preparation, mobile phases, column changes, instrument cleaning, or changes in sensitivity of the system over time, a between-batch correction is performed. Whenever more than one batch is present, a batch correction is performed to minimize batch-to-batch variation in the data, by calculating the median ratio of every compound in all the QC samples per batch as described in eq 2. These batch-to-batch compound-specific correction factors are calculated using eq 2 and multiplied by the area ratios of each compound, as shown in eq 3.

$$f_{\text{correction\_batch\_}n} = \frac{\text{Ratio\_QC\_median\_all\_batches}}{\text{Ratio\_QC\_median\_batch\_}n} \quad (2)$$

$$\text{Batch corrected area ratio} = f_{\text{correction\_batch\_}n} \times \text{area ratio} \quad (3)$$

When trends occur within batches, such as a decrease in sensitivity, mzQuality can correct for this using a within-batch correction. The within-batch correction is performed using a first-order regression.[19]

3.3.4. Compound Quality. Once all of the batch corrections are performed, various metrics are calculated to check the quality of compounds. The RSDqc is calculated using the QC samples both before and after batch correction to evaluate the stability of a compound during measurements. Additionally, the percentage of QC samples in which each compound is detected is calculated to improve confidence in the identified compounds. The RSDqc is calculated per compound according to eq 4. The standard deviation (SD) of the corrected ratios of all QC samples from all batches (SD_corrected_ratio_QC_all_batches) is divided by the mean of all of these ratios (Mean_corrected_ratio_QC_all_batches).

https://doi.org/10.1021/jasms.5c00073
J. Am. Soc. Mass Spectrom. 2025, 36, 1669−1676

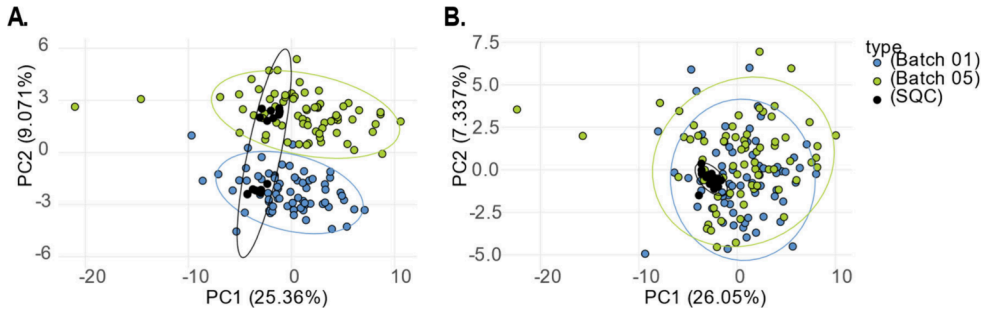


Figure 3. PCA plot before and after between batch correction. (A) PCA plots of batches 1 and 5 before correction. (B) PCA plot of batches 1 and 5 after correction. The samples are grouped by batch and show a 95% confidence interval. After batch correction, the short-term quality control (SQC) samples are aligned in the middle of the two batches.

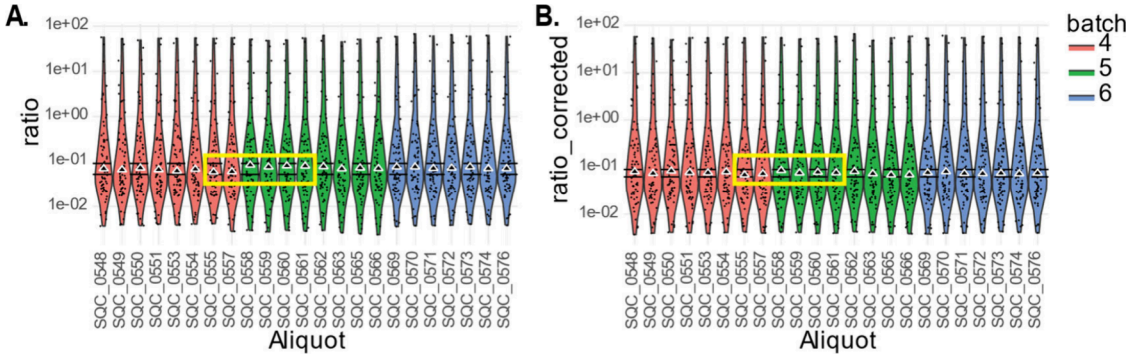


Figure 4. Violin plot of batch 04−06 before and after correction. (A) Violin plot of quality control samples from batch 04−06 before batch correction. (B) Violin plot of quality control samples from batch 04−06 after batch correction. Short-term quality control (SQC) samples 0555, 0557, 0558, 0559, 0560, and 0561 are highlighted to demonstrate the between-batch correction performed by mzQuality.

$$\mathrm{RSD_{qc}} = \frac{\mathrm{SD\_corrected\_ratio\_QC\_all\_batches}}{\mathrm{Mean\_corrected\_ratio\_QC\_all\_batches}} \times 100 \quad (4)$$

Blank samples are used to assess the background signal. The background signal is calculated according to eq 5 by dividing the mean of all the areas in all blank samples (mean_peak area_blanks) by the median of all the areas in study samples (median_peak area_study samples).

$$\text{Background signal} = \frac{\text{Mean\_peak area\_blanks}}{\text{Median\_peak area\_study samples}} \times 100 \quad (5)$$

Based on the recommendations of the metabolomics quality assurance and quality control consortium, we set the threshold for the RSDqc at 30%.[14] For the background signal, we recommend that the value of the e-score should not exceed 40%. However, mzQuality allows users to adjust these settings according to their preferences and analytical methods.

3.4. Output Plots. The example data set passed the initial integrity and validation tests, and the calculations were performed as described. The following mzQuality output is provided to demonstrate the visualization of data useful in evaluating data quality.

3.4.1. Aliquot Plot. The initial step of the visual assessment of the data is to inspect the aliquot plot per batch. Figure 2 shows the aliquot plot, where the horizontal axis shows the samples in injection order. Panel A shows the median area of all compounds; panel B shows the area_IS and panel C shows the area ratio, all plotted against the injection order of the samples. This plot portrays all of the samples as boxes in a box-whisker plot, color-coded per sample type, and offers insights into the distribution of all measured compounds per sample. In panels A and C, the CAL samples and ACAL samples follow an upward trend, consistent with an expected pattern for a calibration curve with increasing concentrations. The boxplots from QC samples of the same type display similar areas. In panel C, the PROC samples show low values and the blank samples are high. The ratio of PROC is expected to be low due to the absence of compounds (area), but there are internal standards present. Conversely, the high values in the blank samples result from the lack of internal standards. Particularly, the aliquot plot can be used to assess the data before any correction is performed and to check for instrumental drift, missing samples, misinjections, or other mishaps such as incomplete batch export from the raw data. Instrumental drift visible in the area (panel A) should not be visible in the ratio plot (panel C). However, extreme instrumental drift may not be corrected or only affect individual compounds, so within-batch correction should then be applied. Figure 2 highlights sample 209 in yellow, showing a notably lower area compared with the other samples in panel A. This could be attributed to either biological or technical issues. Panel C shows the variation of the sample. 209 can be corrected by using IS, making the variation most likely from a technical source and not a biological one.

3.4.2. PCA Plot. A PCA plot can be used to (1) check how representative the QC samples are to the study, (2) visualize possible batch effects, and (3) identify potential sample outliers. Ideally, SQC samples should be centered, as they are aliquots of a pool of all study samples and thus represent their average. Figure 3A shows that, before batch correction, the samples and SQC samples from the two batches were not aligned. Still, after

https://doi.org/10.1021/jasms.5c00073
J. Am. Soc. Mass Spectrom. 2025, 36, 1669−1676

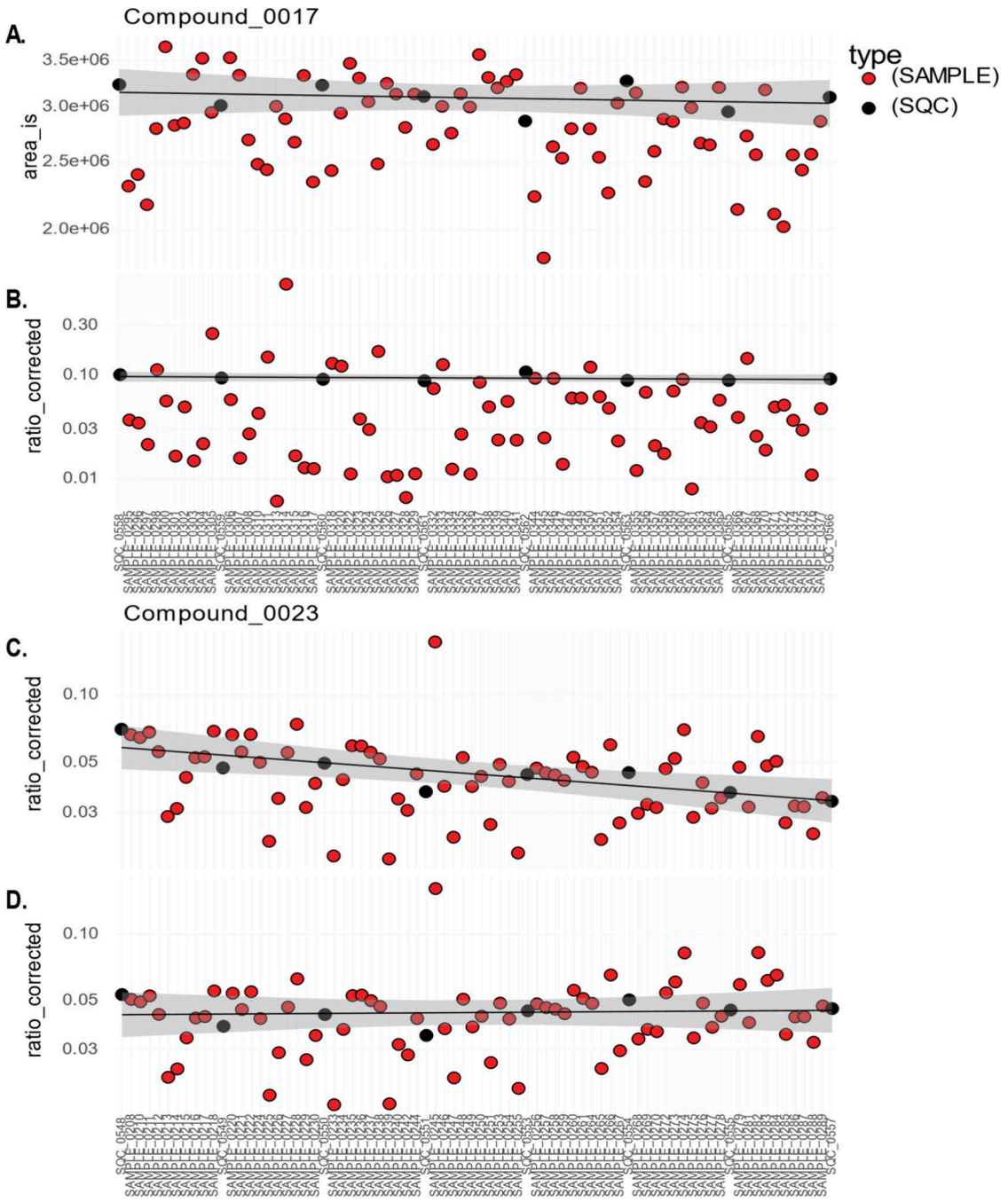


Figure 5. Individual compound plots: the *y*-axis shows the $\log_2$ transformed corrected ratios, while the *x*-axis shows the injection number. Red dots portray the samples, while black dots portray the short-term quality control (SQC). (A) The area_IS of the internal standard related to compound 17 in batch 5. (B) Example of a compound without any trends in batch 5. (C) Example of a downward trend within one batch. (D) Shows how this trend is corrected with the within-batch correction functionality of mzQuality in batch 4.

batch correction (Figure 3B), the SQC samples are aligned and mostly centered within the study samples, with the two confidence interval (CI) ellipses overlapping. To reduce the complexity of the PCA plot, only two batches are shown, while the PCA plots of all batches can be found in Supplementary Figure 1.

Outlier samples can be detected by identifying those that fall far outside the 95% CI ellipse. These outliers should be flagged and investigated for deviations in compound peak areas or deviations from standard experimental procedures. If the peaks are normal and no technical explanation is found, the sample should not be removed from the data set, as the observed variation is most likely due to biological differences.

3.4.3. QC Violin Plot. A violin plot displays QC samples that were not removed in the previously outlined steps, where each dot represents a compound. The plot can be used to investigate QC outliers and deviating batches. Figure 4 illustrates the QC data of three batches before and after the between-batch correction. Assessment of the plot involves (1) examining the shape of the violin plots, which should be consistent over the batches, and (2) investigating the median of the QCs, which should all align on a horizontal line.

https://doi.org/10.1021/jasms.5c00073
J. Am. Soc. Mass Spectrom. 2025, 36, 1669−1676

Deviations in the median ratio are observed in the last two samples of batch 4 and the first three samples of batch 5 (Figure 4A). Although the sample numbers are not sequential due to the batch design, their measurements are consecutive. The observed deviations are corrected by the between-batch adjustment applied by mzQuality (Figure 4B).

3.4.4. Individual Compound Plots. The individual compound plot can display peak area, area_IS, ratio, corrected ratio, retention time, and other numerical values related to compounds plotted against the injection number. While compound plots can visualize all sample types, CAL, ACAL, blank, LQC, and PROC are not included in Figure 5 for clarity.

Figure 5A displays an individual compound plot for an IS. These plots provide a unique ability to visually inspect the IS quality of individual samples within mzQuality.

Figure 5B provides an example of a compound plot without trends over time or other irregularities. Here, the between-batch corrected area ratio is plotted against the injection number of samples, which can reveal trends within and across batches.

Figure 5 C,D demonstrates the effect of within-batch correction before (Figure 5C) a noticeable downward trend over time is observed, which is no longer present in Figure 5D. Within-batch correction is applied to all compounds, so its effects should be checked for each compound.

The additional plots including all sample types can be found in Supplementary Figure 2. These additional plots can be used to monitor trends in the retention time. This is particularly valuable for identifying potential data integration errors. Samples with missing values, as indicated by NA, will not have a data point shown.

After visual inspection of all individual compound plots, it may be necessary to revisit peak integration to address any observed irregularities. If these irregularities can be corrected, new export files should be generated and mzQuality must be run from the beginning. This output folder includes all of the previously mentioned plots and Excel files that can be used for statistical analysis in tools such as R.

The Excel file lists all measured compounds, categorized by confidence level. Compounds are organized into three tabs based on quality and user-defined or default settings: the "High confidence" tab includes compounds with RSDqc values below 15%, the "With Caution" tab includes compounds with RSDqc values ranging between 15% and 30%, and the "Low Signal to Noise" tab including compounds with either background signal >40% or RSDqc values >30%. An example and further details on the content in the output folder are provided in GitHub: https://github.com/hankemeierlab/mzQuality.

4. CONCLUSIONS

Here, we outlined practices to minimize experimental variation and to ensure optimal performance of mzQuality. Additionally, we showcased a typical application of mzQuality with a supplied data set, collected across multiple batches. The software can manage any type of raw data independent of vendor software and does not require the user to have any programming skills. Its intuitive visualization of data enables users to assess their data easily. Moreover, mzQuality assures the user that the data are suitable for statistical and biological interpretation in line with the requirements set by regulatory bodies like the FDA and EMA. By adhering to these recommended practices and by utilizing mzQuality, we anticipate that the metabolomics community will generate higher-quality data, thereby improving the quality of data reported in the literature.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.5c00073.

PCA plots of all example data and compound plots including all sample types of one example compound (PDF)

Table S1: the full batch design of all example data (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Alida Kindt − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands;* orcid.org/0000-0001-6551-6030; Phone: +31715272727; Email: a.s.d.kindt@lacdr.leidenuniv.nl

Authors

Marielle van der Peet − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands;* orcid.org/0000-0001-8155-1485

Pascal Maas − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands*

Agnieszka Wegrzyn − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands;* orcid.org/0000-0003-4872-8626

Lieke Lamont − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands*

Ronan Fleming − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands; Digital Metabolic Twin Centre, School of Medicine, University of Galway, Galway H91 TK33, Ireland*

Constance Bordes − *University of Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR1219, Bordeaux F-33000, France;* orcid.org/0000-0003-3367-1778

Stéphanie Debette − *University of Bordeaux, INSERM, Bordeaux Population Health Research Center, UMR1219, Bordeaux F-33000, France; Department of Neurology, Institute for Neurodegenerative Diseases, Bordeaux University Hospital, Bordeaux F-33000, France; Institut du Cerveau (ICM), Paris Brain Institute, INSERM U1127, UMR CNRS 7225 Paris, Sorbonne Université, Assistance Publique des Hôpitaux de Paris, Paris 75013, France*

Amy Harms − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands;* orcid.org/0000-0002-2931-4295

Thomas Hankemeier − *Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden 2333 CC, The Netherlands;* orcid.org/0000-0001-7871-2073

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.5c00073

Funding

This work was cofunded by the European Union's Horizon Europe Framework Programme, Recon4IMD [grant number 101080997]; the CHDR as part of the SKINERGY trial under NWA-ORC project [NWA grant number 1389.20.182]; X-Omics [NWO grant number 184.034.019]; EXPOSOME-SCAN project [grant number 175.2019.032]; and the NWO Gravitation program EXPOSOME-NL [grant number 024.004.017].

Notes

During the preparation of this work, the author(s) used ChatGPT to improve readability and conciseness. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We sincerely thank Sam Heijer and Jelte Geerlings for their invaluable contributions to analyzing the samples that generated the data for this study.

■ REFERENCES

(1) Chu, X.; Jaeger, M.; Beumer, J.; Bakker, O. B.; Aguirre-Gamboa, R.; Oosting, M.; Smeekens, S. P.; Moorlag, S.; Mourits, V. P.; Koeken, V. A. C. M.; de Bree, C.; Jansen, T.; Mathews, I. T.; Dao, K.; Najhawan, M.; Watrous, J. D.; Joosten, I.; Sharma, S.; Koenen, H. J. P. M.; Withoff, S.; Jonkers, I. H.; Netea-Maier, R. T.; Xavier, R. J.; Franke, L.; Xu, C.-J.; Joosten, L. A. B.; Sanna, S.; Jain, M.; Kumar, V.; Clevers, H.; Wijmenga, C.; Netea, M. G.; Li, Y. Integration of Metabolomics, Genomics, and Immune Phenotypes Reveals the Causal Roles of Metabolites in Disease. *Genome Biol.* **2021**, *22* (1), 198.

(2) Kohler, I.; Hankemeier, T.; van der Graaf, P. H.; Knibbe, C. A. J.; van Hasselt, J. G. C. Integrating Clinical Metabolomics-Based Biomarker Discovery and Clinical Pharmacology to Enable Precision Medicine. *Eur. J. Pharm. Sci.* **2017**, *109*, S15−S21.

(3) Allard, P.-M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V. C.; Roussi, F.; Litaudon, M.; Wolfender, J.-L. Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. *Anal. Chem.* **2016**, *88* (6), 3317−3323.

(4) Han, W.; Li, L. Evaluating and Minimizing Batch Effects in Metabolomics. *Mass Spectrom. Rev.* **2022**, *41* (3), 421−442.

(5) Livera, A. M. D.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J. A.; Castillo, S.; Simpson, J. A.; Speed, T. P. Statistical Methods for Handling Unwanted Variation in Metabolomics Data. *Anal. Chem.* **2015**, *87* (7), 3606−3615.

(6) Liu, Q.; Walker, D.; Uppal, K.; Liu, Z.; Ma, C.; Tran, V.; Li, S.; Jones, D. P.; Yu, T. Addressing the Batch Effect Issue for LC/MS Metabolomics Data in Data Preprocessing. *Sci. Rep.* **2020**, *10* (1), 13856.

(7) Wehrens, R.; Hageman, J. A.; van Eeuwijk, F.; Kooke, R.; Flood, P. J.; Wijnker, E.; Keurentjes, J. J. B.; Lommen, A.; van Eekelen, H. D. L. M.; Hall, R. D.; Mumm, R.; de Vos, R. C. H. Improved Batch Correction in Untargeted MS-Based Metabolomics. *Metabolomics* **2016**, *12*, 88.

(8) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* **2018**, *14* (6), 72.

(9) Brown, M.; Dunn, W. B.; Ellis, D. I.; Goodacre, R.; Handl, J.; Knowles, J. D.; O'Hagan, S.; Spasić, I.; Kell, D. B. A Metabolome Pipeline: From Concept to Data to Knowledge. *Metabolomics* **2005**, *1* (1), 39−51.

(10) Drotleff, B.; Lämmerhofer, M. Guidelines for Selection of Internal Standard-Based Normalization Strategies in Untargeted Lipidomic Profiling by LC-HR-MS/MS. *Anal. Chem.* **2019**, *91* (15), 9836−9843.

(11) Lange, E.; Gröpl, C.; Reinert, K.; Kohlbacher, O.; Hildebrandt, A. High-Accuracy Peak Picking of Proteomics Data Using Wavelet Techniques. *Pac. Symp. Biocomput.* **2006**, 243−254.

(12) *MZmine 2 - Features*; https://mzmine.github.io/features.html#Batch (accessed 2024−03−06).

(13) Naake, T.; Rainer, J.; Huber, W. MsQuality: An Interoperable Open-Source Package for the Calculation of Standardized Quality Metrics of Mass Spectrometry Data. *Bioinformatics* **2023**, *39* (10), No. btad618.

(14) Kirwan, J. A.; Gika, H.; Beger, R. D.; Bearden, D.; Dunn, W. B.; Goodacre, R.; Theodoridis, G.; Witting, M.; Yu, L.-R.; Wilson, I. D. the metabolomics Quality Assurance and Quality Control Consortium (mQACC). Quality Assurance and Quality Control Reporting in Untargeted Metabolic Phenotyping: mQACC Recommendations for Analytical Quality Management. *Metabolomics* **2022**, *18* (9), 70.

(15) Dunn, W. B.; Broadhurst, D. I.; Edison, A.; Guillou, C.; Viant, M. R.; Bearden, D. W.; Beger, R. D. Quality Assurance and Quality Control Processes: Summary of a Metabolomics Community Questionnaire. *Metabolomics* **2017**, *13* (5), 50.

(16) Yang, W.; Schoeman, J. C.; Di, X.; Lamont, L.; Harms, A. C.; Hankemeier, T. A Comprehensive UHPLC-MS/MS Method for Metabolomics Profiling of Signaling Lipids: Markers of Oxidative Stress, Immunity and Inflammation. *Anal. Chim. Acta* **2024**, *1297*, No. 342348.

(17) Baggerly, K. A.; Edmonson, S. R.; Morris, J. S.; Coombes, K. R. High-Resolution Serum Proteomic Patterns for Ovarian Cancer Detection. *Endocr. Relat. Cancer* **2004**, *11* (4), 583−584.

(18) *R: Rosner's Test for Outliers*; https://search.r-project.org/CRAN/refmans/EnvStats/html/rosnerTest.html (accessed 2024−04−11).

(19) van der Kloet, F. M.; Bobeldijk, I.; Verheij, E. R.; Jellema, R. H. Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. *J. Proteome Res.* **2009**, *8* (11), 5132−5141.