Evaluating the effect of data merging and post-acquisition normalization on statistical analysis of untargeted high-resolution mass spectrometry based urinary metabolomics data

Fynn Brix,[†,*] Tobias Demetrowitsch,[†] Julia Jensen-Kroll,[†] Helena Zacharias,[▽,#,⊥] Silke Szymczak,[§] Matthias Laudes,[#,‖] Stefan Schreiber,[#,‖] Karin Schwarz[†]

[†]Institute of Human Nutrition and Food Science, Kiel University, Kiel, Heinrich-Hecht-Platz 10, 24118 Kiel, Germany

[▽]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, 30625 Hannover, Germany

[§]Institute of Medical Biometry and Statistics, University of Luebeck and Medical Centre Schleswig-Holstein, Campus Luebeck, 23562 Luebeck, Germany

[#]Department of Internal Medicine I, University Medical Centre Schleswig-Holstein, Campus Kiel, Kiel, Germany

[⊥]Institute of Clinical Molecular Biology, Kiel University and University Medical Center Schleswig-Holstein, Campus Kiel, 24105 Kiel, Germany

[‖]Institute of Diabetes and Clinical Metabolic Research, Kiel University, Düsternbrooker Weg 17, 24105, Kiel, Germany

ABSTRACT: Urine is one of the most widely used biofluids in metabolomic studies, because it can be collected non-invasively and is available in large quantities. However, it shows large heterogeneity in sample concentration and consequently requires normalization to reduce unwanted variation and extract meaningful biological information. Biological samples like urine are commonly measured with electrospray ionization (ESI) coupled to a mass spectrometer, producing datasets for positive and negative mode. Combining these gives a more complete picture of the total metabolites present in a sample. However, the effect of this data merging on subsequent data analysis, especially in combination with normalization, has not yet been analysed. To address this issue, we conducted a neutral comparison study to evaluate the performance of eight post-acquisition normalization methods under different data merging procedures using 1029 urine samples from the Food Chain plus (FoCus) cohort. Samples were measured by a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS). Normalization methods were evaluated by five criteria capturing the ability to remove sample concentration variation and preserve relevant biological information. Merging data after normalization was generally favourable for quality control (QC) sample similarity, sample classification and feature selection for most of the tested normalization methods. Merging data after normalization and the usage of probabilistic quotient normalization (PQN) in a similar setting are generally recommended. Relying on a single analyte to capture sample concentration differences, like with post-acquisition creatinine normalization, seems to be a less preferable approach, especially when data merging is applied.

Urine is one of the most widely used biofluids in metabolomic studies, because it can be collected non-invasively and is available in large quantities.[1] However, it has a large heterogeneity in sample concentration[2] and volume may change up to 15-fold under normal conditions.[3]
Consequently, numerous normalization methods have been developed to reduce variation originating from unwanted factors.[1] These normalization methods may be categorized as being pre- or post-acquisition.[2] Pre-acquisition methods adjust the sample volumes based on measured reference quantities These parameters can also be used in post-acquisition methods, but requires an additional workflow step and information on sample volume is prerequisite.[4]

Post-acquisition normalization methods are applied after data collection.[5] Several evaluation studies have been conducted for human urine,[2,6–8] serum,[9] and plasma,[10] as well as animal urine,[11,3,12] measured by nuclear magnetic resonance spectroscopy or liquid chromatography-mass spectrometry. Frequently used methods include PQN,[11,13,2,6,8,5] as well as variance stabilization normalization (VSN),[6,8,14] and quantile normalization.[6,7,14].

Most post-acquisition normalization methods adjust for sample-to-sample variation, whereas VSN additionally adjusts for variation on the metabolite level.[7,8] All of the above normalization methods make specific assumptions about the data. Whether or not these assumptions are met for a specific data set may influence the performance of the normalization methods.[15] Thus, the selection of a particular normalization method should be made based on the data characteristics, research question and the subsequent data analysis methods.[13]

Studies comparing post-acquisition normalization methods with urine samples and mass spectrometry (MS) with ESI are either using positive ionization of molecules only,[16–18] or include datasets of both polarities.[19,8,12] However, none used merged datasets from both polarities. Furthermore, most studies are based on small sample sizes with < 100 samples,[19,16,17,2,6,7,20,5] and only few are based on data sets with > 1000 samples.[8,21] The combined effect of data merging and normalization has not yet been evaluated in the workflow for pre-processing metabolomics data. Thus, in this neutral comparison study we aim to objectively evaluate the performance of eight post-acquisition normalization methods under different data merging procedures based on a large-scale urinary metabolomics dataset. Based on our results we will provide recommendations for different data merging procedures and normalization methods in different scenarios.

MATERIALS AND METHODS

Human Urine Samples and Sample preparation. Urine metabolomics profiling was performed in the Food Chain Plus (FoCus) cohort, which has been published recently.[22] The cohort was established in 2011 for population-based research with a focus on metabolic inflammation. The study was approved by the local ethics committee of the Kiel University (A156-03/Date 2011/07/28) and was registered under the clinical trial number DRKS00005285 at the German Clinical Trials Register in Cologne.[22] The average age of the study participants who gave urine is 52 years with biological sex of 40% males and 60% females.

Spot urine samples of 1031 participants were available; two samples were excluded, because creatinine was lacking or data for one analytical method were missing. Samples were diluted 1:500 with methanol and water (50:50, v/v) prior to analysis. The preparation of the quality control (QC) samples followed a procedure modified from a previous publication.[23]

Data acquisition. Data were acquired using a 1260 Infinity HPLC (Agilent, Waldbronn, Germany) for direct injection of samples. The HPLC was linked to an ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS) (7T, SolariXR, Bruker, Bremen, Germany). Mass spectra were acquired with electrospray ionization (ESI) source in both modes (positive and negative ionization) and two methods with a mass range between 65 and 1500 Da (USM, SM), leading to four data sets per sample. The intensity threshold was 10^6 counts. Data were calibrated using an in-house database in the quadratic mode with a tolerated mass error < 0.5 mDa. Detailed instrumental parameters can be found in an earlier publication.[24] Blank samples were injected prior to each batch. Pooled QC samples were injected at the start and end of each batch.

Data processing and merging. Bruker raw data (.d) were processed in the MetaboScape 2021b software (Bruker, Bremen, Germany). Ion deconvolution and other settings are given in Table S1. Most probable chemical formulas were assigned based on accurate measured masses, isotopic patterns, as well as the seven golden rules.[25] Potential compound names were assigned based on the Human Metabolome Database entries (version 5.0) for all human biospecimens.[26] The raw Bucket tables were

exported and further processed in R (version 4.1.2). Figure 1 shows a schematic depiction of the different analysis workflows. Signal correction, peak filtering and imputation of missing values were conducted using the R package statTarget version 1.24.0.[27] Signal correction was applied using QC-based random forest signal correction (QC-RFSC). Peak filtering was conducted by only including compounds that were detected in at least 80% of the samples. Imputation of missing values was done using the k-Nearest Neighbour (kNN) method.[27]

After the signal correction the datasets were merged. Data merging was done in three different ways: merging all four analytical datasets into one ("by-method"), merging positive and negative data into two datasets ("by-polarity") and no merging at all prior to normalisation ("singles"). Merging of compounds detected in multiple datasets was done by selecting the compound with either the lowest number of missing values, the highest median intensity, or the highest mean intensity calculated using all samples. In case of ties the criteria are applied in the described order. Compounds only detected in one of the datasets are simply added to the respective merged dataset. For the next step, each of the merged datasets was corrected with each of the normalisation methods. Subsequently, all datasets were merged into one dataset for comparison.

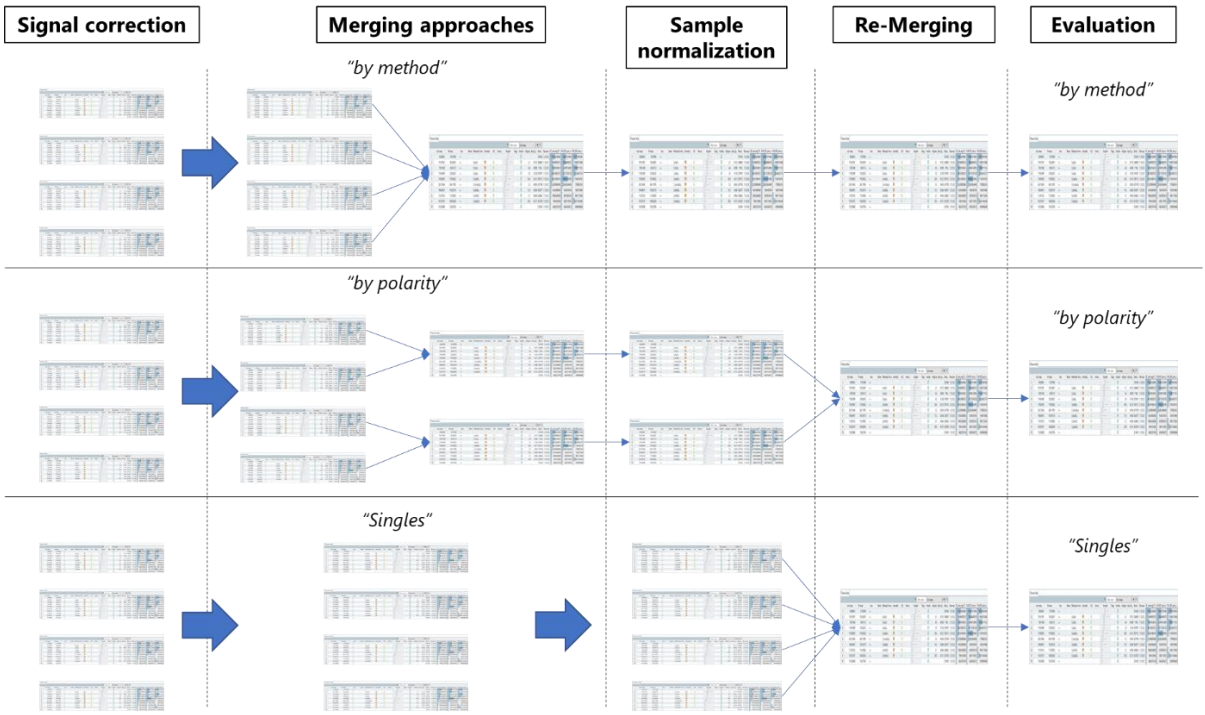

Figure 1: Schematic depiction of the analysis workflows. First, signal corrected data are merged either to one set, two sets or kept as four individual sets. Different normalization methods are applied and data are merged subsequently and evaluated by different criteria.

Evaluated normalization methods. Post-acquisition normalization methods evaluated were: creatinine, sum, MS total useful signal (MSTUS),[12] PQN[28] with median of all QC samples as reference spectrum and all samples as reference, quantile,[29] median, VSN,[30] and cubic spline normalization (CSN)[31] . A description for each method and the R packages which have been used can be found in Table S2. Baseline data refers to data corrected using a QC-based random forest signal correction method.[27]

Evaluation criteria. The performance evaluation of the normalization methods emphasized on removal of sample concentration variation and preservation of biological information. Sample concentration variation removal was assessed by QC clustering in principal component analysis (PCA)

and the median of the relative standard deviations (RSD) of all metabolites in QC samples. QC clustering was determined in a quantitative way by calculating the average auto scaled distances of each QC sample to the centroid of all QC samples in the score plot.

Preservation of biological information was evaluated by training random forest (RF) models using the R package ranger[32] with default parameter settings to predict subjects' sex. To identify compounds differentiating for sex a Web of Science literature search was conducted. For literature search keywords and the resulting 41 compounds see Tables S3 and S4.

RESULTS AND DISCUSSION

Common post-acquisition normalization methods were evaluated under different merging procedures by the ability to remove sample concentration variation and to preserve biological information. The first criterion was tested via reduction of QC sample variation and the latter via comparing prediction accuracy, number of significant metabolites, and matches with previously reported differing metabolites for biological sex.

Comparison of merging procedures. All normalization methods except for creatinine achieved the lowest average distance among QC samples after the "singles" merging procedure (Table 1). Furthermore, almost all normalization methods except for creatinine exhibited a monotonic decline of the QC distance. This trend is also reflected in the total distance score for all normalisation methods combined, which ranged from 85 for "by-method" to 56 for "by-polarity" and down to 26 for "singles" merging procedure. Results for the RSD of QCs were similar to those of the QC distance, since all normalization methods expect for one (here CSN) achieved the lowest average RSD value after "singles" merging. In addition, a monotonic decline in the RSD was also observed for five of the normalization methods, namely the MSTUS, creatinine, PQN(QC), sum, and median normalization. A systematic decline was also observed, calculating the total of the RSD values. Merging procedures "by-method" and "by-polarity" did not differ much, with values of 120 and 118 respectively. However, this value decreased to 107 for the "singles" merging procedure. The smallest change in the QC distance and average RSD in QCs was found for VSN normalized data. This is because the variance stabilization reduces variability and the data is on a different scale after the normalization.

Table 1: Average distance of QCs in the PCA score plots and average RSD of features in QCs, for each normalization method and merging procedure[a]

Normalization methods	PCA QC aDist			QC RSD		
	By-method	By-polarity	Singles	By-method	By-polarity	Singles
Baseline	2.57	2.6	2.61	5.55	5.55	5.55
Creatinine	**0.6**	4.65	1.91	11.98	10.58	8.33
Sum	2.35	1.4	0.75	6.94	6.00	4.27
MSTUS	2.06	1.41	1.21	6.51	5.16	5.13
Median	2.02	1.32	0.56	5.73	5.22	4.39
PQN	1.93	1.61	0.60	6.84	7.26	4.42
PQN(QC)	2.03	**1.28**	**0.53**	5.72	5.15	4.31
Quantile	32.22	16.32	7.5	53.14	54.56	52.28
CSN	36.93	22.50	7.9	17.68	18.14	17.73
VSN	3.06	3.04	3.00	**0.83**	**0.85**	**0.82**
Total:	*85.77*	*56.13*	*26.57*	*120.92*	*118.47*	*107.23*

[a]QC aDist, Average distance of QCs; QC RSD, average RSD of features in QCs. Bold letters indicate best values for the respective merging procedure (column).

The balanced accuracy of the random forest models predicting the sex of subjects using the differently merged and normalized data is shown in Table 2. The highest balanced accuracy was achieved by models built using data from the "singles" merging procedure. MSTSU and VSN were the exceptions, which had the best values after merging according to polarity. Also, the majority of normalization methods attained the most significant metabolites for predicting sex (Table 5) after the "singles" merging procedure. Moreover, there is also a positive increasing trend for the average number of significant metabolites calculated including all normalization methods for each respective merging procedure. Normalization methods benefitting from "singles" merging include creatinine, PQN(QC), sum, median and CSN. Merging "by-polarity" was the best procedure for MSTUS normalization. Merging "by-method" resulted in the highest number of significant features for PQN and VSN normalized data.

Table 2: Test set based average balanced accuracy for predicting the subjects' sex[a]

Normalization methods	**Average balanced accuracy (SD)**		
	By method	By polarity	Singles
Baseline	74.68 (3.50)	75.58 (2.99)	74.67 (3.05)
Creatinine	69.81 (2.82)	71.91 (1.31)	73.05 (3.90)
Sum	76.33 (2.60)	76.02 (3.59)	76.49 (3.08)
MSTUS	74.77 (4.46)	75.77 (3.58)	74.26 (3.34)
Median	76.26 (3.70)	76.05 (3.56)	**77.98 (3.37)**
PQN	76.35 (2.71)	75.91 (3.61)	77.42 (2.93)
PQN(QC)	75.86 (3.19)	76.14 (2.83)	76.64 (3.24)
Quantile	**76.65 (3.02)**	**76.96 (2.91)**	77.62 (3.18)
CSN	76.35 (2.19)	76.34 (3.43)	77.46 (2.99)
VSN	74.75 (3.29)	75.97 (3.63)	75.38 (3.93)
Average:	*75.18 (2.04)*	*75.66 (1.37)*	*76.10 (1.67)*

[a]Bold letters indicate best values for each merging procedure. Merging procedures are compared row-wise and normalization methods column-wise.

To summarize the above results, a comparison was made for each normalization method as to which merging procedure lead to the best values for each of the evaluation criteria (Table S6). "Singles" merging led to the highest values for four out of five evaluation criteria for all normalization methods except for MSTUS. Median normalization achieved best values with the same merging procedure ("singles") for all evaluation criteria. VSN and MSTUS normalization had the highest prediction values with polarity merging and for RSD and PCA, "singles" merging led to better results.

Table 3: Number of shared significant metabolites for predicting the subjects' sex[a]

Normalization methods	# of shared significant metabolites		
	By method	By polarity	Singles
Baseline	98	96	95
Creatinine	107	110	116
Sum	103	106	108
MSTUS	103	105	97
Median	102	111	**122**
PQN	**123**	**116**	112
PQN(QC)	107	104	116
Quantile	106	100	105
CSN	105	113	117
VSN	98	97	89
Total:	*1137*	*1145*	*1164*

[a]Significant metabolites were identified as described in the Material and Methods section. Bold letters indicate best values for the respective merging procedure.

The above results showed that the "singles" merging procedure was optimal for almost all normalisation methods. One possible explanation is, that with separate normalization of the datasets, the assumptions of the normalization methods are better met. This is supported by the fact that methods with similar assumptions benefit from the same merging procedure. The methods sum, MSTUS, median, and both PQN variants all share the premise that only a small proportion of the compounds (i.e. metabolites) is up and down regulated in equal frequencies.[6,10] These methods all perform the best with "singles" merging procedures, except for one of the evaluation criteria (Table S5). The MSTUS method seems to favour merging "by-polarity", however, the performance for this particular method may also depend on the number of compounds detected in all samples, which differs between merging procedures. Better performance for single merging for CSN and quantile normalization may be due to compound intensities being more comparable from sample to sample, if datasets are kept separate for normalization.[7]

Another influencing factor may be, that compound signal levels could differ between analytical methods or ESI modes, due to different ionization efficiencies.[33] Therefore, a single compound, like in creatinine normalization, may not be representative enough for the whole dataset, which includes peaks generated with differing ionizations or analytical methods. This would be a possible explanation for the bad performance of creatinine normalization. Nevertheless, creatinine is still frequently used.[1,5]

Reported gender related metabolites. In order to evaluate the potential loss of biological information by normalization and merging, the number of matches of the significant metabolites with the reported sex-specific metabolites was calculated and compared to the number of the baseline data (Table 5). A total of eight studies[34–41] were used for the compilation of the sex related metabolites and initially 65 compounds were included. A chemical formula was identified for 56 of the compounds and 41 compounds were matched by chemical formula in the study datasets. It has to be noted that it is not expected for any normalization method to match with all the differential metabolites as they are not established biomarkers themselves. The "singles" merging procedure led to the highest number of matches for eight out of nine normalization methods. For PQN, VSN, and sum normalization, the numbers were equal between singles and "by-polarity" merging procedures. For PQN(QC), "singles" and "by-method" merging were the best merging procedures. For the MSTUS method, the merging "by-method" was the optimal procedure.

Table 5: Number of matches among the significant metabolites with reported sex metabolites[a]

Normalization methods	# matches with reported sex metabolites		
	By method	By polarity	Singles
Baseline	7	9	8
Creatinine	5	4	8
Sum	8	9	9
MSTUS	11	10	8
Median	10	8	**12**
PQN	11	**12**	**12**
PQN(QC)	**12**	9	**12**
Quantile	10	9	**12**
CSN	9	10	11
VSN	7	9	9
Total:	*96*	*94*	*106*

[a]Bold letters indicate best values for the respective merging procedure.

Comparing the overlap of the significant features with the reported sex related metabolites showed that, generally, normalization methods with a high number of significant metabolites also showed a high number of matches with sex metabolites, compared to methods with a lower number of significant metabolites (Table 3). Creatinine is the exception, showing a high number of significant metabolites, which is not reflected by the number of matches for the sex metabolites. This may be indicative of possibly more low-quality features being selected among the significant metabolites for creatinine normalized data in comparison to the other. PQN and PQN(QC) achieved the highest number of matches and showed stable performance across merging procedures. Creatinine and VSN normalized data led to a low number of matches across merging procedures. Creatinine normalization did not improve baseline data for all merging procedures.

Gender/sex have been used to evaluate urinary sample normalization methods in the past, e.g., for sample clustering in PCA,[42] relation of model variance,[1] and as group variable to determine differential metabolites.[15] PQN normalization gave the highest matches of significant metabolites with reported sex metabolites, however, none of the normalization methods was superior. Similarly, Li and colleagues[8] also found comparable performance of the normalization methods PQN, CSN, MSTUS, VSN, and quantile normalization in terms of the overlap between experimentally validated biomarkers and spiked-in biomarkers with determined statistically significant metabolites. Quantile normalization performed slightly lower based on a spike-in dataset, which was not observed in our study. It is possible, that intensity differences due to the spiking-in of compounds with different concentrations are diminished in quantile normalization, because of the assignment of average values during the normalization.

In this study, all normalization methods exceeded or at least equalled the baseline data in terms of matches with the reported gender metabolites except for creatinine and median normalization. Kohl and colleagues[7] evaluated the retention of genuine biological information by relating variation of expected constant features with that of varied spiked-in features. In line with this work, PQN performed the best and was comparable to the non-normalized data. Quantile, CSN, and VSN performed fairly comparable and did not match the non-normalized ones, however, the spiked-in signals still clearly stood out. Mervant and colleagues[1] also used gender and assessed biological information when comparing normalization methods. They evaluated the explained variance

associated with sex after normalization using a modified partial least squares (PLS) method. Contrary to the results here, they found post-acquisition creatinine to slightly increase variance related to sex, while PQN and MSTUS did not increase variance, compared to the reference data. Possible reasons for differing results may include differences in the PQN application, statistical workflow, and that explained variance may not be directly correlate with matches for reported sex metabolites.

Comparison of normalization methods. To determine which normalization methods performed optimal considering all evaluation criteria and merging procedures, individual ranks for each evaluation criterion were assigned (Table 4). These ranks were equally weighted by the same factor and the resulting weighted ranks summed up. The method with the lowest weighted sum is placed on rank 1 overall.

Table 4: Ranking of normalization methods based on evaluation criteria for each merging procedure[a]

Method	Rank	Weighted Sum	PCA	RSD	RF (BA[1])	RF (sig mets[2])	RF (lit matches[3])
Merged by method							
PQN	1	2.80	**2**	6	**2.5**	**1**	**2.5**
PQN(QC)	2	3.30	4	**3**	6	**2.5**	**1**
Median	3	4.90	**3**	4	5	8	4.5
MSTUS	4	5.20	5	5	7	6.5	**2.5**
Quantile	5	6.10	9	10	**1**	4	4.5
Sum	6	6.30	6	7	4	6.5	7
Creatinine	7	6.70	**1**	8	10	**2.5**	10
CSN	8	6.90	10	9	**2.5**	5	6
VSN	9	7.20	8	**1**	8	9.5	8.5
Baseline	10	7.40	7	**2**	9	9.5	8.5
Merged by polarity							
PQN(QC)	1	3.84	**1**	**2**	**3**	7	6.2
PQN	2	4.20	5	7	7	**1**	**1**
Median	3	4.40	**2**	4	4	**3**	9
MSTUS	4	4.70	4	**3**	8	6	**2.5**
Sum	5	5.04	**3**	6	5	5	6.2
CSN	6	5.50	10	9	**2**	**2**	**2.5**
VSN	7	5.84	7	**1**	6	9	6.2
Baseline	8	7.24	6	5	9	10	6.2
Quantile	9	7.24	9	10	**1**	8	6.2
Creatinine	10	8.20	8	8	10	4	10
Singles							
Median	1	1.79	**2**	4	**1**	**1**	**2.5**
PQN(QC)	2	3.71	**1**	**3**	5	**3.5**	**2.5**
PQN	3	3.79	**3**	5	4	5	**2.5**
Sum	4	5.07	4	**2**	6	6	6.5
CSN	5	5.43	10	9	**3**	**2**	5
Quantile	6	5.21	9	10	**2**	7	**2.5**
VSN	7	6.79	8	**1**	7	10	6.5
MSTUS	8	8.00	5	6	9	8	9
Creatinine	9	8.50	6	8	10	**3.5**	9
Baseline	10	8.14	7	7	8	9	9

[a]BA, balanced accuracy; sig mets, significant metabolites; lit matches, matches with literature metabolites for sex. Bold letters indicate best values for each evaluation criterion.

Methods with top ranks for QC clustering in PCA include PQN, PQN(QC), median, MSTUS and sum normalization (Table 4). The QC distance values for these methods were similar across merging procedures (Table 1). Normalization methods with low ranking performance include VSN, quantile, CSN. The results for the RSD were very similar to those of the PCA analyses. PQN(QC), PQN, median, sum and MSTUS showed similar performance and are at the top of the ranking across merging procedures. Creatinine, CSN and quantile normalized data consistently achieved lower ranks and none of them achieved a lower RSD than the baseline data. Remarkably, for merging "by-method", none of the normalization methods (except VSN) were able to achieve a lower RSD than the value of 5.55 from the QC-RFSC corrected data. However, VSN normalized data are on a generalised logarithm scale with base 2 and thus not immediately comparable to the average RSD of the other normalization methods.

The balanced accuracy values of the sex prediction were similar across normalization methods, expect for creatinine normalization, which yielded the lowest values (Table 2). Creatinine normalized data did not achieve a higher balanced accuracy than the baseline data, which was also true for MSTUS if merged by "singles" procedure. Quantile and CSN had the highest accuracies across the three merging procedures. The number of significant metabolites across all normalization methods were also fairly similar. Only PQN with merging "by-method" exhibited more significant metabolites (123), followed by creatinine with 107 (Table 3). CSN, creatinine, and PQN normalization led to a high number of significant metabolites across all merging procedures compared to the other methods. VSN showed the lowest number of significant metabolites across all merging procedures.

PQN and PQN(QC) normalizations were among the best methods, resulting in the lowest weighted sum of individual ranks for each evaluated criterion across all merging procedures (Table 4). This is in line with other studies,[11,9,13,6,8,43] which also found PQN to perform best and recommended it as optimum normalization method. Median normalization despite its simplicity was also among the top methods across merging procedures. Other studies have also found Median normalization comparable to PQN in terms of sample clustering in PCA and pooled RSD using test samples,[15] as well as RSD in QC samples,[43] which is in accordance with findings of this work. The similarity in the performance between PQN and median normalization may be due to the fact that both methods operate in a similar manner by relating each sample to a median spectrum.[43] The performance of a normalization method may depend on how well the assumptions of the respective method are met by the data.[15] Therefore, one possible explanation for the performance of PQN and median may be, that the current data meets the assumptions of these two methods more, compared to those of the other methods. Quantile normalization varied in its overall ranking across merging procedures, but showed particular good performance for the sample classification, which is consistent with other research.[7] Post-acquisition creatinine normalization showed low performance in agreement with earlier studies.[16,17,7,1,20,36,5,18,12,42] In the PCA analysis, creatinine normalized QC samples showed greater dispersion in comparison to other normalization methods, in line with previous reports.[17,20,12] Sum normalization showed a mediocre performance in this work, ranging in the middle of the tested normalisation methods. Sum normalization may be susceptible to compounds with large abundance,[10] which would explain the spread of QC samples in PCA. Therefore, some authors questioned the usage of sum normalization for metabolomics data.[6,10]

CONCLUSION

The present study shows for the first time that data merging has an effect on normalization performance and subsequent analysis steps and must be considered when planning the data analysis. Merging data after normalization was generally favourable for QC similarity, sample classification and feature selection for most of the tested normalization methods. PQN and Median normalization showed the best performance overall, considering all tested criteria. Based on this, several

recommendations can be provided. Merging data after normalization ("singles") and the usage of PQN in a similar setting are generally recommended. PQN is preferred here over median normalisation because of the more suitable assumptions made about the data. Relying on a single analyte to capture sample concentration differences, like with post-acquisition creatinine normalization, seems to be a less preferable approach, especially when data merging is applied. The results of this study may have broader implications, since other biological matrices like saliva, sweat or faeces also show heterogeneity in sample concentration or metabolite signals.

■ ASSOCIATED CONTENT

*S Supporting Information

Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*F.B. E-mail: fbrix@foodtech.uni-kiel.de. Tel.: +49 (0)431 880 5365.

Author Contributions

The paper was written through contributions of all authors. All authors have given approval to the final version of the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGEMENTS

This work was supported by the German Ministry of Education and Research (BMBF) grant 0315540C and by the Cluster of Excellence "Precision Medicine in Chronic Inflammation" of the German Research Foundation grant EXC2167. HUZ is supported by the BMBF within the framework of the e:Med research and funding concept grant 01ZX1912A.

References

(1) Mervant, L.; Tremblay-Franco, M.; Jamin, E. L.; Kesse-Guyot, E.; Galan, P.; Martin, J.-F.; Guéraud, F.; Debrauwer, L. *Metabolomics : Official journal of the Metabolomic Society* **2021**, DOI: 10.1007/s11306-020-01758-z.

(2) Gagnebin, Y.; Tonoli, D.; Lescuyer, P.; Ponte, B.; Seigneux, S. de; Martin, P.-Y.; Schappler, J.; Boccard, J.; Rudaz, S. *Analytica chimica acta* **2017**, DOI: 10.1016/j.aca.2016.12.029.

(3) Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier-Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Analytical chemistry* **2011**, DOI: 10.1021/ac201065j.

(4) Wu, Y.; Li, L. *Journal of chromatography. A* **2016**, DOI: 10.1016/j.chroma.2015.12.007.

(5) Rosen Vollmar, A. K.; Rattray, N. J. W.; Cai, Y.; Santos-Neto, Á. J.; Deziel, N. C.; Jukic, A. M. Z.; Johnson, C. H. *Metabolites* **2019**, DOI: 10.3390/metabo9100198.

(6) Hochrein, J.; Zacharias, H. U.; Taruttis, F.; Samol, C.; Engelmann, J. C.; Spang, R.; Oefner, P. J.; Gronwald, W. *Journal of proteome research* **2015**, DOI: 10.1021/acs.jproteome.5b00192.

(7) Kohl, S. M.; Klein, M. S.; Hochrein, J.; Oefner, P. J.; Spang, R.; Gronwald, W. *Metabolomics : Official journal of the Metabolomic Society* **2012**, DOI: 10.1007/s11306-011-0350-z.

(8) Li, B.; Tang, J.; Yang, Q.; Cui, X.; Li, S.; Chen, S.; Cao, Q.; Xue, W.; Chen, N.; Zhu, F. *Scientific reports* **2016**, DOI: 10.1038/srep38881.

(9) Di Guida, R.; Engel, J.; Allwood, J. W.; Weber, R. J. M.; Jones, M. R.; Sommer, U.; Viant, M. R.; Dunn, W. B. *Metabolomics* **2016**, DOI: 10.1007/s11306-016-1030-9.

(10) Wulff, J. E.; Mitchell, M. W. *ABB* **2018**, DOI: 10.4236/abb.2018.98022.

(11) Chen, J.; Zhang, P.; Lv, M.; Guo, H.; Huang, Y.; Zhang, Z.; Xu, F. *Analytical chemistry* **2017**, DOI: 10.1021/acs.analchem.6b05152.
(12) Warrack, B. M.; Hnatyshyn, S.; Ott, K.-H.; Reily, M. D.; Sanders, M.; Zhang, H.; Drexler, D. M. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2009**, DOI: 10.1016/j.jchromb.2009.01.007.
(13) Filzmoser, P.; Walczak, B. *Journal of chromatography. A* **2014**, DOI: 10.1016/j.chroma.2014.08.050.
(14) Walach, J.; Filzmoser, P.; Hron, K., Chapter Seven - Data Normalization and Scaling. In *Comprehensive Analytical Chemistry : Data Analysis for Omic Sciences: Methods and Applications;* Jaumot, Joaquim; Bedia, Carmen; Tauler, Romà, Eds.; Elsevier, 2018.
(15) Yu, H.; Huan, T. *Bioinformatics* **2022**, DOI: 10.1093/bioinformatics/btac355.
(16) Chetwynd, A. J.; Abdul-Sada, A.; Holt, S. G.; Hill, E. M. *Journal of chromatography. A* **2016**, DOI: 10.1016/j.chroma.2015.12.056.
(17) Cook, T.; Ma, Y.; Gamagedara, S. *Journal of pharmaceutical and biomedical analysis* **2020**, DOI: 10.1016/j.jpba.2019.112854.
(18) Vogl, F. C.; Mehrl, S.; Heizinger, L.; Schlecht, I.; Zacharias, H. U.; Ellmann, L.; Nürnberger, N.; Gronwald, W.; Leitzmann, M. F.; Rossert, J.; Eckardt, K.-U.; Dettmer, K.; Oefner, P. J. *Analytical and bioanalytical chemistry* **2016**, DOI: 10.1007/s00216-016-9974-1.
(19) Chen, G.; Liao, H.; Tseng, Y. J.; Tsai, I.; Kuo, C. *Analytica chimica acta* **2015**, DOI: 10.1016/j.aca.2015.01.022.
(20) Nam, S. L.; La Mata, A. P. d.; Dias, R. P.; Harynuk, J. J. *Metabolites* **2020**, DOI: 10.3390/metabo10090376.
(21) Yang, Q.; Bo, L.; Zhu, F., Eds. *Comparison of Standardization Approaches Applied to Metabolomicds Data*; ACM: New York, 2017.
(22) Geisler, C.; Schlicht, K.; Knappe, C.; Rohmann, N.; Hartmann, K.; Türk, K.; Settgast, U.; Schulte, D. M.; Demetrowitsch, T.; Jensen-Kroll, J.; Pisarevskaja, A.; Brix, F.; Gruber, B.; Rimbach, G.; Döring, F.; Rosenstiel, P.; Franke, A.; Schreiber, S.; Henning, C. H. C. A.; Lieb, W.; Nöthlings, U.; Schwarz, K.; Laudes, M. *European journal of epidemiology* **2022**, DOI: 10.1007/s10654-022-00924-y.
(23) Demetrowitsch, T. J.; Petersen, B.; Keppler, J. K.; Koch, A.; Schreiber, S.; Laudes, M.; Schwarz, K. *Bioanalysis* **2015**, DOI: 10.4155/bio.14.270.
(24) Seoudy, A. K.; Schlicht, K.; Kulle, A.; Demetrowitsch, T.; Beckmann, A.; Geisler, C.; Türk, K.; Rohmann, N.; Hartmann, K.; Brandes, J.; Schulte, D. M.; Schreiber, S.; Schwarz, K.; Holterhus, P.-M.; Laudes, M. *Neuroendocrinology* **2023**, DOI: 10.1159/000529146.
(25) Kind, T.; Fiehn, O. *BMC bioinformatics* **2007**, DOI: 10.1186/1471-2105-8-105.
(26) Wishart, D. S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B. L.; Berjanskii, M.; Mah, R.; Yamamoto, M.; Jovel, J.; Torres-Calzada, C.; Hiebert-Giesbrecht, M.; Lui, V. W.; Varshavi, D.; Varshavi, D.; Allen, D.; Arndt, D.; Khetarpal, N.; Sivakumaran, A.; Harford, K.; Sanford, S.; Yee, K.; Cao, X.; Budinski, Z.; Liigand, J.; Zhang, L.; Zheng, J.; Mandal, R.; Karu, N.; Dambrova, M.; Schiöth, H. B.; Greiner, R.; Gautam, V. *Nucleic acids research* **2022**, DOI: 10.1093/nar/gkab1062.
(27) Luan, H.; Ji, F.; Chen, Y.; Cai, Z. *Analytica chimica acta* **2018**, DOI: 10.1016/j.aca.2018.08.002.
(28) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Analytical chemistry* **2006**, DOI: 10.1021/ac051632c.
(29) Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. *Bioinformatics* **2003**, DOI: 10.1093/bioinformatics/19.2.185.
(30) Lin, S. M.; Du, P.; Huber, W.; Kibbe, W. A. *Nucleic acids research* **2008**, DOI: 10.1093/nar/gkm1075.
(31) Workman, C.; Jensen, L. J.; Jarmer, H.; Berka, R.; Gautier, L.; Nielser, H. B.; Saxild, H.-H.; Nielsen, C.; Brunak, S.; Knudsen, S. *Genome biology* **2002**, DOI: 10.1186/gb-2002-3-9-research0048.

(32) Wright, M. N.; Ziegler, A. *J. Stat. Soft.* **2017**, DOI: 10.18637/jss.v077.i01.
(33) Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A. *Analytical chemistry* **2017**, DOI: 10.1021/acs.analchem.7b00096.
(34) Kochhar, S.; Jacobs, D. M.; Ramadan, Z.; Berruex, F.; Fuerholz, A.; Fay, L. B. *Analytical biochemistry* **2006**, DOI: 10.1016/j.ab.2006.02.033.
(35) Psihogios, N. G.; Gazi, I. F.; Elisaf, M. S.; Seferiadis, K. I.; Bairaktari, E. T. *NMR in biomedicine* **2008**, DOI: 10.1002/nbm.1176.
(36) Rasmussen, L. G.; Savorani, F.; Larsen, T. M.; Dragsted, L. O.; Astrup, A.; Engelsen, S. B. *Metabolomics* **2011**, DOI: 10.1007/s11306-010-0234-7.
(37) Rist, M. J.; Roth, A.; Frommherz, L.; Weinert, C. H.; Krüger, R.; Merz, B.; Bunzel, D.; Mack, C.; Egert, B.; Bub, A.; Görling, B.; Tzvetkova, P.; Luy, B.; Hoffmann, I.; Kulling, S. E.; Watzl, B. *PloS one* **2017**, DOI: 10.1371/journal.pone.0183228.
(38) Thévenot, E. A.; Roux, A.; Xu, Y.; Ezan, E.; Junot, C. *Journal of proteome research* **2015**, DOI: 10.1021/acs.jproteome.5b00354.
(39) Trimigno, A.; Khakimov, B.; Savorani, F.; Tenori, L.; Hendrixson, V.; Čivilis, A.; Glibetic, M.; Gurinovic, M.; Pentikäinen, S.; Sallinen, J.; Garduno Diaz, S.; Pasqui, F.; Khokhar, S.; Luchinat, C.; Bordoni, A.; Capozzi, F.; Balling Engelsen, S. *Molecular nutrition & food research* **2019**, DOI: 10.1002/mnfr.201800216.
(40) Wang, T.; Tang, L.; Lin, R.; He, D.; Wu, Y.; Zhang, Y.; Yang, P.; He, J. *Molecular genetics & genomic medicine* **2021**, DOI: 10.1002/mgg3.1738.
(41) Zhang, S.; Liu, L.; Steffen, D.; Ye, T.; Raftery, D. *Metabolomics* **2012**, DOI: 10.1007/s11306-011-0315-2.
(42) Yan, Z.; Yan, R. *Journal of chromatography. A* **2016**, DOI: 10.1016/j.chroma.2016.03.023.
(43) Noonan, M. J.; Tinnesand, H. V.; Buesching, C. D. *BioEssays : news and reviews in molecular, cellular and developmental biology* **2018**, DOI: 10.1002/bies.201700210.