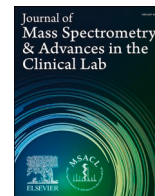




Contents lists available at [ScienceDirect](https://www.sciencedirect.com)
**Journal of Mass Spectrometry and
 Advances in the Clinical Lab**

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



Graphical Review

Small molecule biomarker discovery: Proposed workflow for LC-MS-based clinical research projects



S. Rischke^a, L. Hahnefeld^{a,b}, B. Burla^c, F. Behrens^{b,d}, R. Gurke^{a,b,*}, T.J. Garrett^{e,*}

^a pharmazentrum frankfurt/ZAFES, Institute of Clinical Pharmacology, Johann Wolfgang Goethe University, Theodor Stern-Kai 7, 60590 Frankfurt am Main, Germany

^b Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, and Fraunhofer Cluster of Excellence for Immune Mediated Diseases CIMD, Theodor-Stern-Kai 7, 60596 Frankfurt am Main, Germany

^c Singapore Lipidomics Incubator, Life Sciences Institute, National University of Singapore, Singapore, Singapore

^d Division of Rheumatology, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60596 Frankfurt am Main, Germany

^e Department of Pathology, Immunology and Laboratory Medicine and Southeast Center for Integrated Metabolomics, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:

LC-MS-Based Clinical Research

Biomarker Discovery Study

Lipidomics

Metabolomics

ABSTRACT

Mass spectrometry focusing on small endogenous molecules has become an integral part of biomarker discovery in the pursuit of an in-depth understanding of the pathophysiology of various diseases, ultimately enabling the application of personalized medicine. While LC-MS methods allow researchers to gather vast amounts of data from hundreds or thousands of samples, the successful execution of a study as part of clinical research also requires knowledge transfer with clinicians, involvement of data scientists, and interactions with various stakeholders.

The initial planning phase of a clinical research project involves specifying the scope and design, and engaging relevant experts from different fields. Enrolling subjects and designing trials rely largely on the overall objective of the study and epidemiological considerations, while proper pre-analytical sample handling has immediate implications on the quality of analytical data. Subsequent LC-MS measurements may be conducted in a targeted, semi-targeted, or non-targeted manner, resulting in datasets of varying size and accuracy. Data processing further enhances the quality of data and is a prerequisite for in-silico analysis. Nowadays, the evaluation of such complex datasets relies on a mix of classical statistics and machine learning applications, in combination with other tools, such as pathway analysis and gene set enrichment. Finally, results must be validated before biomarkers can be used as prognostic or diagnostic decision-making tools. Throughout the study, quality control measures should be employed to enhance the reliability of data and increase confidence in the results.

The aim of this graphical review is to provide an overview of the steps to be taken when conducting an LC-MS-based clinical research project to search for small molecule biomarkers.

Introduction

Liquid chromatography – mass spectrometry (LC-MS) aided biomarker discovery is a complex and extensive process that requires interdisciplinary expertise, horizontal knowledge transfer between clinicians, analysts, data scientists, and other stakeholders (e.g., patients), and detailed planning. This process can be divided into five interwoven phases, with quality control (QC) measures supervising the other phases

and facilitating confidence in findings and reproducibility. The other stages include clinical trial (including sample collection and pre-processing), sample preparation (e.g., extraction), LC-MS analyses, data processing, and evaluation. The following graphical review outlines a general workflow for LC-MS-based clinical research projects that are intended to search for small molecule biomarkers or biomarker profiles.

Abbreviations: HILIC, Hydrophilic interaction liquid chromatography; HRMS, High resolution mass spectrometry; LC-MS, Liquid chromatography – mass spectrometry; MRM, Multiple reaction monitoring; PCA, Principal component analysis; QA, Quality assurance; QC, Quality control; RF, Random Forest; RP, Reversed phase; SVA, Support vector machine; (U)HPLC (Ultra-), High pressure liquid chromatography.

* Corresponding authors at: pharmazentrum frankfurt/ZAFES, Institute of Clinical Pharmacology, Johann Wolfgang Goethe University, Theodor Stern-Kai 7, 60590 Frankfurt am Main, Germany (R. Gurke).

E-mail addresses: gurke@med.uni-frankfurt.de (R. Gurke), tgarrrett@ufl.edu (T.J. Garrett).

<https://doi.org/10.1016/j.jmsacl.2023.02.003>

Received 9 November 2022; Received in revised form 14 February 2023; Accepted 14 February 2023

Available online 17 February 2023

2667-145X/© 2023 THE AUTHORS. Publishing services by ELSEVIER B.V. on behalf of MSACL. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Study and trial design

Clinical research projects typically originate from observations of pre-clinical research or epidemiological data that suggest a need for a better understanding of pathophysiological processes, with the goal of improved diagnosis, prediction of progression, or selection of therapy in certain diseases [Fig. 1]. A plan is laid out based on literature, clinical relevance, and feasibility, including patient feedback, and funding is secured from either public organizations or industry [1]. It is important to discuss the implementation of potential biomarker screenings in terms of equipment, expertise, and capacity between end users and researchers early on. An ethics council must then evaluate the study protocol and give their approval. Before the first samples are collected, analytical scientists will develop or select the methods to be used in the project and specify the pre-analytical requirements for the compounds or compound classes of interest. Depending on the research goals, different trial designs should be evaluated, taking into account various strategies for cohort randomization (e.g., simple randomization, block randomization or stratified randomization) and blinding (e.g., double/triple blinding) [2,3]. Meanwhile, subject recruitment and stratification will occur at the clinical site and samples will be taken and processed following predefined standard operating procedures [4]. Comprehensive subject data (e.g., age, gender, known diseases and medication) for later evaluation must be documented, while national and international standards for data privacy must be adhered to. Sampling protocols should carefully regulate sampling to control for pre-analytical variables. Pre-analytical factors are known to distort analyte levels and disrupt meaningful analysis, so emphasis should be put on sample handling that ensures sample integrity [5]. Sample documentation should include unique sample IDs, data on sample collection and the respective participant, using either unique sample-IDs or barcodes. Sample matrices, regulations in shipping biological specimens, sample stability, as well as distance between the clinical and analytical site must all be taken into account when planning the clinical research project [6,7].

Pre-analytics - sampling and sample preprocessing

Sample preprocessing, including the generation of aliquots, should take place at the clinical site(s). Aliquot volumes should be selected appropriately to avoid unwanted freeze–thaw cycles. Further processing may also occur in the analytical laboratory before transferring samples to their final storage or starting the measurement. Special consideration should be placed on preserving sample stability in this phase as well. Analyte concentrations can be altered by simple chemical reactions, such as hydrolysis, enzymatic metabolism, or mass exchange between cells, or even the intrinsic instability of the compound of interest [8]. Therefore, the mode and temperature of intermediate storage, as well as the acceptable amount of processing delay should be predetermined. Sample handling steps for every sample need to be documented for tracking potential quality defects. For instance, in the case of anti-coagulated whole blood samples, the time and temperature until centrifugation and separation of plasma, as well as centrifugation parameters and final storage temperature must be documented [Fig. 2] [9–12]. In cases where the protocol of sample collection is not within the scope of the research (e.g., samples derived from biobanks), sample quality can be assessed retrospectively. Especially for blood specimens, various markers for sample quality defects have been described. For example, hemolysis in blood samples can be assessed by determining the hemolytic index [13], while other metabolomics biomarkers have been proposed for the identification of other pre-analytical risk factors, such as prolonged storage at room temperature [9,14]. Without consensus on the best mode of detection of stability issues, documentation of sample collection remains an important tool for tracking potential quality defects. Aliquoting preprocessed matrices allows for obtaining standardized volumes from individual samples, as well as preparing technical replicates for different measurements and diluting matrices for the process of analyte extraction, which might be especially necessary for highly abundant compounds. The most important reason for generating small-volume aliquots is the prevention of additional freeze/thaw cycles, which can have strong effects on analyte concentration [15,16].

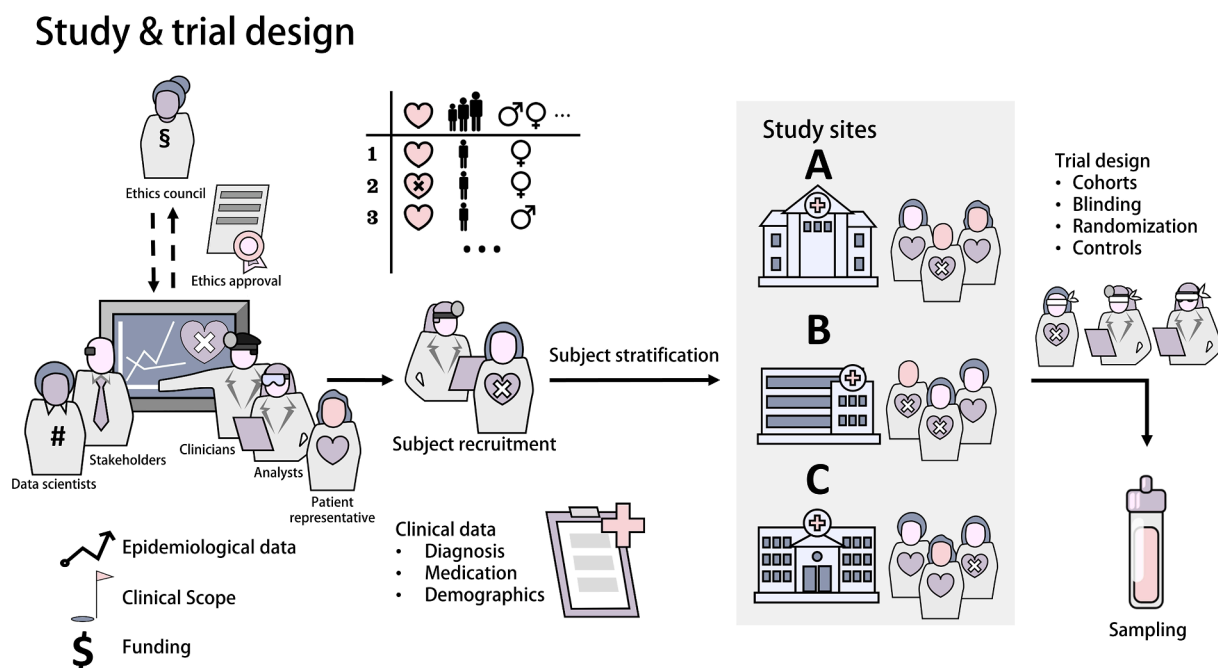


Fig. 1. Before generating any sample, a clinical study for discovering biomarkers requires careful planning, including obtaining ethics approval, defining a research scope involving stakeholders such as clinicians, analysts, data scientists, investors, and patients, and obtaining sufficient funding. Subject recruitment and stratification are instrumental to subsequent evaluations. Documenting meta data such as clinical (e.g., BMI), individual (e.g., smoking habits), and demographic information (e.g., age) on study subjects is highly relevant for trial design, as well as subsequent data analyses. The trial design depends on a multitude of factors, including modes of randomization and blinding, as well as more practical considerations such as sampling plans and study sites.

Sampling & Sample Preprocessing

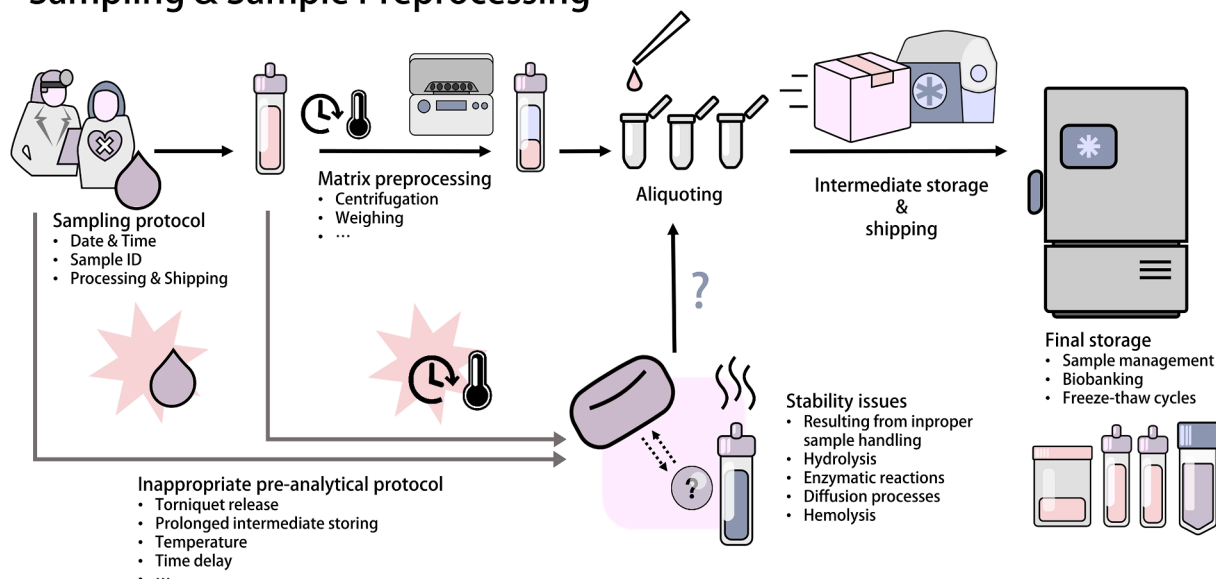


Fig. 2. Sampling should follow a predefined protocol aligned with quality standards for the analytes of focus in the study. The pre-analytical phase in the life cycle of samples has the potential to significantly alter analyte concentrations. To ensure accurate results, well-defined procedures and relevant documentation on sample processing should be in place. Matrix preprocessing should be conducted at the sampling site, as whole blood cannot be stored or transported effectively without significant ex-vivo changes; it must be processed promptly into a plasma sample. When multiple measurements or technical replicates are planned for individual samples, aliquoting is required and should be conducted directly after sampling and before dry-ice shipping and final storage, usually at < -70 °C.

Sample quality impairments, such as evaporation or oxidation, should be considered when generating very small aliquots. The effect of aliquot volume on the duration of thawing and subsequent stability implications should be considered as well [17]. As measurements are not necessarily performed immediately after sample preprocessing, biospecimens should be frozen at storage temperatures below -70 °C. When freeze–thaw cycles are kept to a minimum, deep freezing facilitates long-term storage and collection of sufficiently high sample numbers for performing LC-MS analytics [6]. Sample stability concerns related to the measurement should be pre-conceived during LC-MS method development and validation.

Planning of LC-MS-analysis

LC-MS is able to quantify a wide variety of small molecules, including lipids or polar metabolites. The choice of methods depends largely on research requirements and the accessibility of compounds for analysis. Depending on whether the clinical study aims to answer highly specific or more general research questions, measurement can be conducted in targeted or non-targeted modes. Targeted LC-MS analysis involves the analysis of a predefined set of substances, often related by substance class or biological pathway, whereas non-targeted analysis records a broad set of mass traces that are identified and quantified later and usually result in comprehensive data sets [18]. Targeted analysis is often focused on low-concentration compounds, like lipid mediators, which require specific method optimization. LC-MS/MS is most commonly used for targeted analysis, due to its high sensitivity, selectivity, and robustness. Targeted analysis can also include absolute quantification, which requires the acquisition of reference substances and internal standards that are preferably isotopically labeled and allow for external calibration on an individual analyte level [19]. Absolute quantitative data allows for drawing conclusions regarding physiological reference intervals and simplifies data integration from multiple sites or studies. In the absence of substance-specific reference standards, relative quantification can be applied, either comparing the results to internal standards or to other samples (or reference materials) like a control group [20]. Non-targeted LC-HRMS analysis typically generates

relative quantitative datasets of an extensive number of analytes, but focuses on species of higher abundance, due to a limited dynamic range [21]. Unlike targeted analysis, no predefinition of analytes is required for non-targeted approaches. To identify analytes, inferences are made from retention time and mass traces using previous experimental data and/or MS/MS libraries. To accelerate data evaluation, the analytes reviewed in a non-targeted analysis can be restricted post-acquisition to previously identified compounds, which is often referred to as semi-targeted analysis [22]. While the classical workflow of multiple reaction monitoring (MRM) allows for the adjustment of collision energy and de-clustering potential to optimize ion fragmentation and enhance sensitivity for a specific set of analytes [23], non-targeted approaches rely on a fixed combination of instrument parameters for a, generally, much broader analyte coverage [24]. Yet certain metabolites might still be elusive to the dynamic range of non-targeted setups. MS/MS approaches with a wide range of (scheduled) MRM transitions can serve as a compromise between the more reliable quantification of analytes provided by targeted analyses and the more extensive analyte coverage of non-targeted approaches [25].

Special care should be taken when planning the batch sequence for LC-MS analysis, regardless of whether it is targeted, semi-targeted, or non-targeted [Fig. 3]. Randomizing samples can mitigate the risk of bias due to signal drift [26]. Further considerations should be given to the placement of blank samples, calibration standards, QC samples, long-term control samples, and bridging samples (if applicable). Bridging samples are technical replicates of study samples between batches, which can be used to track the robustness of measurements over multiple batches. The batch size, among other factors, depends on the overall sample size, duration of measurement, and reasonable assumptions of sample stability [27].

The extraction of analytes from sample matrices refers to the isolation of compounds of interest while excluding or minimizing those that could interfere with the measurement. Analyte extraction can take different forms, ranging from protein precipitation, liquid–liquid extraction, to solid-phase extraction [21]. An internal or surrogate standard, which is distinguishable from the analyte yet behaves similarly in measurement (e.g., an isotopically labeled variant), is processed

Planning of quantitative and relative quantitative LC-MS analytics

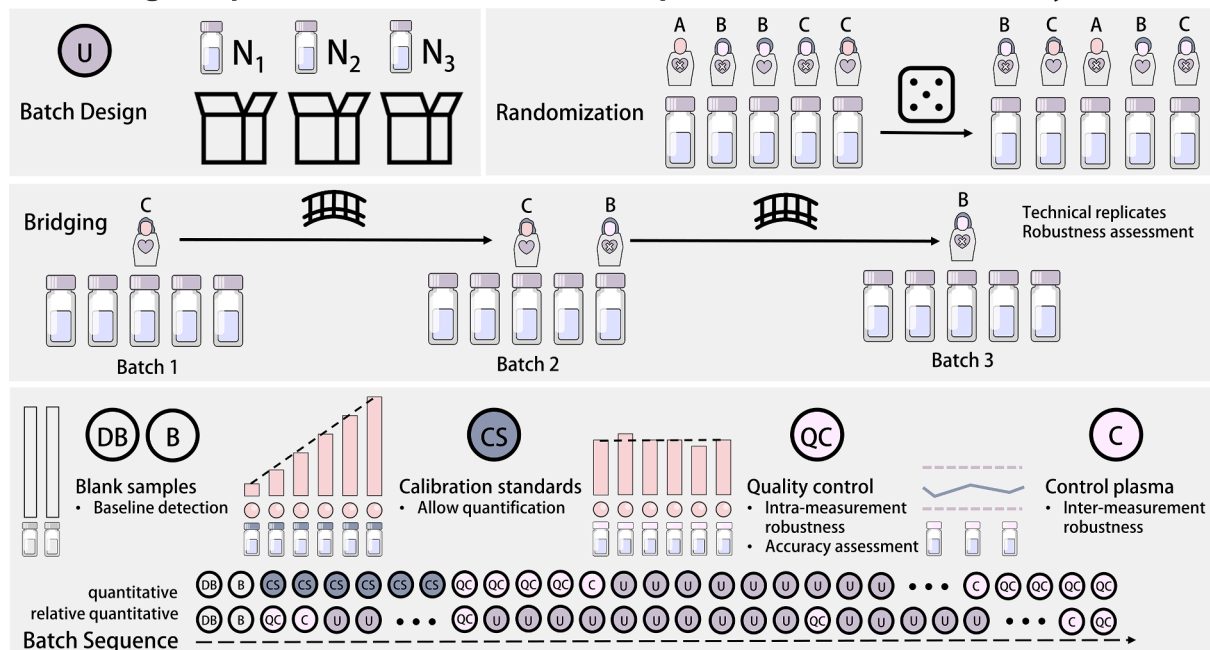


Fig. 3. Meticulous planning of LC-MS measurements avoids pitfalls, like selection biases, and can be instrumental in quality control of the measurement. Initial batch design refers to dividing samples into different analytical runs. Different kinds of randomization can come into play, when aiming to mitigate the risk of individual samples being subjected to signal drift. While simple randomization is suitable for independent samples, a stratified cluster randomization might be more suitable for samples with nested dependence (e.g. time series data for individual subjects). Bridging refers to deliberately measuring technical replicates of samples, in order to connect batches by some proportion of shared measurements, allowing for comparability and assessment of robustness. The batch sequence (e.g., order of sample types) is subject to the mode of measurement. Depending on whether the measurement is absolute quantitative or relative quantitative, different combinations of blank samples, inter- and intra-measurement quality control samples, calibration standards and subject related unknown samples need to be considered. (Used abbreviations: U = “Sample of unknown concentration”; DB = “Double blank (without IS)”; B = “Blank (with IS)”; CS = “Calibration standard”; QC = “Quality control”; C = “Control plasma”).

within the same sample. Since the amount of internal standards is held constant across samples, it is suitable for assessing analytical variability or matrix effects on a sample-by-sample basis [28]. In addition to samples, calibration standards (in the case of quantitative analysis), quality control (QC) samples, and long-term control samples are also prepared.

Targeted LC-MS-analysis

In the case of targeted and absolute quantitative LC-MS analytics, calibration standards and QC samples at different concentration levels are necessary [29]. Quantitative analysis is used to compare compound-specific calibration standards to an internal standard in order to enable the calculation of absolute analyte concentrations and to ensure confidence in compound identification. Retention time, fragmentation patterns, and peak shape of reference standards allow for drawing conclusions on analyte behavior. It is ideal to use isotopically labeled internal standards as analogues of the compound of interest. Calibration standards are derived by diluting and extracting reference substances in an analyte-free matrix to create a reference frame of concentrations (calibration curve), which can then be compared to samples of unknown concentration in order to calculate their relative concentrations [Fig. 3] [30]. Reference standard-based QC samples are processed with known concentrations spanning the range of measurable concentrations defined by the calibration curve, allowing for an assessment of accuracy and measurement system stability. In targeted analysis, these QC samples are usually placed at the beginning, intermittently at regular intervals, and at the end of the batch sequence so their accuracy and coefficient of variance can be used to determine measurement robustness and identify (and possibly correct for) drift effects within the batch [31]. Long-term control samples are similarly used as other QC samples, but concentration levels of technical replicates can be tracked in control charts over

multiple measurements of the same or different studies in order to assess consistency of the LC-MS setup [32]. Thus, while regular QC samples serve as a control within the measurement, long-term control samples serve as a quality control method for the whole analytical process. Bridging samples are also technical replicates of individual samples that must be planned and prepared during sample aliquoting; one sample from each pair is measured in the batch following the first one. This is especially useful for assessing technical variations (i.e., batch effects) across multiple batches within the study. The selection for bridging samples should be done preemptively and randomized [33]. After completing the planning phase of the measuring sequence, extraction of analytes can begin. The measurement method, LC-column, and solvents should be specific to the chemical properties of the investigated analytes [Fig. 4]. Reversed phase (RP) columns are considered standard for most separation methods and are especially suitable for separation of lipophilic substances; however, polar metabolites may require Hydrophilic Interaction Liquid Chromatography (HILIC) phases in order to achieve suitable selectivity [35]. For targeted analysis, tandem mass spectrometers are often used due to their high selectivity (and wide linear range) and sensitivity, which is well-suited for analyzing a predefined set of analytes. While many targeted methods use absolute quantification, targeted screening methods also allow for analyzing hundreds of analytes without reference standards using relative quantification and a measurement setup similar to non-targeted methods [36].

Semi- and Non-targeted LC-MS-analysis

In non-targeted analysis, features with unique combinations of mass-to-charge and retention time are aligned, annotated, and compared based on relative quantification. High resolution mass spectrometry is typically used to allow for compound annotation or identification based

Sample preparation & LC-MS analytics

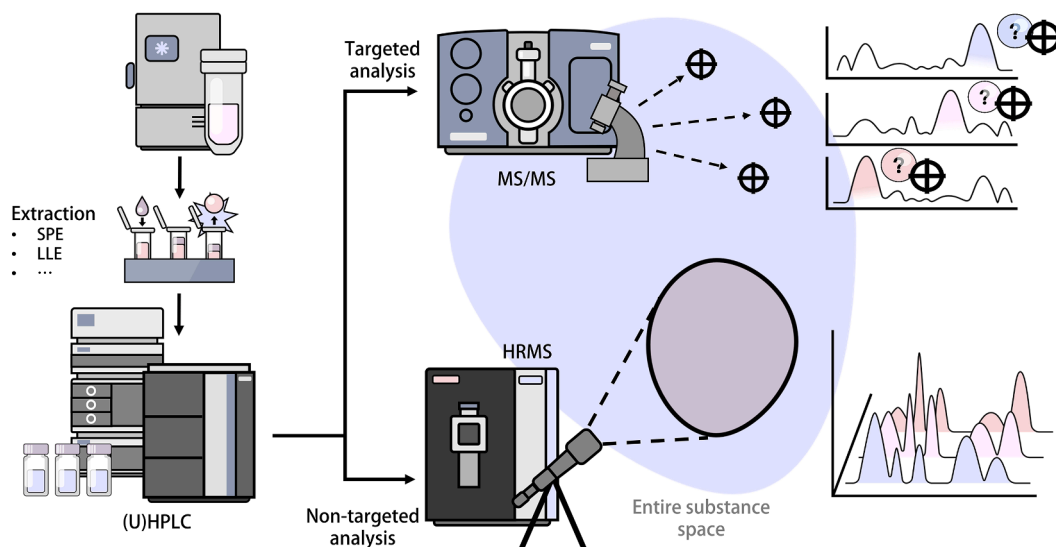


Fig. 4. The process of LC-MS analysis begins with the extraction of analytes that are specific to the class. Each extraction method includes one or more purification steps. Internal or surrogate standards are added to the samples, which are processed alongside the analytes in question. This allows the signal of the analytes of interest to be related to the surrogate signal, as they are all subject to the same influences during extraction, such as matrix effects and other measurement-specific variations in sensitivity. In an often used setup, (U)HPLC is used to separate analytes, while mass spectrometry facilitates their detection. Targeted analysis using tandem mass spectrometry is limited to a predefined set of analytes, but, when combined with matching reference standards, it allows for absolute quantification. Non-targeted analysis using high resolution mass spectrometry has fewer restrictions on what analytes can be detected, allowing for simultaneous analysis of hundreds or thousands of features. However, without compound-specific reference standards, quantification is only relative. Semi-targeted analysis is a compromise between these two methods, usually focusing on hundreds of well-characterized analytes while granting broad insights at reduced processing times. As reference standards are limited in availability and economical viability, semi-targeted analysis is mostly only relative quantitative.

on exact mass, isotopic distribution, (relative) retention time, and by comparison of MS/MS fragmentation patterns to reference spectra [24]. This expands the range of potentially observable analytes compared to targeted analysis; however, it also increases the chance of false identification. Quantification is limited by the fact that many analytes are related to only few surrogate internal standards or relative sample concentrations are calculated in relation to all other samples. The signal intensity of compounds can vary based on carbon chain length, the number of double bonds, and structural differences [37]. The amount of relatively quantified features can easily reach into the thousands; therefore, internal standards cannot be selected on an individual analyte level. Absolute quantitative analysis with reference standard calibration is also not suitable for the same reasons. Long-term control samples can still be used for tracking system performance. QC samples, used for assessing measuring robustness within an analysis run, also do not require a range of concentrations; rather, they are replicates from a representative sample pool distributed over the batch at regular intervals (e.g., every 8 to 20 samples).

Data integration and processing

Signal interpretation uses specialized software, either specific to the manufacturer or offered by third-party vendors, to identify and integrate signal peaks for each analyte and its respective internal standard. This is instrumental in calculating the concentration, as it relies on the area under the curve. The evaluation software usually automates significant proportions of peak annotation and integration when referencing a database of earlier measurements [38]. Exceptions include non-targeted analysis, where it might not be possible to find a suitable internal standard for all features, yet surrogates are still used to verify technical variation. Absolute concentrations are calculated using a calibration curve [30] and values that deviate from the line should be deemed indistinct and may be imputed in later steps. QC samples are utilized to evaluate measurement robustness, a key parameter in biomarker

discovery [39]. These concentrations should align with other QC samples (e.g., pooled QC samples), and their accuracy and coefficient of variance serves as a measure of quality for the analysis run [40,41]. Despite this, the resulting dataset might still be skewed and can be transformed further depending on its planned application [Fig. 5]. Missing values can be imputed, although the choice of method usually requires knowledge as to why the values are missing [42]. Erratically missing values may be imputed by more sophisticated methods like kNN or Random Forest Imputation [43], while values that fall below the lower limit of quantification can often be imputed with half of said limit, thus accounting for their approximate scale. Additionally, sample data can be normalized and scaled by various methods in order to remove systematic and statistical biases. Transformations towards normal distribution of measurements can be performed to satisfy the assumptions of statistical and machine learning models [34]. Normalization techniques include approaches of general applicability that aim for comparability of samples (e.g., Probabilistic Quotient Normalization [44] or Quantile Normalization [45]), as well as techniques that target specific biases like batch effect correction methods (e.g., LOESS normalization [46], SVA [47] or ComBat [48,49]). HRMS datasets usually take up large amounts of hard drive storage (e.g., dozens of gigabytes per 100 samples), and have very high complexity (e.g., randomization, data hierarchy, time series). Therefore, data management and archiving in LC-MS, as well as other study data, is crucial for traceability of the study.

Data analysis and Clinical translation

Data analysis specific to the clinical scope of the study requires domain knowledge when pursuing specific questions. Before cohorts are evaluated, data sets may need to be adjusted for covariates (e.g., age distribution among groups) to avoid confounding [50,51]. A mix of classical statistics and machine learning applications is employed for data investigation [Fig. 6] [52], with classical statistics focusing on explaining phenomena, while machine learning methods focus on

Data Integration & Processing

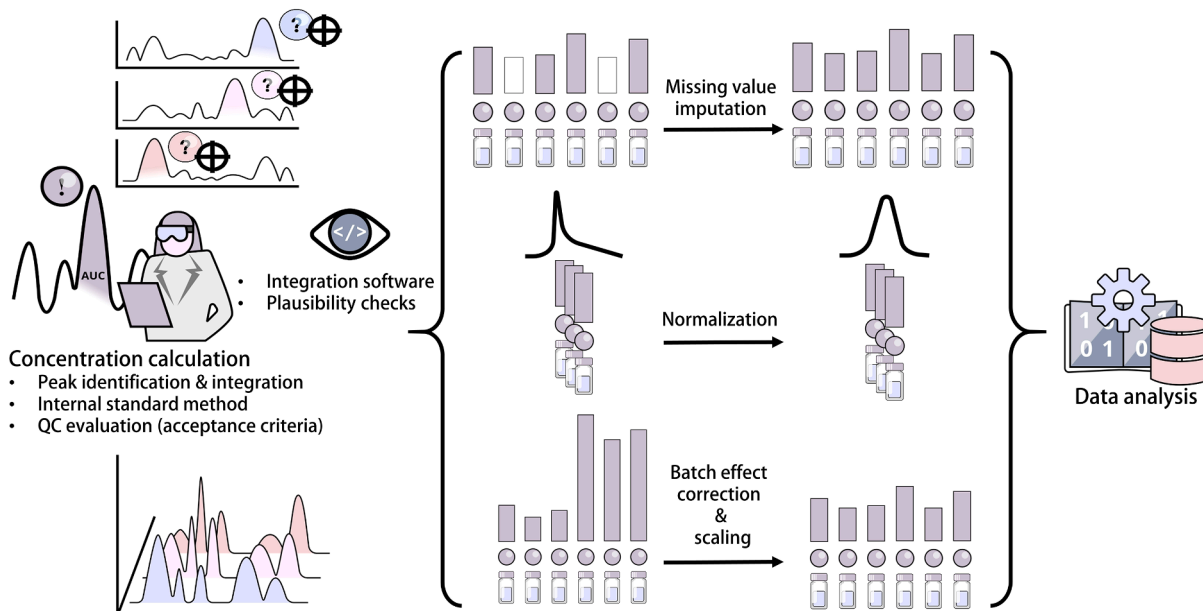


Fig. 5. Peak identification depends on the mass traces and retention time of the respective analytes. Peak integration enables relative quantification. Relative quantification (non-targeted analysis) compares the peak areas of multiple analytes with one or multiple internal standards or a reference sample, and does not use a calibration curve of reference standards, which is necessary for targeted measurements. Further data processing might be needed after obtaining the results. Missing values must be imputed, outliers potentially removed, samples may need to be scaled for comparability, normalized to meet the assumptions of statistical models, and batch effects must be identified and corrected retrospectively.

Data Analysis

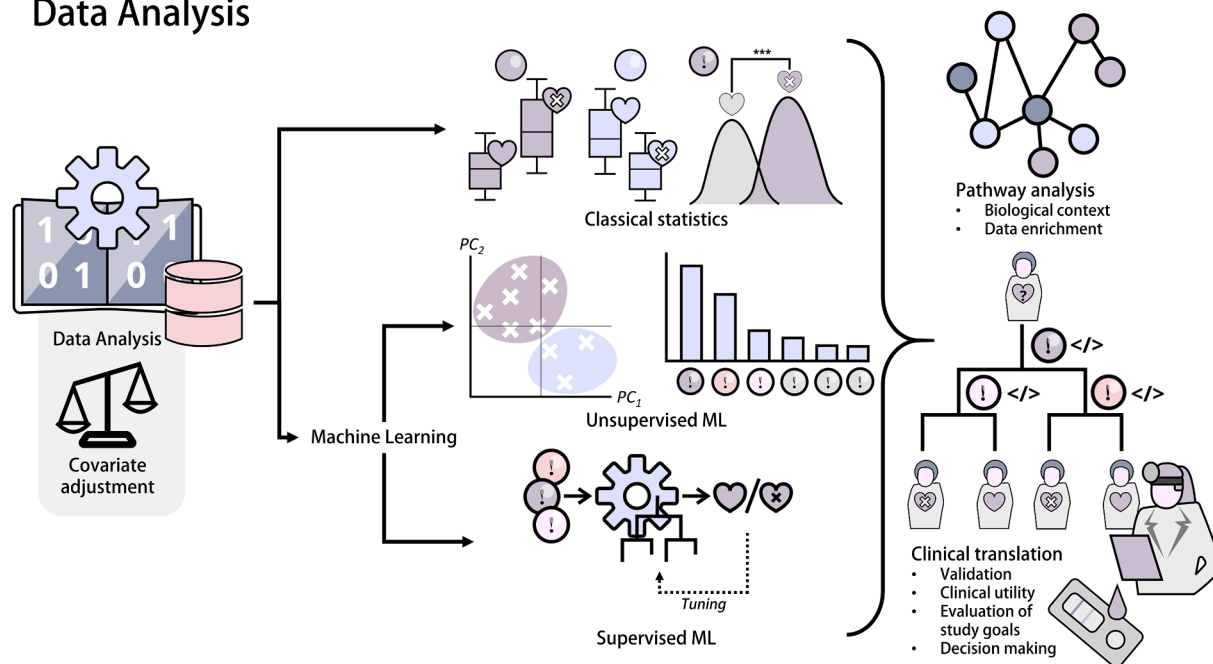


Fig. 6. Data analysis can take place between the fields of classical statistics and machine learning, which often overlap. Statistics obtain p-values, usually in uni- or bivariate analysis, to claim significance, while machine learning approaches strive for the best possible test metrics, such as optimal prediction accuracy. Machine learning can be further divided into unsupervised and supervised approaches. Unsupervised approaches act independently of pre-defined sample groups to develop hypotheses and explore data (although inferences are also possible). In contrast, supervised models usually aim for optimal regression or classification in the context of the research scope. Therefore, both supervised models and classical statistics can be used to confirm or reject hypotheses. Pathway analyses can integrate the obtained data into a broader biological context. Finally, resulting findings and potential biomarkers or biomarker panels need to be clinically validated, demonstrate clinical benefit, and be practical in order to be translated into clinical routine.

prediction [53]. Unsupervised machine learning algorithms offer a general way of learning underlying structures of datasets; principal component analysis (PCA) and hierarchical clustering are instrumental

in data exploration. Besides information on patterns within the data, they are able to offer techniques of variable selection (i.e., dimensionality reduction), which is crucial in identifying potential biomarkers

from hundreds or thousands of measured analytes. Supervised machine learning algorithms can put selected variables to the test and rely on powerful regression and classification algorithms, like the Random Forest (RF) or Support Vector Machines (SVM) [54]. Model validation, such as Cross-validation, must be facilitated to assess supervised machine learning model performance and exclude model overfitting [55]. Optimization of supervised machine learning models is usually performed through a process of model parameter tuning and model assessment [54]. Pathway analysis helps to embed findings in a broader biological context and form hypotheses and make causal assumptions [56].

Besides analytical and clinical validity, biomarkers or biomarker panels need to prove their utility by being practical and broadly applicable in a clinical context. However, validating biomarkers poses methodological and general challenges. A common problem in initial biomarker research is the lack of power due to small sample sizes, which increases the risk for confounders and reduces the applicability of machine learning methods. A high number of potential biomarkers, generated from a small sample size can thus be a result of overfitting [57]. Awareness of this is of particular interest, as there is no consensus on the best way of feature selection and machine learning methods are often selected for isolating a set of variables [58]. Furthermore, metabolites are especially prone to vary due to various covariates, such as age, gender, and diet. Therefore, if feature selection steps result in a high number of analytes, special caution needs to be taken to avoid such biases (e.g., by stratifying the cohort properly). Additionally, due to the complexity of many diseases and subsequent systemic changes, individual biomarkers are unlikely to emerge for diagnostic or prognostic purposes [59]. To overcome this issue, validation calls for increasing the sample size with an independent cohort to confirm selectivity, stability,

and repeatability of the measurement and raise statistical power [60]. Moreover, there needs to be an emphasis on the utility of potential biomarkers for their end users throughout the research process. Insufficient interdisciplinary exchange can obscure the lack of clinical relevance, which is not only influenced by predictive power, but also by cost-effectiveness in the clinical laboratory (in terms of instruments, work space, expertise, and staff allocation) and benefit for physicians and their patients in making therapy decisions [61]. While many biomarkers fail at this stage, those that are successful can improve clinical decision making and/or elucidate novel therapeutic targets [46,47].

Quality control & quality assurance

Published potential biomarkers found in LC-MS studies routinely fail in clinical translation, due to inadequate reporting, low quality control, low reproducibility and overfitting of models to available data sets [62]. Other issues that hinder the implementation of valuable biomarkers include biases in cohort selection and evaluation, as well as insufficient independent validation of research results. Quality management measures should be put in place to provide quality control (QC) and quality assurance (QA) throughout the entire research process, so as to improve the reliability and reproducibility of the data and experiments [Fig. 7]. This can be achieved by providing specific instructions for each step of the analytical process and requiring frequent documentation, which will enable traceability [36]. Documentation for quality control purposes should also include recording metadata that describe the study results. To enhance reproducibility across studies, national and international guidelines should be developed to harmonize quality control protocols [63].

Quality control & Quality assurance

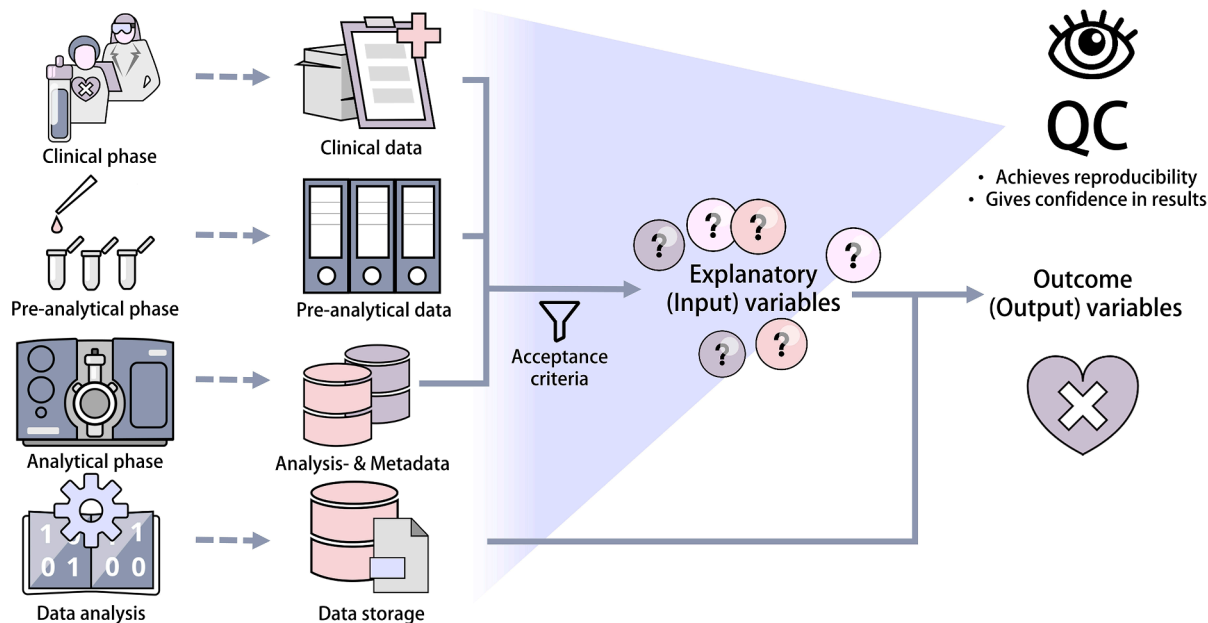


Fig. 7. Quality control and quality assurance accompany the entire process of clinical research projects. Standard operating procedures, as well as good documentation practices, are essential for quality assurance in all stages of research. Data management of clinical data, metadata, data on (pre-)processing, analysis, and evaluation of results allows for traceability across the entire study. Clinical data, documentation on sample processing, analytical phases, and detailed reporting on the data analysis facilitate comprehension of the results and are necessary for reproducibility and validity of the study.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft Sonderforschungsbereich SFB 1039/Z01 “Krankheitsrelevante Signaltransduktion durch Fettsäurederivate und Sphingolipide”, by Fraunhofer Cluster of Excellence for Immune Mediated diseases CIMD and the HIPPOCRATES project. HIPPOCRATES has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement no. 101007757. The JU receives support from the European Union’s Horizon 2020 research and innovation program and EFPIA. BB is supported by grants from the National University of Singapore via the Life Sciences Institute.

Declaration of Competing Interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmsacl.2023.02.003>.

References

- [1] Bhandari, Mohit M.D., M.Sc., F.R.C.S.C.*, Schemitsch, Emil H. M.D., F.R.C.S.C.†, Planning a Randomized Clinical Trial - An Overview, *Techniques in Orthopaedics* 19 (2004) 66–71.
- [2] K. Broglio, Randomization in Clinical Trials: Permuted Blocks and Stratification, *J. Am. Med. Assoc.* 219 (2018) 2223–2224, <https://doi.org/10.1001/jama.2018.6360>.
- [3] A. Hróbjartsson, I. Boutron, Blinding in randomized clinical trials: imposed impartiality, *Clin. Pharmacol. Ther.* 90 (2011) 732–736, <https://doi.org/10.1038/clpt.2011.207>.
- [4] R.L. Schmidt, R.E. Factor, Understanding sources of bias in diagnostic accuracy studies, *Arch. Pathol. Lab. Med.* 137 (2013) 558–565, <https://doi.org/10.5858/arpa.2012-0198-RA>.
- [5] R. Lehmann, From bedside to bench-practical considerations to avoid pre-analytical pitfalls and assess sample quality for high-resolution metabolomics and lipidomics analyses of body fluids, *Anal. Bioanal. Chem.* (2021), <https://doi.org/10.1007/s00216-021-03450-0>.
- [6] G. Lippi, J.J. Chance, S. Church, P. Dazzi, R. Fontana, D. Giavarina, K. Grankvist, W. Huisman, T. Kouri, V. Palicka, M. Plebani, V. Puro, G.L. Salvagno, S. Sandberg, K. Sikaris, I. Watson, A.K. Stankovic, A.-M. Simundic, Preanalytical quality improvement: from dream to reality, *Clin. Chem. Lab. Med.* 49 (2011) 1113–1126, <https://doi.org/10.1515/CCLM.2011.600>.
- [7] T. Hyötyläinen, M. Oresić, Bioanalytical techniques in nontargeted clinical lipidomics, *Bioanalysis* 8 (2016) 351–364, <https://doi.org/10.4155/bio.15.244>.
- [8] S. Nishiumi, M. Suzuki, T. Kobayashi, M. Yoshida, Differences in metabolite profiles caused by pre-analytical blood processing procedures, *J. Biosci. Bioeng.* 125 (2018) 613–618, <https://doi.org/10.1016/j.jbiosc.2017.11.011>.
- [9] B. Kamlage, S.G. Maldonado, B. Bethan, E. Peter, O. Schmitz, V. Liebenberg, P. Schatz, Quality markers addressing preanalytical variations of blood and plasma processing identified by broad and targeted metabolite profiling, *Clin. Chem.* 60 (2014) 399–412, <https://doi.org/10.1373/clinchem.2013.211979>.
- [10] D. Vuckovic, Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry, *Anal. Bioanal. Chem.* 403 (2012) 1523–1548, <https://doi.org/10.1007/s00216-012-6039-y>.
- [11] G. Liebisch, *Biochimica et biophysica acta*, *BBA* 62 (2017) 636, [https://doi.org/10.1016/0006-3002\(62\)90265-2](https://doi.org/10.1016/0006-3002(62)90265-2).
- [12] S. Naz, D.C. Moreira dos Santos, A. García, C. Barbas, Analytical protocols based on LC-MS, GC-MS and CE-MS for nontargeted metabolomics of biological tissues, *Bioanalysis* 6 (2014) 1657–1677, <https://doi.org/10.4155/bio.14.119>.
- [13] G. Lippi, E.J. Favaloro, M. Franchini, Haemolysis index for the screening of intravascular haemolysis: a novel diagnostic opportunity? *Blood Transfus.* 16 (2018) 433–437, <https://doi.org/10.2450/2018.0045-18>.
- [14] X. Liu, M. Hoene, P. Yin, L. Fritsche, P. Plomgaard, J.S. Hansen, C.T. Nakas, A. M. Niess, J. Hudemann, M. Haap, M. Mendy, C. Weigert, X. Wang, A. Fritsche, A. Peter, H.-U. Häring, G. Xu, R. Lehmann, Quality Control of Serum and Plasma by Quantification of (4E,14Z)-Sphingadienine-C18-1-Phosphate Uncovers Common Preanalytical Errors During Handling of Whole Blood, *Clin. Chem.* 64 (2018) 810–819, <https://doi.org/10.1373/clinchem.2017.277905>.
- [15] O. Fliniaux, G. Gaillard, A. Lion, D. Cailleu, F. Mesnard, F. Betsou, Influence of common preanalytical variations on the metabolic profile of serum samples in biobanks, *J. Biomol. NMR* 51 (2011) 457–465, <https://doi.org/10.1007/s10858-011-9574-5>.
- [16] S. Cuhadar, M. Koseoglu, A. Atay, A. Dirican, The effect of storage time and freeze-thaw cycles on the stability of serum samples, *Biochem. Med. (Zagreb)* 23 (2013) 70–77, [10.11613/bm.2013.009](https://doi.org/10.11613/bm.2013.009).
- [17] C. Pizarro, I. Arenzana-Rámila, N. Pérez-del-Notario, P. Pérez-Matute, J. M. González-Sáiz, Thawing as a critical pre-analytical step in the lipidomic profiling of plasma samples: New standardized protocol, *Anal. Chim. Acta* 912 (2016) 1–9, <https://doi.org/10.1016/j.aca.2016.01.058>.
- [18] P. Reveglia, C. Paolillo, G. Ferretti, A. de Carlo, A. Angiolillo, R. Nasso, M. Caputo, C. Matrone, A. Di Costanzo, G. Corso, Challenges in LC-MS-based metabolomics for Alzheimer’s disease early detection: targeted approaches versus untargeted approaches, *Metabolomics* 17 (2021) 78, <https://doi.org/10.1007/s11306-021-01828-w>.
- [19] H.C. Köfeler, T.O. Eichmann, R. Ahrends, J.A. Bowden, N. Danne-Rasche, E. A. Dennis, M. Fedorova, W.J. Griffiths, X. Han, J. Hartler, M. Holčápek, R. Jirásko, J.P. Koelmel, C.S. Ejsing, G. Liebisch, Z. Ni, V.B. O’Donnell, O. Quehenberger, D. Schwudke, A. Shevchenko, M.J.O. Wakelam, M.R. Wenk, D. Wolrab, K. Ekroos, Quality control requirements for the correct annotation of lipidomics data, *Nat. Commun.* 12 (2021) 4771, <https://doi.org/10.1038/s41467-021-24984-y>.
- [20] A. Triebel, B. Burla, J. Selvalatchmanan, J. Oh, S.H. Tan, M.Y. Chan, N.A. Mellet, P. J. Meikle, F. Torta, M.R. Wenk, Shared reference materials harmonize lipidomics across MS-based detection platforms and laboratories, *J. Lipid Res.* 61 (2020) 105–115, <https://doi.org/10.1194/jlr.D119000393>.
- [21] D. Petras, I. Koester, R. Da Silva, B.M. Stephens, A.F. Haas, C.E. Nelson, L.W. Kelly, L.I. Aluwihare, P.C. Dorrestein, High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter, *Front. Mar. Sci.* 4 (2017) 405, <https://doi.org/10.3389/fmars.2017.00405>.
- [22] J.M. Onorato, R. Langish, A. Bellamine, P. Shipkova, Applications of HILIC for targeted and non-targeted LC/MS analyses in drug discovery, *J. Sep. Sci.* 33 (2010) 923–929, <https://doi.org/10.1002/jssc.200900659>.
- [23] J. Zhou, Y. Yin, Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry, *Analyst* 141 (2016) 6362–6373, <https://doi.org/10.1039/c6an01753c>.
- [24] A. Ribbenstedt, H. Ziarrusta, J.P. Benskin, Development, characterization and comparison of targeted and non-targeted metabolomics methods, *PLoS One* 13 (2018) e0207082.
- [25] N. Munjoma, G. Isaac, A. Muazzam, O. Cexus, F. Azhar, H. Pandha, A.D. Whetton, P.A. Townsend, I.D. Wilson, L.A. Gethings, R.S. Plumb, High Throughput LC-MS Platform for Large Scale Screening of Bioactive Polar Lipids in Human Plasma and Serum, *J. Proteome Res.* 21 (2022) 2596–2608, <https://doi.org/10.1021/acs.jproteome.2c00297>.
- [26] P. Jonsson, A. Wuolikainen, E. Thysell, E. Chorell, P. Stattin, P. Wikström, H. Antti, Constrained randomization and multivariate effect projections improve information extraction and biomarker pattern discovery in metabolomics studies involving dependent samples, *Metabolomics* 11 (2015) 1667–1678, <https://doi.org/10.1007/s11306-015-0818-3>.
- [27] A.M. Evans, C. O’Donovan, M. Playdon, C. Beecher, R.D. Beger, J.A. Bowden, D. Broadhurst, C.B. Clish, S. Dasari, W.B. Dunn, J.L. Griffin, T. Hartung, P.-C. Hsu, T. Huan, J. Jans, C.M. Jones, M. Kachman, A. Kleensang, M.R. Lewis, M.E. Monge, J.D. Mosley, E. Taylor, F. Tayyari, G. Theodoridis, F. Torta, B.K. Ubhi, D. Vuckovic, Dissemination and analysis of the quality assurance (QA) and quality control (QC) practices of LC-MS based untargeted metabolomics practitioners, *Metabolomics* 16 (2020) 113, <https://doi.org/10.1007/s11306-020-01728-5>.
- [28] E.H. Wilkes, M.J. Whitlock, E.L. Williams, A data-driven approach for the detection of internal standard outliers in targeted LC-MS/MS assays, *J. Mass Spectrom. Adv. Clin. Lab* 20 (2021) 42–47, <https://doi.org/10.1016/j.jmsacl.2021.06.001>.
- [29] A.H.P. America, J.H.G. Cordewener, Comparative LC-MS: a landscape of peaks and valleys, *Proteomics* 8 (2008) 731–749, <https://doi.org/10.1002/pmic.200700694>.
- [30] Y. Zhao, G. Liu, J.X. Shen, A.-F. Aubry, Reasons for calibration standard curve slope variation in LC-MS assays and how to address it, *Bioanalysis* 6 (2014) 1439–1443, <https://doi.org/10.4155/bio.14.71>.
- [31] C.J. Briscoe, M.R. Stiles, D.S. Hage, System suitability in bioanalytical LC/MS/MS, *J. Pharm. Biomed. Anal.* 44 (2007) 484–491, <https://doi.org/10.1016/j.jpba.2007.03.003>.
- [32] D. Broadhurst, R. Goodacre, S.N. Reinke, J. Kuligowski, I.D. Wilson, M.R. Lewis, W. B. Dunn, Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, *Metabolomics* 14 (2018) 72, <https://doi.org/10.1007/s11306-018-1367-3>.
- [33] O.D. Buhule, R.L. Minster, N.L. Hawley, M. Medvedovic, G. Sun, S. Viali, R. Deka, S.T. McGarvey, D.E. Weeks, Stratified randomization controls better for batch effects in 450K methylation analysis: a cautionary tale, *Front. Genet.* 5 (2014) 354, <https://doi.org/10.3389/fgene.2014.00354>.
- [34] K. Ortmayr, T.J. Causon, S. Hann, G. Koellensperger, Increasing selectivity and coverage in LC-MS based metabolome analysis, *TrAC Trends Anal. Chem.* 82 (2016) 358–366, <https://doi.org/10.1016/j.trac.2016.06.011>.
- [35] E.-M. Harrieder, F. Kretschmer, S. Böcker, M. Witting, Current state-of-the-art of separation methods used in LC-MS based metabolomics and lipidomics, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1188 (2022), 123069, <https://doi.org/10.1016/j.jchromb.2021.123069>.
- [36] B. Burla, M. Arita, M. Arita, A.K. Bendt, A. Cazenave-Gassiot, E.A. Dennis, K. Ekroos, X. Han, K. Ikeda, G. Liebisch, M.K. Lin, T.P. Loh, P.J. Meikle, M. Oresić, O. Quehenberger, A. Shevchenko, F. Torta, M.J.O. Wakelam, C.E. Wheelock, M. R. Wenk, MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines, *J. Lipid Res.* 59 (2018) 2001–2017, <https://doi.org/10.1194/jlr.S087163>.

- [37] E.L. Schymanski, H.P. Singer, J. Slobodnik, I.M. Ipolyi, P. Oswald, M. Krauss, T. Schulze, P. Haglund, T. Letzel, S. Grosse, N.S. Thomaidis, A. Bletsou, C. Zwiener, M. Ibáñez, T. Portolés, R. de Boer, M.J. Reid, M. Onghena, U. Kunkel, W. Schulz, A. Guillon, N. Noyon, G. Leroy, P. Bados, S. Bogialli, D. Stipančev, P. Rostkowski, J. Hollender, Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis, *Anal. Bioanal. Chem.* 407 (2015) 6237–6255, <https://doi.org/10.1007/s00216-015-8681-7>.
- [38] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, *TrAC Trends Anal. Chem.* 82 (2016) 425–442, <https://doi.org/10.1016/j.trac.2016.07.004>.
- [39] H.G. Gika, G.A. Theodoridis, M. Earll, I.D. Wilson, A QC approach to the determination of day-to-day reproducibility and robustness of LC-MS methods for global metabolite profiling in metabolomics/metabolomics, *Bioanalysis* 4 (2012) 2239–2247, <https://doi.org/10.4155/bio.12.212>.
- [40] M. Kaza, M. Karaźniewicz-Lada, K. Kosicka, A. Siemiątkowska, P.J. Rudzki, Bioanalytical method validation: new FDA guidance vs. EMA guideline. Better or worse? *J. Pharm. Biomed. Anal.* 165 (2019) 381–385, <https://doi.org/10.1016/j.jpba.2018.12.030>.
- [41] M. Ghorasaini, Y. Mohammed, J. Adamski, L. Bettcher, J.A. Bowden, M. Cabruja, K. Contrepolis, M. Ellenberger, B. Gajera, M. Haid, D. Hornburg, C. Hunter, C. M. Jones, T. Klein, O. Mayboroda, M. Mirzaian, R. Moaddel, L. Ferrucci, J. Lovett, K. Nazir, M. Pearson, B.K. Ubhi, D. Raftery, F. Riols, R. Sayers, E.J.G. Sijbrands, M. P. Snyder, B. Su, V. Velagapudi, K.J. Williams, Y.B. de Rijke, M. Giera, Cross-Laboratory Standardization of Preclinical Lipidomics Using Differential Mobility Spectrometry and Multiple Reaction Monitoring, *Anal. Chem.* 93 (2021) 16369–16378, <https://doi.org/10.1021/acs.analchem.1c02826>.
- [42] W.-C. Lin, C.-F. Tsai, Missing value imputation: a review and analysis of the literature (2006–2017), *Artif. Intell. Rev.* 53 (2020) 1487–1509, <https://doi.org/10.1007/s10462-019-09709-4>.
- [43] R. Wei, J. Wang, M. Su, E. Jia, S. Chen, T. Chen, Y. Ni, Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data, *Sci. Rep.* 8 (2018) 663, <https://doi.org/10.1038/s41598-017-19120-0>.
- [44] R. Di Guida, J. Engel, J.W. Allwood, R.J.M. Weber, M.R. Jones, U. Sommer, M. R. Viant, W.B. Dunn, Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling, *Metabolomics* 12 (2016) 93, <https://doi.org/10.1007/s11306-016-1030-9>.
- [45] J. Lee, J. Park, M. Lim, S.J. Seong, J.J. Seo, S.M. Park, H.W. Lee, Y.-R. Yoon, Quantile normalization approach for liquid chromatography-mass spectrometry-based metabolomic data from healthy human volunteers, *Anal. Sci.* 28 (2012) 801–805, <https://doi.org/10.2116/analsci.28.801>.
- [46] M. Rusilowicz, M. Dickinson, A. Charlton, S. O'Keefe, J. Wilson, A batch correction method for liquid chromatography-mass spectrometry data that does not depend on quality control samples, *Metabolomics* 12 (2016) 56, <https://doi.org/10.1007/s11306-016-0972-2>.
- [47] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, *Bioinformatics* 28 (2012) 882–883, <https://doi.org/10.1093/bioinformatics/bts034>.
- [48] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, *Biostatistics* 8 (2007) 118–127, <https://doi.org/10.1093/biostatistics/kxj037>.
- [49] Y. Zhao, L. Wong, W.W.B. Goh, How to do quantile normalization correctly for gene expression data analyses, *Sci. Rep.* 10 (2020) 15534, <https://doi.org/10.1038/s41598-020-72664-6>.
- [50] J. Witte, V. Didelez, Covariate selection strategies for causal inference: Classification and comparison, *Biom. J.* 61 (2019) 1270–1289, <https://doi.org/10.1002/bimj.201700294>.
- [51] A. Srivastava, D.J. Creek, Discovery and Validation of Clinical Biomarkers of Cancer: A Review Combining Metabolomics and Proteomics, *Proteomics* 19 (2019) e1700448.
- [52] Breimann L., Statistical modelin: The two cultures (with comments and a rejoinder by the author), *Statistical Science* 16 199–231.
- [53] A.-L. Boulesteix, M. Schmid, Machine learning versus statistical modeling, *Biom. J.* 56 (4) (2014) 588–593.
- [54] K.M. Kingsmore, C.E. Puglisi, A.C. Grammer, P.E. Lipsky, An introduction to machine learning and analysis of its use in rheumatic diseases, *Nat. Rev. Rheumatol.* 17 (2021) 710–730, <https://doi.org/10.1038/s41584-021-00708-w>.
- [55] F. Maleki, N. Muthukrishnan, K. Ovens, C. Reinhold, R. Forghani, Machine Learning Algorithm Validation: From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment, *Neuroimaging Clin. N. Am.* 30 (2020) 433–445, <https://doi.org/10.1016/j.nic.2020.08.004>.
- [56] M.A. Wörheide, J. Krumsiek, G. Kastenmüller, M. Arnold, Multi-omics integration in biomedical research - A metabolomics-centric review, *Anal. Chim. Acta* 1141 (2021) 144–162, <https://doi.org/10.1016/j.aca.2020.10.038>.
- [57] N.P. Long, S.J. Yoon, N.H. Anh, T.D. Nghi, D.K. Lim, Y.J. Hong, S.-S. Hong, S. W. Kwon, A systematic review on metabolomics-based diagnostic biomarker discovery and validation in pancreatic cancer, *Metabolomics* 14 (2018) 109, <https://doi.org/10.1007/s11306-018-1404-2>.
- [58] Vladimir Tolstikov, A. James Moser, Rangaprasad Sarangarajan, Niven R. Narain and Michael A. Kiebish, Current Status of Metabolomic Biomarker Discovery: Impact of Study Design and Demographic Characteristics.
- [59] B.J. Smith, L.C. Silva-Costa, D. Martins-de-Souza, Human disease biomarker panels through systems biology, *Biophys. Rev.* 13 (2021) 1179–1190, <https://doi.org/10.1007/s12551-021-00849-y>.
- [60] E.S. Nakayasu, M. Gritsenko, P.D. Piehowski, Y. Gao, D.J. Orton, A.A. Schepmoes, T.L. Fillmore, B.I. Frohnert, M. Rewers, J.P. Krischer, C. Ansong, A.M. Suchy-Dicey, C. Evans-Molina, W.-J. Qian, B.-J.-M. Webb-Robertson, T.O. Metz, Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation, *Nat. Protoc.* 16 (2021) 3737–3760, <https://doi.org/10.1038/s41596-021-00566-6>.
- [61] C.R. Marchand, F. Farshidfar, J. Rattner, O.F. Bathe, A Framework for Development of Useful Metabolomic Biomarkers and Their Effective Knowledge Translation, *Metabolites* 8 (2018), <https://doi.org/10.3390/metabo8040059>.
- [62] J.P.A. Ioannidis, P.M.M. Bossuyt, Waste, Leaks, and Failures in the Biomarker Pipeline, *Clin. Chem.* 63 (2017) 963–972, <https://doi.org/10.1373/clinchem.2016.254649>.
- [63] H.C. Köfeler, R. Ahrends, E.S. Baker, K. Ekroos, X. Han, N. Hoffmann, M. Holcapek, M.R. Wenk, G. Liebisch, Recommendations for good practice in MS-based lipidomics, *J. Lipid Res.* 62 (2021), 100138, <https://doi.org/10.1016/j.jlr.2021.100138>.