English - Or. English

12 May 2021

Metabolomics Reporting Framework (MRF)

Draft May 2021

This version of the MRF is currently under development by the Expert Group on Metabolomics and is intended to be presented in draft form for the OECD Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST) and Working Party and Hazard Assessment (WPHA) joint 'omics session in June 2021. The MRF is currently tested through case studies and will be distributed to the EAGMST for more formal review in the near future once members of the EG have revised and agreed to the text after the case studies. Given this is a draft that is subject to revision, members of EAGMST and WPHA are invited to acknowledge progress and ask questions for clarification or provide general comments, rather than specific technical comments on text in the draft. Please note that this is a draft document, and members are requested to not circulate the document. This document is accompanied with a Reporting Template in xl format.

JT03476433

Table of Contents

1. Introduction

1.1 Background, Objective and Scope

1.1.1 Background

While a mature technology for research purposes, the application of metabolomics to regulatory toxicology has been limited to date. The ability of metabolomics to measure downstream molecular phenotypes that are predictive of toxicity paves the way for applications in hazard (or adverse outcome) identification, chemical grouping to inform biologically-based read-across of toxicity, identifying metabolic points-of-departure, and potentially other areas (refs).

Regulatory acceptance of (metabol)omics data remains the greatest challenge to its wider use in regulatory decision-making, in part due to the lack of minimal reporting standards for a metabolomics experiment. In 2017, the MEtabolomics standaRds Initiative in Toxicology (MERIT) project was launched, bringing together a team of international experts from industry, government agencies, regulators and academia to start to address this need. Best practice guidelines and minimal reporting standards for the acquisition, processing and statistical analysis of untargeted metabolomics and targeted metabolite data in the context of regulatory toxicology were subsequently developed and published (ref c).

In 2018, the Organisation for Economic Cooperation and Development (OECD) Extended Advisory Group for Molecular Screening and Toxicogenomics (EAGMST) undertook an extension of this work with the objective to develop a Metabolomics Reporting Framework (MRF) Guidance Document (GD). The purposes of this GD are to foster and encourage further international regulatory uptake of metabolomics data, specifically in the context of the OECD's program on chemical safety that focuses on industrial chemicals and biocides. This project was conducted in parallel with the development of a Transcriptomics Reporting Framework (TRF), with each project supported by a team of experts from industry, government agencies, regulators and academia.

1.1.2 Objective

This MRF guidance document is intended to describe the essential information that should be reported when a metabolomics technology is applied in the context of regulatory decision-making, so as to enable assessment of the quality of a metabolomics study from its design through the collection, analysis, and reporting of data. Adherence to such a reporting framework is also anticipated to maximise the likelihood that the results can be reproduced and potentially reused in the form of a knowledge base.

The essential information to be reported includes the experimental design, quality assurance and quality control (QA/QC), sampling of biological specimens, sample processing and extraction of metabolites, data acquisition and processing of untargeted, targeted and hybrid assays assays, annotation and/or identification of the metabolites (i.e., analytes), and statistical analyses specific to

the regulatory application. The specific elements used in a report submitted to a regulator will depend on the context of use of the metabolomics assay.

1.1.3 Scope

The scope of the MRF is in large part defined by its intended applications in regulatory toxicology. Hence the MRF focuses on the application of the most mature, stable and proven metabolomics technologies only. Due to the familiarity within the regulatory community of defined endpoints, both targeted metabolite analysis (i.e. of a small number of pre-defined metabolic biomarkers) and untargeted metabolomics (i.e. broad metabolite profiling of a given sample) are included, along with hybrid approaches of targeted and untargeted.

The MRF is designed to address the needs of two main types of end-users: firstly, regulators - to enable compliance checking of regulatory submissions by conveying essential aspects of the study design, findings and conclusion of regulatory relevance, and secondly to researchers - to promote transparency, data sharing and the meta-analysis of larger datasets (when such opportunities arise, respecting the confidentiality of industry owned data).

Aligned with the TRF, the MRF is constructed around a modular structure that facilitates the updating of individual technologies and allows the development of additional modules for new technologies. The modular structure is described in section 1.2. This narrative guidance document should be used in parallel with the minimal reporting guidelines presented in tabular spreadsheet format, one per module, to facilitate ease of reporting. Although only select reporting elements are indicated as ‘mandatory’, it should be noted that a more comprehensive submission will enhance confidence and potential use in regulatory decision-making.

1.2 Modular Structure of MRF

Each metabolomics study will be reported using the following five types of module: Summary Report (SR), Toxicology Experiment Module (TEM), Data Acquisition & Processing Reporting Module(s) (DAPRM), Data Analysis Reporting Module(s) (DARM), and Application Reporting Module(s) (ARM), which are arranged as summarised in Figure 1. To report a metabolomics study, scientists select the relevant reporting modules (and hence reporting elements), minimally comprising of an SR, TEM, one DAPRM module, one DARM module and one ARM module.

- **Summary Report (SR)**

This module describes a subset of mandatory reporting elements in order to provide a high level overview of the regulatory toxicology and metabolomics experiment. There is one SR per study.

- **Toxicology Experiment Module (TEM)**

This module serves to capture and report the key descriptors of the *in vivo* or *in vitro* toxicology study from which samples are derived for metabolomics analysis. There is one TEM per study.

- **Data Acquisition & Processing Reporting Modules (DAPRMs)**

These modules serve to capture and report descriptions of the metabolomics assays, data acquisition and associated data processing, prior to statistical analysis, e.g. using untargeted LC-MS metabolomics (see Figure 1). There is one or more DAPRM per study (a minimum of one is mandatory), dependent upon the number of metabolomics assays applied to the samples from the study.

- **Data Analysis Reporting Modules (DARMs)**

These modules serve to capture and report descriptions of the statistical analysis that is often undertaken in a metabolomics study, e.g. for the purposes of discovering differentially abundant metabolites. There is one or more DARM per study (a minimum of one is mandatory), dependent upon the type(s) of data analysis applied to the metabolomics datasets.

- **ApplicationReporting Modules (ARMs)**

These modules serve to capture and report descriptions of the application of the metabolomics data to a regulatory application, e.g. for the purposes of chemical grouping to enable read-across (see Figure 1).

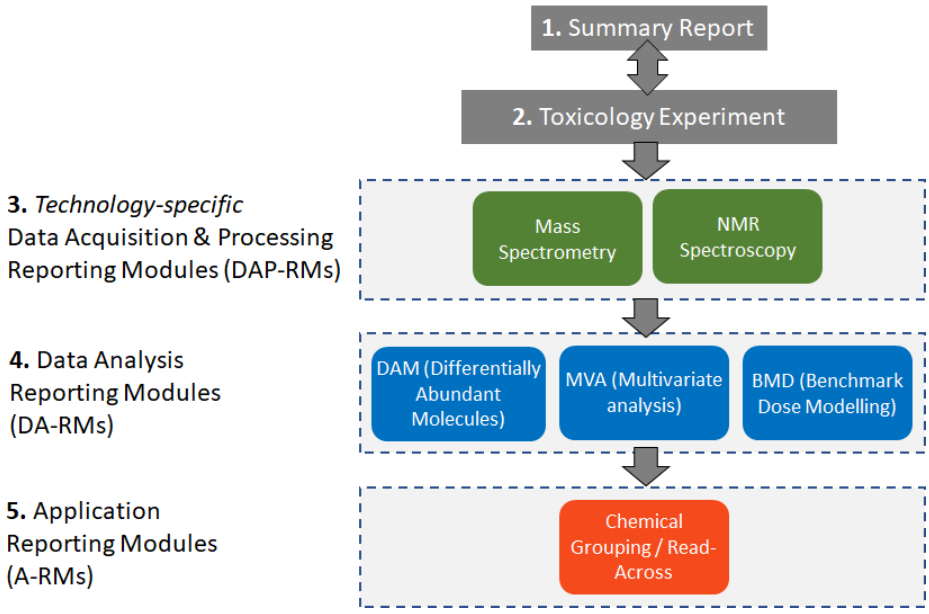

Figure 1. Modular Structure of the Metabolomics Reporting Framework (MRF). (requires updating)

1.3 Example Use Cases using Modular Reporting

To demonstrate the use of the modular reporting format, two examples *(requires updating)* are illustrated here: a hybrid LC-MS metabolomics study applied to chemical grouping (Figure 2) and targeted LC-MS/MS applied to quantify specific metabolic biomarkers (Figure 3).

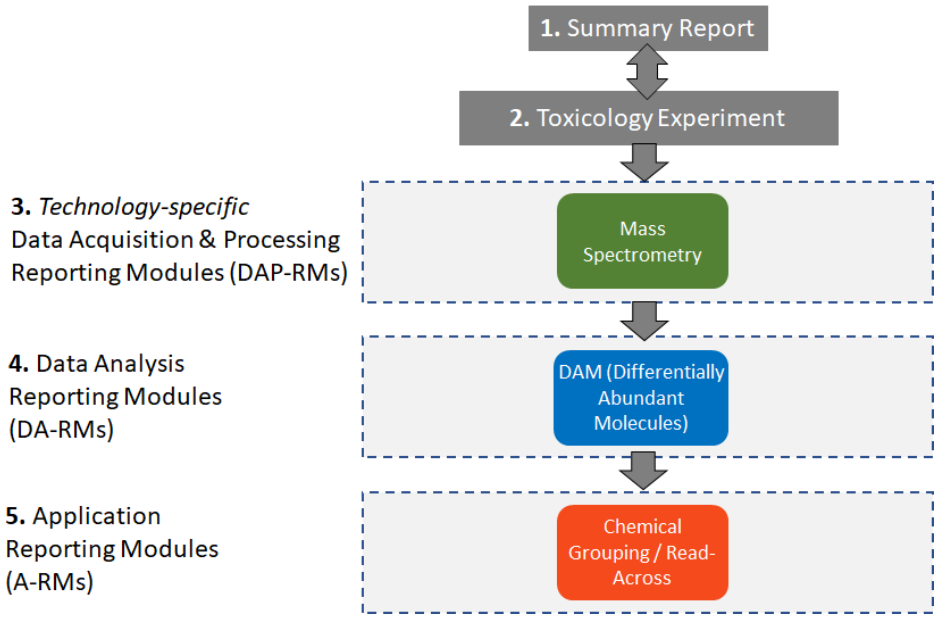

Figure 2. Modular reporting structure when describing the application of hybrid LC-MS metabolomics to chemical grouping. (requires updating)

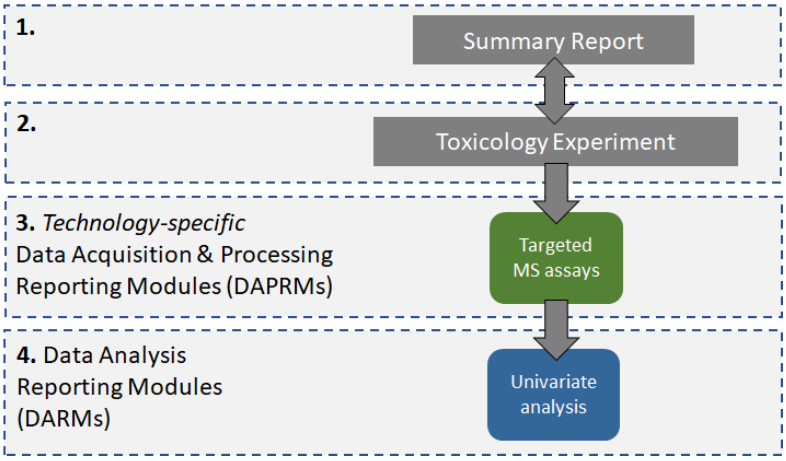

Figure 3. Modular reporting structure when describing the application of targeted LC-MS/MS to measure specific metabolic biomarkers. (requires updating)

Although the MRF guidance document is focused on *what* information should be reported, it is prudent to consider *how* such data and metadata will be reported. Different regulatory authorities will have differing requirements for metabolomics data management. For example, some authorities widely use OECD Harmonised Reporting (OHT) templates for describing data and metadata from regulatory toxicology tests, while others do not. This guidance document describes the reporting of data and metadata in a sufficiently flexible manner that can meet both eventualities. For example, data could be submitted as a single package of modules, as presented in Figures 2 and 3. Alternatively, for the situation where the reporting needs to fit with existing OHTs and data structures, the reporting

could be separated into two parts, one termed the *Regulatory compliance path* and the other the *Complete data path*, as proposed by the MERIT guidelines [ref].

- The *regulatory compliance path* would allow compliance standards to be checked - i.e., only critical information describing the data source, data quality, results and interpretation from the metabolomics experiment are reported. This could include the Summary Report and selected reporting elements from a DAPRM(s), DARM(s) and ARM(s). The majority of information described in the Toxicology Experiment Module may already be reported via an existing OHT.
- The *complete data path* would report the full metabolomics data and metadata to an access-controlled specialised metabolomics data repository, comprising of a DAPRM(s), DARM(s) and ARM(s). This could include some or all information from the Summary Report and Toxicology Experiment Module.

The benefits of this sub-structured reporting include ensuring that regulators are not swamped with unnecessary technical details, yet those details are available for deeper examination if required; minimising the number of amendments that will be required to existing reporting mechanisms to allow the addition of metabolomics findings; and maximising the re-use of existing specialist metabolomics data repositories.

2. Summary Report

This module describes a subset of mandatory reporting elements in order to provide a high level summary of the application of metabolomics to a regulatory toxicology study. With the exception of the reporting element in section "2.1 Study identifiers", each section is derived from reporting elements reported elsewhere within the MRF GD. One summary report should be reported per study.

2.1 Study identifiers

REPORT:

1. Unique study identifier
2. Link to standardised toxicology dataset (e.g. link to OECD test dataset). Mandatory, if available.
3. Omics complete dataset (e.g. link to MetaboLights, ArrayExpress accession number). Mandatory, if available.

2.2. Toxicology Experiment Module summary

2.2.1 Study rationale

Derived from section 3.1.

REPORT:

1. Objective(s) - to ensure clarity in the purpose for which the omics approach is proposed, which may include a specific regulatory application
2. Background information - for example, the objectives may be anchored to an existing AOP or to other mechanistic information on the molecular endpoints being measured

2.2.2 Test and control Items

Derived from section 3.2

REPORT:

1. Test item name
2. Test item-specific identifiers
    a. CAS
    b. SMILES
    c. IUPAC name
    d. Additional information where available, e.g. InChi, Distributed Structure-Searchable Toxicity (DSSTox) substance identifier (DTXSID), etc.
    e. Sources of identifier2.2.3 Test system characteristics

Derived from section 3.3

REPORT:

1. General characteristics of the test system or subject:
   EITHER in vivo
   a. Animal species
   b. Strain
   c. Sex
   d. Age during study
   OR in vitro
   e. Cell type
   f. Origin (animal species/organ/tissue)

2.2.4 Study design

Derived from section 3.4

REPORT:

1. Dose selection
   a. Dose levels
   b. Dose intervals
2. Number of biological replicates per treatment condition

2.2.5 Treatment Conditions

Derived from section 3.5

REPORT:

1. Route of administration
2. Vehicle description
3. Exposure duration / schedule

2.2.6 Study exit & sample collection

Derived from section 3.6

REPORT:

1. Type of biological sample (*in vitro*: cells, cell media, etc.; *in vivo*: biofluid, cells, tissue, organ, organism, etc.)
2. Quench solvent

2.3 Data Acquisition & Processing Reporting Module(s) summary

2.3.1 Metabolite extraction from biofluids, cells and tissues

Derived from sections 4.1.1.1. / 4.2.1.1.

REPORT:

1. Extraction method general description

2.3.2 Mass spectrometry / NMR spectroscopy assay type(s)

Derived from sections 4.1.3.1. / 4.2.3.1.

REPORT:

1. Mass spectrometry / NMR spectroscopy assay type
2. Mass spectrometry / NMR spectroscopy assay type description

2.3.3 Demonstration of quality of metabolomics analysis

Derived from sections 4.1.5. / 4.2.5.

REPORT:

1. Was a System Suitability QC measured and Instrument performance report submitted?
2. Was an Intrastudy QC measured and an Intrastudy QC reproducibility report submitted?
3. Was an Intralaboratory QC measured and an Intralaboratory QC reproducibility report submitted?
4. Was an Interlaboratory QC measured and an Interlaboratory QC reproducibility report submitted?
5. Was a process blank measured and applied to the data processing?

2.3.4 Analytical confidence in the identity of each feature and/or metabolite

Derived from sections 4.1.6.3. / 4.2.6.3.

REPORT:

1. Summary of MSI level(s) of identification (1-4), across all features and/or metabolites

2.3.5 Analytical confidence in the quantity of each feature and/or metabolite

Derived from sections 4.1.6.5. / 4.2.6.5

REPORT:

1. Summary of level(s) of quantification (1-3), across all features and/or metabolites

2.4. Data Analysis Reporting Module(s) summary

2.4.1 Data analysis approach

Derived from section 5.

REPORT

1. DARM(s) used

2.4.2 Rationale for data analysis approach

Derived from section 5.

REPORT

1. What is purpose of applying the DARM(s)

3. Toxicology Experiment Module

The Toxicology Experiment Module serves to capture and report the key descriptors of the *in vivo* or *in vitro* toxicology study from which samples are derived for metabolomics analysis. One TEM should be reported per study.

3.1 Study Rationale

A clear and concise report of the study rationale is necessary to understand the suitability of the experimental design for the regulatory question being addressed, including the selection of experimental model, sex, target tissue, dosing regimen, etc. These fundamental aspects of the experimental design are clearly dependent on the study rationale.

Over the past two decades metabolomics has increasingly been applied to chemical toxicity and disease studies, revealing new insights into mode(s) and mechanism(s) of action, disease markers, and toxicity signatures in human and environmental health (ref MERIT paper). Despite these successes, metabolomics has experienced only limited application to regulatory toxicology (van Ravenzwaay et al., 2016; Buesen et al., 2014; ECETOC, 2013, 2010, 2010a). Recently, efforts have intensified to address this underutilization. As mentioned above, the Metabolomics standaRds Initiative in Toxicology (MERIT) project proposed a number of best practices for maximizing the utility of metabolomics datasets to regulators and thus facilitating the use of metabolomics in regulatory toxicology. In addition, a set of specific applications were recommended to address the most immediate and practical contributions for metabolomics to regulatory toxicology. Specifically, the MERIT team recommended the following four applications: 1) deriving points of departure via benchmark dosing, 2) discovery of chemical mode(s) of action and molecular key events, 3) chemical grouping and read-across, and 4) cross-species extrapolation of toxicity pathways. Not surprisingly, the value of the results collected from studies conducted for these and other test applications will depend largely on the appropriate use of experimental designs and analytical approaches.

Experimental designs for each of the applications described above will differ. For example, if the identification of differentially expressed metabolites (i.e. statistical testing) is required for a mode of action analysis, appropriate sample sizes per experimental group are required. In contrast, establishing similarities in metabolomic profiles to support chemical groupings for read-across through unsupervised clustering approaches may be done with a smaller number of biological replicates, but may require either the availability or production of a database of metabolomic profiles against which comparisons can be made. The use of benchmark-dose analysis to identify a metabolomic point of departure benefits from a larger number of dose groups than typically used in toxicological studies but this can be offset by smaller sample sizes per group. In addition, the purpose of the study may govern the analytical platform used in deriving the metabolomic profiles (e.g. use of mass spectrometry for the detection of metabolites at low concentrations) which will impact the choices made for downstream analyses.

Overall, a clear rationale describing why a metabolomic study was undertaken is required to assess the suitability of the experimental design and its intended use for regulatory decision making. This section is intended to be a narrative description of the reasoning for the study and its design. Details of the study parameters are to be provided in later sections.

REPORT:

1. Background information

   Provide necessary background information for the end user to understand the rationale for why the study was undertaken, including the regulatory question(s).

2. Objectives

   Clearly define the objectives of the metabolomic study toward informing the regulatory question. Describe whether the metabolomic study results are intended to be interpreted in isolation or in combination with results from other studies.

3. Test Guideline compliance

   If appropriate, please refer to which OECD Test Guidelines have been followed in the performance of the method.

4. Mechanistic understanding

   Briefly describe any prior toxicological, mode of action, or mechanistic information that is useful to understanding the study rationale (e.g. established mechanism of action and its relationship to the toxicological effect of interest).

5. Model selection

   For animal studies, briefly explain how and why the selected animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology. Provide a rationale for tissue or organ selection for the study.

   Provide a rationale for the species and strain used.

   For *in vitro* studies, briefly describe the biological relevance of the test system used in relation to the tissue/organ/species of interest.

6. Dose level and dose interval selection

   Provide a brief rationale for the selection of the employed doses levels and dose-intervals. For example, selection for *in vivo* experimentation may be based on known toxicological effects or molecular changes documented for the test article identified in prior studies, allowing for "read-across" between metabolomic data generated and other in-life findings, or clinical pathological changes and pathological observations. Note that dose interval information should include the time of exposure during the day. Similarly, in the case of *in vitro* experiments, if a relationship is sought to exposures *in vivo*, test material concentrations may be chosen using a quantitative *in vitro*/*in vivo* extrapolation (IVIVE) rationale. Similarly, a rationale for *in vitro* and *in vivo* dose intervals should be provided.

7. Route of administration

Where relevant, provide a rationale for the choice of route of administration, referring to objectives of the study, potential route of human exposure, the physical and chemical characteristics of the test item and the relevance for the evaluated endpoint.

8. Time point selection

   Provide a brief rationale for the exposure durations and sampling time points. Metabolome profiles are dependent on both the duration of exposure and the time of sample collection. Metabolite changes may adapt over time, with early time points reflecting molecular initiating and early key events, and later time points reflecting pathological changes or adaptation. For *in vivo* studies, the interval between the final dosing and sample procurement should be specified. The same is true for sampling-time post-treatment in *in vitro* studies.

9. Samples and replicates

   Provide a clear rationale for the choice of:

   a. Biological replicate number, based on the scientific question posed and statistical power calculations predicting adequate coverage of biological variability.

   b. Number of technical and analytical replicates, based on accepted and/or published standards for the assay and compliance with statistical power calculations.

10. Limitations

    To facilitate regulatory evaluation, when appropriate, indicate the study limitations that could affect the outcome or the interpretation of the results. These can include technical or mechanistic limitations in relation to known modes of action. For example, if the culture conditions include poorly described systems that would be a source of uncontrollable variability (such as large inter-individual variability in urine concentrations); another limitation of cell systems may be the information about the metabolic capacity of the cells or the known metabolic pathways which can be reflected in the cell type; such information should be available to the evaluator. Likewise, some of the test articles used might have physicochemical properties (lipophilicity, volatility, etc.) that might lead to a cell exposure that is different from the expected exposure (through interaction with plastic or proteins in culture plates and medium) or that produce large and confounding signals in NMR or MS spectra. In the case of *in vivo* studies, discussion of limitations should include any potential source of bias of the animal model or imprecision associated with the result.

3.2 Test and Control Items

According to the *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring*, a test item is defined as the subject of a study, and is also associated with "test compound", "test substance", "test article", or other similar terms to describe the item being tested (OECD 2018.a). Studies submitted for analysis to regulatory agencies should in the spirit of good laboratory practice (GLP) report test item transportation, receipt, identification, labelling (see section 2.8 Sample Identification Codes), sampling, handling, storage and characterization. Information regarding the test item characterization is needed to inform potential route of exposure, as well as physicochemical properties that might influence the study (i.e., solubility, volatility, etc.).

Regulatory scientists must have all test items, vehicle, and control identification and characterization information in order to accurately interpret metabolomic study results. The following information should be reported for all test items: A) test substance, B) vehicle, and C) controls, including: test item name, mixture formulation composition, preparation of test item, physicochemical properties, chemical stability (OECD 2018.b), commercial source, and substance-specific identifiers. Additional information for nanomaterial test items should also be provided according to the 2016 OECD Workshop Report on *Physical-Chemical Parameters: Measurements and Methods Relevant for the Regulation of Nanomaterials* (OECD 2016.a).

REPORT:

1. Test item name

2. If test item is a mixture, report formulation composition
    a. Identify substances that make up the mixture
    b. Percentage of substances (if known)

3. Preparation of test item (composition)
    a. Concentration of test items
    b. Concentration of diluent(s)
    c. Identification of impurities

4. Physicochemical properties
    a. Appearance/physical state/colour
    b. Molecular weight
    c. Melting point/freezing point
    d. Boiling point
    e. pH
    f. Viscosity
    g. Density
    h. Vapor pressure
    i. Partition coefficient
    j. Water solubility
    k. Solubility in organic solvents/fat solubility
    l. Particle size distribution/fiber length and diameter distribution
    m. Additional physicochemical information (i.e., agglomeration, porosity, etc.)

5. Chemical stability
    a. Stability in organic solvents and identity of relevant degradation products
    b. Storage stability and reactivity towards container material
    c. Stability: thermal, sunlight, metals, if relevant
    d. Stability: dissociation constant, if relevant

6. Commercial source
    a. Vendor
    b. Manufacture ID
    c. Lot (Batch) number
    d. Purity
    e. Salt form
    f. Expiration date
    g. Storage conditions

7. Test item-specific identifiers
    a. CAS
    b. SMILES
    c. IUPAC name
    d. Additional information where available, e.g. InChIKey, InChI string, Distributed Structure-Searchable Toxicity (DSSTox) substance identifier (DTXSID), etc.
    e. Sources of identifiers

3.3 Test System Characteristics

Metabolomic studies suitable for use in a regulatory context may be performed in a wide range of model species and/or *in vitro* test systems. In fact, a notable strength of the metabolomics approach is its amenability to nearly any species of interest without consideration for the extent of sequencing for that species genome. Despite this, the bulk of vertebrate metabolomic studies in toxicity testing have focused on rodent models. However, the field of metabolomics is now replete with studies that make use of alternative species' (i.e. fish, invertebrates, etc.) that can be employed to inform cross-species extrapolation of toxicity pathways.

Similar to traditional toxicity testing, it is critical that regulatory scientists applying metabolomic data for risk assessment be provided with comprehensive information regarding the characteristics of the test system from which the data are derived. Test system refers to the biological system that is exposed to the test items to obtain experimental data. There are numerous examples in the literature demonstrating differential susceptibility of different species, strains within a species and sexes to chemical toxicity. Likewise, *in vitro* test systems derived from different species, tissues or even individuals vary in terms of relative sensitivity to toxicant exposure. The end user must be equipped with detailed and accurate information regarding the

test species or *in vitro* test system used to generate the metabolomic data in order to critically evaluate the results and accurately compare the results across studies and data types.

With respect to *in vivo* toxicology studies, researchers should include relevant taxonomic information (i.e. species and strain), sex, age (at onset of dosing and at study termination) and commercial source of all individuals included in a study. If determination of sex was not included in the study design (such as in the case of some types of alternative species studies), or pooled samples from multiple individuals were examined, then this should be explicitly described by the researcher. Researchers should also include detailed information on the housing conditions for all individuals included in a study including number of individuals housed per cage, type of bedding, type of food, type of water provided, food and water accessibility (i.e. ad libitum or defined quantities), light / dark cycle, relative humidity and other housing conditions the researcher may deem relevant for study interpretation. In general, information following the ARRIVE guidelines should be included (Kilkenny, Browne et al. 2010).

With respect to *in vitro* toxicology studies, researchers should include relevant information on culture type including species, strain (if applicable), sex of the organism, and organ or tissue from which the cells were derived. Researchers should include detailed information on culture conditions used to conduct the study as applicable, including complete media formulations, culturing vessel, growth substrate, passage number, donor lot, source (including commercial vendor or academic source), incubator conditions and proof of cell line authentication if available (OECD 2018C). In studies using complex, multicellular culture models (e.g. 3D cell models, organoids, organ-on-chip, etc.), the researchers should report what types of cells the cultures are expected to contain, cite relevant literature characterizing the model system and describe any other relevant characteristic that might not be listed here.

REPORT:

1. General characteristics of the test system or subject:

In vivo

   a. Animal species
   b. Strain
   c. Sex
   d. Age during study
   e. Developmental stage
   f. Individual weights/lengths at start
   g. Supplier
   h. Any interventions that were carried out before or during the experiment
   i. Quality criteria before use
   j. Health status and acclimation prior to study start
   k. Randomization of animals to groups
   l. Identification
   m. Other …

In vitro

   a. Cell type (cell line or primary cells, tumor cells, etc…)

b. Origin (animal/organ/tissue)
c. Cell passage number
d. Differentiation stage
e. Absence of mycoplasma
f. Metabolic competence
g. Cell banking
h. Supplier
i. Quality criteria before use
j. Other

2. Housing, husbandry and culture conditions

In vivo

a. Type of facility (e.g. specific pathogen free [SPF])
b. Type of cage or housing
c. Bedding material
d. Number of cage companions
e. Tank shape (for fish) and its material
f. Breeding programme
g. Light/dark cycle
h. Temperature
i. Quality of water (e.g. for aquatic toxicity tests)
j. Type of food (ingredients in food as detailed as possible)
k. Access to food and water
l. Environmental enrichment
m. Methods for fertilization/collection of eggs, if applicable
n. Other appropriate characteristics

In vitro

a. Incubation characteristics:
    - Temperature
    - CO_2/O_2 conditions
    - Humidity
    - Other
b. State of the cells before use:
    - Viability (including test used)
    - Quality control

- Morphology
- Recommended confluency of use
- Other

c. Culture media (in case of multiple)

d. Use of serum (with details such as species of origin, age, sex, etc)

e. Use of antibiotics

f. Use of feeder cells

g. Use of matrixes or scaffolds

3.4 Study Design

Study designs are based on the assessment of all available information for factors that have the potential to influence study results. A detailed description of all elements and parameters included in the study design will increase transparency and confidence in metabolomic data, and as a result, will have greater utility in regulatory assessment. The following guidance for reporting study design using *in vitro* and *in vivo* systems to generate metabolomic data is based on previously published OECD guidance documents for the respective areas (OECD 2014; OECD 2017). Much of the reporting guidance is based on the application of good laboratory practice (GLP) principles and good *in vitro* method practice (GIVIMP) according to an internationally accepted definition, ensuring mutual acceptance of data (MAD) across OECD countries (OECD 1998; OECD 2018.c). Although GLP-compliant study protocols have been developed for most areas of regulatory risk assessment, there are no absolute requirements for their application in the generation of metabolomic data submitted for regulatory purposes (EPA 2009; FDA 2015). In general, study designs should provide experimental detail, standard operating procedure (SOP) information, and statistical design information in equivalence to the sentiment of GLP study design, but not necessarily requiring all aspects of traceability etc., which are generally required for GLP auditing. However, a study running under GLP is preferred, as is the standard for regulatory studies following OECD Test Guidelines. In addition to these recommendations, those published recently by the MERIT project (Viant 2019), and which describe in detail a metabolomics study design, should be followed to the fullest extent possible. Finally, the report should detail any 3R (reduction, refinement and replacement) arguments underlying the study design, e.g. choice between *in vivo* and *in vitro* tests systems, statistical powering (see also Section 2.1 on study rationale and Section 2.2 on test system characteristics) (EC 2010).

REPORT:

1. 3Rs considerations

Briefly describe how the study addresses the 3R principles.

2. Dose selection

a. Dose levels

Indicate each of the dose levels/concentrations used in the study and identify the matched vehicle/solvent controls to be used.

b. Dose intervals

Indicate:

- Dose interval (acute single or chronic dosing).
- Time of day of dosing (in vitro: time of dosing after seeding).
- Frequency of dosing.

3. Description of the test method instruments, equipment, and reagents

Provide a full description of the instruments and equipment used for the collection and processing of samples for metabolomic analysis, with details concerning:

a. Commercial source, detailing the suppliers/manufacturers of instrumentation, other laboratory equipment and reagents relevant to the study.

b. Manufacturer's instrument model identification.

c. Manufacturer's reagent and kit information.

d. Any special safety/handling requirements.

4. Types of treatments

The study design report must include a description of the type(s) of treatment including:

a. Controls: Defined as experimental samples derived from animals or cells treated with their respective dose formulation, in the absence of test article. All control types should be reported (i.e., positive control, negative control, vehicle control, blank, etc.) and following the criteria described in Section 2.2. Test and control items.

b. Pre-treatments: Where necessary, a description of pre-treatments involving metabolic activation, for example of specific cytochromes(s) P450, should be provided.

c. Acclimation: A brief description of animal acclimation should be reported to include the length of the acclimation period, health status of the test system, and environmental conditions. Relevant quarantine conditions should be described, where applicable.

d. Types of replicates: The report should clearly define the number of biological replicates (samples derived from individual animals or cell samples), utilized for each dose level. This should be clearly delineable from any technical replicates generated, (the sample processed more than once) and/or analytical replicates (the same sample analysed more than once) (Blainey, Krzywinski et al. 2014).

5. Numbers of individuals/samples per treatment

a. Clearly describe the number of biological replicates in each treatment condition.

b. Describe the number of technical and analytical replicates.

6. Statistical design

a. Exposure design

Describe the various statistical approaches (Chow 2014) used in the study design to prevent exposure bias e.g.:

- Randomized block
- Latin square
- Incomplete block.

b. Sampling schemes

Describe sampling schemes used to prevent sample collection bias, and to ensure proper sample labelling post-collection, using methods such as:

- Sequential
- Stratified
- Systematic
- Randomized
- Ranked set

c. Sample blinding

Describe sample blinding approaches following the sample collection, used to prevent experimental bias in downstream sample processing. This should include a unique identification (as explained in section 2.8 Sample Identification Codes) that does not represent the sample or treatment type.

7. Observations/examinations during treatment

Where appropriate, include details of other experimental observations used in the experiments generating the metabolomic samples, including:

a) In-life cage-side or clinical observations, feed consumption, water consumption and body weight in for *in vivo* experiments

b) Toxicokinetics

c) Histopathology and organ weight

d) Clinical pathology in *in vivo* experiments

e) Reason animals were removed from the study

f) Cytological analyses in *in vivo* experiments

g) Cytobiological examinations in *in vitro* experiments (such as cell morphology and cytotoxicity testing)

h) Transcriptomics, proteomics, etc. analysis in *in vivo* and *in vitro* experiments

8. Timetable

Detail the timetables used to perform the study protocol with respect to:

a) Treatments

b) Sample collections

c) Time since last dose administered

d) Time to sample extractions

3.5 Treatment Conditions

In the context of the Adverse Outcome Pathway (AOP) framework (http://www.oecd.org/chemicalsafety/testing/projects-adverse-outcome-pathways.htm), reasoned, well-defined exposures through relevant routes of administration result in changes of endogenous metabolite levels and pathways representative of key events possibly related to the final adverse outcome. The objective of well-designed metabolomic studies informs the selection of the treatment conditions, which, in turn, impact the final outcome. Metabolomic analyses may produce different results depending on the route of administration, the dose levels and the time and schedule of the exposure. Thus, a thorough description of treatment conditions is necessary for interpreting metabolomic results.

For both *in vivo* and *in vitro* metabolomic studies (as in any study), it is understood that the amount of chemical that reaches the target will affect the outcome. If available provide information on tissue dosimetry either measured or modelled.

The present guidance for reporting treatment conditions in metabolomic studies is based mainly on previously published OECD documents (OECD 2014; OECD 2017; OECD 2018.c; OECD 2018.a) and standard harmonized templates for reporting of information derived from *in vivo* or *in vitro* studies for the risk assessment of chemicals (OECD 2016; OECD 2018). Also the ARRIVE guidelines for reporting animal research were taken into account (Kilkenny, Browne et al. 2010). Following GLP-like requirements for accurate, comprehensive reporting will help in the evaluation of the relevance of data deriving from metabolome studies.

REPORT:

1. Route of administration

   Indicate the selected route of administration.

   In vivo

   a. Oral
      - Gavage
      - Diet
   b. Dermal
   c. Inhalation (specify the type of exposure)
   d. Implantation
   e. Parenteral
   f. Exposure regime, if applicable (static, semi-static, flow through)

g. Other

In vitro

a. Direct addition of the test item or a preparation of the test item to cultures

b. Substitution of culture medium

c. Exposure at the air-liquid interface

d. Other

Describe any modifications of the standard culture/housing conditions occurring before and during the test item exposure (refer to 'Housing, husbandry and culture conditions' in section 'Test System Characteristics'), in addition to:

In vivo

a. Fasting period

b. Anesthesia and/or analgesia

c. Other

In vitro

a. Serum free or serum depleted medium

b. Use of items for limiting media evaporation

c. Other

2. Test item preparation

Provide the dose/concentration levels and the dose/concentration level spacing. Describe all the steps leading to the test item preparation for administration to the test system and any modifications to the original procedure, e.g. problems with chemical solubility. For dose levels/concentration to be used, see paragraph 2.20:

a) Dilution in a vehicle

b) Preparation steps (warming, grinding, ……)

c) Separation steps (centrifugation, decantation, filtration,…..)

d) Extraction steps (for specific test items, such as medical devices)

e) Storage conditions

f) Stability during storage

g) Expiration date

h) Whether nominal or measured concentrations were used, if applicable

i) Analytic controls on measured concentrations, if applicable

j) Dosing solution homogeneity

Describe the procedures used to assess:

a. Stability of the test substance under test conditions

b. Solubility and stability in the solvent/vehicle

c. Reactivity of the test substance with the solvent/vehicle or the cell culture medium, if applicable

d. Impact of separation and extraction steps on integrity, homogeneity, concentration and stability of the prepared test item

For *in vivo* studies, provide details about:

a. Procedures for test substance formulation/diet preparation

b. Procedures for generation of test atmosphere and chamber description

c. Achieved concentration

d. Stability and homogeneity of the preparation

e. Test material intake for dietary or drinking water studies Conversion from diet/drinking water or inhalation test substance concentration (ppm) to the actual dose (mg/kg body weight/day), if applicable.

3. Vehicle description and delivery volume

If the test chemical is dissolved or suspended in a suitable vehicle, provide all the relevant information on the vehicle, the delivery volume and the final concentration of the vehicle in the test item preparation. Indicate the maximum volume of liquid that has been administered by gavage or injection. The use of volume exceeding the suggested volume should be justified. In terms of reporting, refer to section 'Test and Control Items').

4. Exposure duration / schedule

Indicate exposure duration (hours, days, weeks or months) and the frequency of the administration.

Indicate the recovery (or depuration) period (in days, weeks, months, if any) after the exposure to the test substance.

5. Housing / culture conditions during treatment

If different from the normal maintenance (refer to section 'Test System Characteristics'), please report on the environmental conditions recorded during treatment. Describe also the undesired deviations from the housing/culture conditions established in the study plan that occurred during the treatment and/or the observation time and their possible impact on the study results.

a. Temperature

b. Humidity

c. CO_2 %

d. pH

e. Availability and quality of nutrients

f. Other

3.6 Study Exit & Sample Collection

Metabolomics studies can be conducted using samples from animal studies and from *in vitro* studies. When properly performed, metabolome profiling of such a sample produces a "snapshot" of its endogenous metabolite levels and allows assessment of their alterations due to perturbation of biological processes. It is widely understood that proper sample preparation is a key step in metabolomics studies and adequate care must be taken to ensure sample fidelity. Several factors with regard to preparation for sampling need to be taken into account because of their potential for causing alterations in the metabolome, which may confound biological interpretation of the data. This applies to both animal studies and *in vitro* studies, although the type of biological sample drives the selection of the subsequent handling steps.

Animal studies should always be conducted in strict accordance with ethical principles and regulations. When terminating an animal study, euthanasia must be performed using appropriate techniques and equipment to ensure death is induced in a manner that is as painless and stress-free as possible. Consequently, anaesthetics are commonly used in these procedures. Sometimes, analgesic drugs are administered during and/or after surgical procedures in laboratory animals. Several studies have demonstrated that anaesthesia and euthanasia may impact omics studies (HK 2004; Overmyer, Thonusin et al. 2015; Nakatsu, Igarashi et al. 2017). Also, for *in vitro* studies, the methods used for harvesting the samples may influence the metabolome (Ramirez et al. 2017). A detailed report on the methodologies used for collecting and storing specimens will allow reviewers to appropriately evaluate the quality of the metabolomic studies.

Additionally, biotic and abiotic information at the time of harvesting should be collected to allow for assessment of sample fidelity. This also includes the conditions under which samples are stored until further processing. As metabolome data are sensitive to environmental conditions, information should be reported as accurately as possible. A more detailed overview of information to be included in the report is listed below.

REPORT:

1. Type of biological sample (*in vitro*: single cell, cell culture, 2D or 3D culture, single or multi type cell culture; *in vivo*: biofluid, cells, tissue, organ, organism)

In vivo

a) Anesthetic used: substance (e.g. isoflurane, ether), dosage, route of administration

b) Analgesic used: substance, dosage, route of administration

c) Method of euthanasia (e.g. carbon dioxide asphyxiation, exsanguination)

d) Phenotypic characteristics (e.g. body weight, organ weight)

e) Methods used for collection of biological sample (e.g. dissection, isolation of tissues or organs, etc.)

In vitro

a) Cell density at time of harvesting

b) Growth phase/stage (e.g. cell cycle phase, if available)

c) Number of culture passage

d) Morphology

e) Methods used for collection of biological sample (e.g. removal of media, wash (see below), quench (see below), scrape into sampling vial, etc.)

2. Sampling vial

a) Type of vial or tube

b) Chemicals within the sample tube (EDTA, heparin, etc.)

c) Chemicals added to preserve sample (nitrogen, argon, etc.)

3. Washing

The primary purpose of washing a sample is to remove the contamination of one type of metabolome by a second type. For example, prior to the extraction of adherent cells to study the intracellular metabolome, it is important to remove (via washing) metabolites present in the cell media. Washing is particularly important for an untargeted LC-MS assay as it is a sensitive analytical method.

a) Wash solvent(s)
b) Washing procedure (including temperature)

4. Quenching

The primary purpose of quenching is to preserve the levels of metabolites in the isolated sample as similarly to their levels in the living system.

a) Quench solvent

b) Quenching procedure (including temperature)

5. Pooling (or aliquoting) of samples

6. Sample storage and transport

Transport and storage conditions prior to metabolite extraction are important factors in the reliability of metabolite measurements. Storage temperature, time and the number of freeze-thaw cycles can all affect the stability of endogenous metabolites.

a) Post sample collection handling, prior to metabolite extraction

b) Storage temperature and duration

c) Transportation method (e.g. between experimental facilities)

d) Number of freeze-thaw cycles

3.7 Sample Identification Codes

Sample management is a critical component of regulatory and non-regulatory experimentation which should be carefully planned. To aid in laboratory organization and management, a laboratory information management system (LIMS) can be used to consolidate laboratory tasks, such as: sample management, laboratory work-flows and protocols, documentation, management of laboratory stocks and solutions, and

clinical data (List, Schmidt et al. 2014). Samples used for metabolomic experimentation should be given a unique identification code and information stored in a secure LIMS where available.

A standard operating procedure (SOP) should be established to ensure identification, tracking, unbiased testing, and data collection records for each sample. The sample identification code generation should be produced in the spirit of good laboratory practice (GLP) in order to maintain proper records of samples and their associated method of experimentation (OECD 1998). The code identification of each unique sample should be securely linked to test item information, experimental study number, and experimental metadata.

REPORT:

1. Laboratory information management system (LIMS)
    a. Software used for information management

2. Unique code for each sample
    a. Method of code generation

3. Information stored for each sample code:
    a. Sample name
    b. Sample source
    c. Sample storage
    d. Characterization
    e. Stability
    f. Vehicle and/or buffer
    g. Test item information
    h. Receiving date
    i. Test date
    j. Metadata
    k. Experimental results

3.8 Supporting Data Streams

Metabolomics studies can be used to address different types of regulatory questions. In order to be able to fully appreciate a metabolomics study and its resulting data, a clear and concise report is required. The framework described in this guidance ensures that all essential information is captured to allow for this detailed understanding.

However, there may be situations where even a higher level of detail is needed to allow for use of metabolomic data for regulatory decision-making. Moreover, data may be re-used for other regulatory questions or for a similar regulatory question at a later point in time. To benefit optimally from the data generated, additional information should be reported to the extent possible. This information can range

from OECD Test Guidelines for a particular animal study to methodological Standard Operating Procedures (SOP) to scientific publications in which analysis of (a subset of) the data has been described. Toxicity or cytotoxicity experiments necessary to establish the appropriate doses/concentrations can also be reported here.

4. MRF Technology-specific Data Acquisition & Processing Reporting Modules

Applications of metabolomics in research utilise a broad array of analytical technologies, typically involving a separation technology (e.g. chromatography) and a detection device (e.g. mass spectrometry). As this OECD reporting framework focuses on the application of metabolomics in regulatory toxicology, only the most mature, stable and proven technologies are considered. Based on two international surveys (refs), the most widely applied analytical methods for metabolite analysis comprise of liquid chromatography mass spectrometry (LC-MS; [e.g. Rick Nature Protocols]), gas chromatography mass spectrometry (GC-MS; [Beale et al. Review of recent developments in GC–MS approaches to metabolomics-based research (2018) Metabolomics, 14:152]), direct infusion mass spectrometry (DIMS; [e.g. Ralf Nature Protocols]) and nuclear magnetic resonance spectroscopy (NMR; [ref An early Nicholson paper + W.B. Dunn, N.J.C. Bailey, H.E. Johnson. 2005. Measuring the metabolome: current analytical technologies. The Analyst, 130: 606-625 + Smolinska et al., 2012. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Analytica Chimica Acta 750: 82-97]). These technologies can be applied in different configurations, depending on the question being addressed by the study, and are often broadly categorised as either untargeted, targeted or a hybrid of one or more assay types (sometimes termed semi-targeted).

Untargeted assay - no analytes are pre-selected for measurement; any metabolites that are detected can be (i) unknown (MSI level 4), annotated (MSI levels 2-3) or identified (MSI level 1) [ref], and (ii) are all relatively quantified (i.e. relative between different samples such as treated vs. controls). In practice the majority of features measured in an untargeted assay will typically be either unknown (MSI level 4) or annotated (MSI levels 2-3).

Targeted assay - all analytes are pre-selected for measurement. Traditionally this would mean all metabolites that are detected are (1) identified to MSI level 1, and are (2) quantified using a metabolite-specific standard. Traditional assays require metabolite-specific standards acquired in the same laboratory using the same analytical methods where the absolute concentration of metabolites in the biological sample are measured (taking into account any matrix effects). See FDA guidelines [couple of refs to check: "Bioanalytical Method Validation - Guidance for Industry" and "Analytical Procedures and Methods Validation for Drugs and Biologics - Guidance for Industry" - both in this folder https://drive.google.com/open?id=1ZSTRHvVDq9hBJykvtxUUkiko_dE2DeI1. also check FDA guidance "In vitro diagnostic multivariate index assays" should also be mentioned.] for traditional targeted assays to measure metabolic biomarkers. However, depending on the availability and use of reference standards in the assay, targeted assays can also "semi-quantify" metabolites rather than provide absolute quantification. Moreover, targeted assays can also be used to reliably detect (e.g. retention time, *m/z*, MRM) and relatively quantify 'analytically known' metabolites, the annotation of which is unclear at the time of measurement (MSI level 4).

Hybrid assays - combine components of untargeted and targeted analyte selection into a single assay. They typically have a higher information content than untargeted assays by attempting to measure pre-selected (and often toxicologically important) metabolic biomarkers..

Section 4 is divided into mass spectrometry metabolomics - detailing all the required reporting fields for widely used untargeted, targeted and hybrid assays for LC-MS, GC-MS and DIMS (section 4.1) and NMR metabolomics (section 4.2, covering only untargeted assays).

The reporting described in this module is generalisable and adaptable, because (1) it is not our intention to prescribe the types of assays that the regulatory toxicology community should use; (2) technologies will

continue to evolve and we want the MRF to remain relevant; (3) metabolomics assays often do not fit well into predefined boxes; and (4) the assays themselves do not need to form the basis of regulatory standards, it is the utility and confidence in the quality of the data produced from the assays that should form the basis of acceptability criteria.

4.1 Mass Spectrometry Metabolomics Module

Mass spectrometry is the most widely used analytical tool for untargeted and targeted measurements of metabolic phenotypes [ref?]. In this **Data Acquisition and Processing Reporting Module** we describe the reporting required for a **mass spectrometry-based metabolomics study**, including (1) sample processing, metabolite extraction and preparation of standards, (2) analytical QA/QC, (3) acquisition and (4) processing of mass spectrometry data, (5) demonstrating the quality of the data, and (6) the data matrices produced by this technology, including metabolite annotation/identification and intensities.

4.1.1. Sample processing, metabolite extraction and addition of chemical reference standards

While the specific protocols for metabolite extraction will depend on the sample type (i.e., biofluid, cells, tissue, etc.) being measured, a reporting framework that can capture the most important information about a diverse range of methods is presented here.

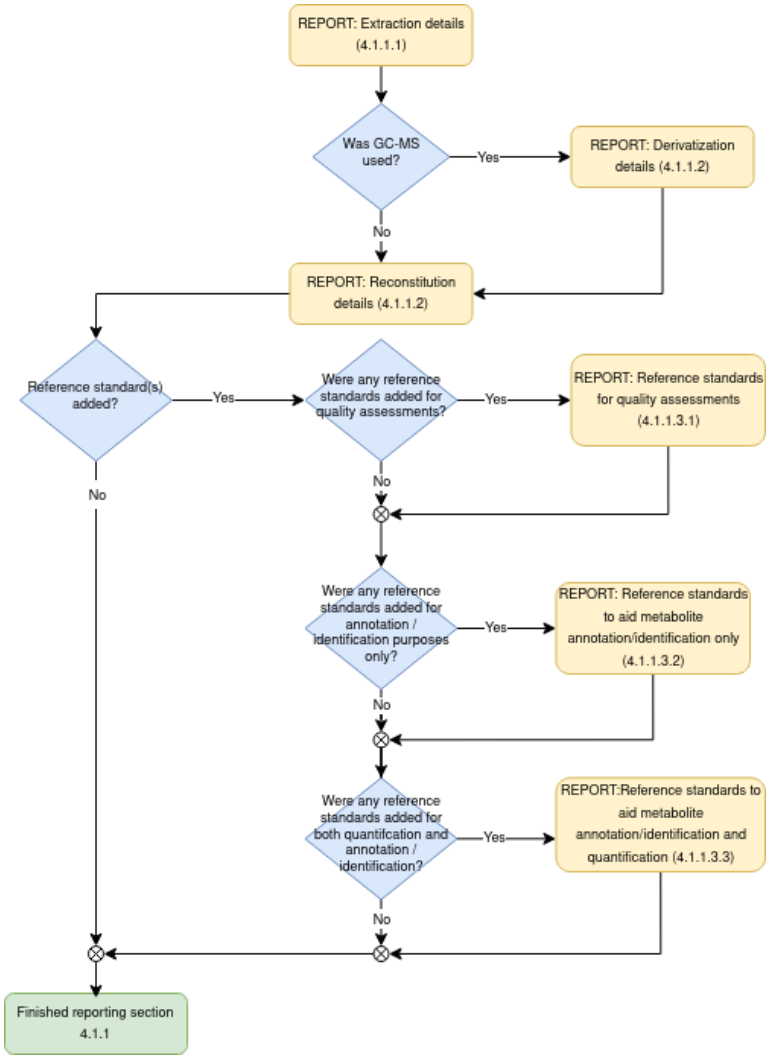

Figure 1: Workflow for mass spectrometry metabolomics reporting - Sample processing, metabolite extraction and addition of chemical reference standards

4.1.1.1. Metabolite extraction from biofluids, cells and tissues

Extraction methods differ for liquid, cellular and tissue sample types, though typically require addition of an organic solvent to denature any proteins present and therefore stop any enzymatic activity that would otherwise change the metabolome.

REPORT:

1. Extraction method general description
2. Sample volume(s)
3. Solvent(s) used
4. Means of agitation/maceration
5. Temperatures and times
6. Post extraction handling, *e.g.*, storage temperature and duration of extracts.

4.1.1.2. Extract concentration and reconstitution in solvent for mass spectrometry analysis

For many sample types, before analysis, the metabolite extract is evaporated to dryness and reconstituted in solvents that (1) facilitate their ionisation, and (2) are compatible with the LC-MS or GC-MS mobile and stationary phases.

REPORT (Only if GC-MS used):

1. Derivatization method general description
2. Reagents and reaction (including incubation temperature and time)
3. Clean-up/partitioning (if used).

REPORT:

4. Evaporation and reconstitution method general description
5. Final reconstitution solvent(s) and final volume (if used)
6. Storage temperature (if relevant)
7. Duration of reconstituted extracts (if relevant)

4.1.1.3. Addition of reference standard(s)

Reference standards can be prepared for a variety of purposes including standards for quality assessment and to correct for variability, standards used for aiding identification/annotation, and standards used for both identification and quantification. Key considerations include if the reference standard was used to annotate, identify and/or quantify an actual metabolite or used to annotate and/or quantify a (set of) chemically related metabolite(s) (e.g. a class of metabolites); if the standard used is a surrogate of the metabolite of interest; also if the reference standard was spiked directly into the biological sample or not.

This document uses the following definition for *surrogate* from Gómez-Sáez et al "a pure substance (compound or element) similar to [the] analyte of interest in chemical composition, separation and measuring which is taken to be representative of the native analyte; this must be absent or in a negligible initial concentration in the sample."

REF: Cuadros-Rodríguez L, Bagur-González MG, Sánchez-Vinas M, González-Casado A, Gómez-Sáez AM. Principles of analytical calibration/quantification for the separation sciences. Journal of Chromatography A. 2007 Jul 27;1158(1-2):33-46. http://depa.fquim.unam.mx/amyd/archivero/calibracionanalitica_6118.pdf

4.1.1.3.1. Internal reference standards for assessing the quality of sample extraction and/or mass spectrometry detection (quality assessment reference standards)

Internal standards could correct the experimental variability (e.g. extraction and/or the instrument performance) for each sample. Specifically, if the internal standard is added at the start of the extraction procedure (pre-spiked) it can provide information on the extraction (optionally matched against a post-spiked representative sample extract) - and serve as an 'extraction standard'. If the internal standard is added after the last step of the extraction procedure (typically added to the reconstitution solvent) it can correct instrument performance - and serve as an 'injection standard'. Both an extraction standard and injection standard can be used in the same assay.

REPORT:

1. Quality assessment reference standard general preparation description
2. For each standard used - report the following:
    a. Standard use (extraction or injection standard)
    b. Name
    c. PubChem cid (optional)
    d. HMDB id (optional)
    e. KEGG id (optional)
    f. CAS (optional)
    g. InChi (optional)
    h. InChiKey (optional)
    i. SMILES (optional)
    j. Purity (if known)
    k. Supplier
    l. Concentration
    m. Point of addition in sample preparation procedure

4.1.1.3.2. Reference standards to aid metabolite annotation/identification **only**

This refers to reference standards prepared to either accurately identify one or more individual metabolites or annotate metabolite classes in a biological sample. These standards are only used for annotation and identification purposes and not used for quantification. All the standards should ideally be acquired on the same instrument type and method as applied to measure the biological samples. The standards can be spiked internally into a biological sample (typically through the spiking of an isotopically-labelled reference material) or spiked into an external alternative matrix (e.g., surrogate or working solvent only)

REPORT:

1. Reference standards for metabolite annotation/identification general preparation description
2. For each standard or each named panel of standards used - report the following:
    a. Spiked internally into the biological sample or alternative external matrix?
    b. Metabolite or metabolite class standard?
    c. Was a surrogate used? If yes - then provide detail
    d. Name
    e. PubChem cid (optional)
    f. HMDB id (optional)

g. KEGG id (optional)
h. CAS (optional)
i. InChi (optional)
j. InChiKey (optional)
k. SMILES (optional)
l. Purity(ies) (if known)
m. Supplier
n. Concentration(s)
o. Point of addition in sample preparation procedure

Existing in-house as well as external (both public and commercial) mass spectral libraries of standards can also be used for annotation purposes. In these cases the full details regarding the reference standard preparation might not be known or available but the source of the mass spectral library and version should be reported in Section 4.1.4.10.

4.1.1.3.3. Reference standards to aid metabolite annotation/identification **and** quantification

This refers to reference standards prepared to either identify and quantify one or more metabolites, or annotate and (semi-) quantify one or more metabolite classes. All the standards need to be acquired on the same instrument type and method as the biological samples. To quantify, single or multi-point calibration data is required.

Traditionally, these reference standards have mostly been made 'in-house' by individual laboratories and are largely defined by the specific metabolites or metabolic pathways of interest. More recently, commercial kits containing a panel of reference standards have become available, often in convenient multi-well plate formats. Identical reference standards should be available for all targeted mass spectrometry assays within a study (both intra- and interlaboratory), with mixtures ideally being prepared by/sourced from one laboratory. Furthermore, the ability to report full reference standard content and associated methods for metabolite identification and quantification should be ensured before commencing a project.

Often considered the most reliable approach in metabolomics for calculating absolute abundances using LC-MS or GC-MS is the inclusion of isotopically-labelled standards and the use of SRM (single reaction monitoring) or MRMs (multiple reaction monitoring, i.e. the selected monitoring of multiple product ions from one or more precursor ions) for each metabolite that is measured. *Here, known amounts of an array of isotope-labelled relevant metabolites are spiked into each sample. Absolute quantification can then be achieved by comparison of the peak areas of the labelled and unlabelled versions of each metabolite.* This method also gives an additional level of confidence to metabolite identification. Guidelines for traditional targeted assays for quantification using metabolite standards are available from the FDA (couple of refs to check: "Bioanalytical Method Validation - Guidance for Industry" and "Analytical Procedures and Methods Validation for Drugs and Biologics - Guidance for Industry" - both in this folder https://drive.google.com/open?id=1ZSTRHvVDq9hBJykvtxUUkiko_dE2DeI1).

Throughout this guidance document the term "absolute quantification" refers only to the process when a reference standard has been used to measure an actual metabolite (or a metabolite surrogate that is very similar to the actual metabolite, e.g. a stable isotope labelled metabolite) and any matrix effects have been

accounted for. All other quantification methods using reference standards, described in this guidance document, are considered "semi-quantification" and if **no** reference standards are used only "relative-quantification" can be achieved. See Table 1 for a description of the quantification methods using reference standards.

For LC-MS and DIMS, to account for the 'matrix effect' when performing quantification, the reference standard typically needs to be internally spiked into the biological sample matrix. For GC-MS, matrix effects are generally less prominent, so the reference standard does not always need to be "internally spiked" to account for such effects but this will depend on the compound being used for the reference standard. Full details on the assessment of the matrix effect should be detailed in section 4.1.2.6**.**

Table 1: Summary of quantification methods based on the type of reference standards.

		How similar is the reference standard to the metabolite of interest?	
		Actual metabolite or a surrogate that is **very** similar to the actual metabolite (e.g. stable isotope labelled metabolite)	A surrogate that is similar/related to the actual metabolite (e.g. the surrogate is the same metabolite class)
Have matrix effects been accounted for?	Yes	**Absolute quantification**	**Semi quantification**
	No	**Semi quantification**	**Semi quantification**

REPORT:

1. Reference standards to aid metabolite annotation/identification and quantification preparation details
2. For each standard each standard or each named panel of standards used - report the following:
    a. Use of standard (absolute quantification or semi quantification (see Table 1)
    b. Has matrix effect been accounted for? Provide detail
    c. Was a surrogate used? Provide detail including how similar the reference standard is to the metabolite of interest
    d. Spiked internally into the biological sample or alternative external solution?
    e. Metabolite or metabolite class standard?
    f. Calibration methodology (choose one) [Ref Cuadros-Rodríguez L 2007]
        i. External calibration
        ii. Matrix matched calibration
        iii. Calibration by internal normalisation
        iv. Standard addition calibration
        v. Internal calibration
    g. Name
    h. PubChem CID (optional)
    i. HMDB id (optional)
    j. KEGG id (optional)
    k. CAS (optional)
    l. InChI (optional)

m. InChIKey (optional)
n. SMILES (optional)
o. Purity (if known)
p. Supplier
q. Concentration(s)
r. Point of addition in sample preparation procedure

4.1.2. Analytical quality assurance and preparation of quality control samples

When conducting a metabolomics study using LC-MS, GC-MS or DIMS it is essential to have a quality assurance (QA) framework and use quality control (QC) samples, as described in the MERIT best practice guidelines [ref]. Here, the reporting of the analytical QA/QC is described for LC-MS, GC-MS and DIMS instrument set up and calibration, and for analysis of a set of biological and QC samples. The QC results are reported in the section 4.1.5 - *Demonstration of quality of metabolomics analysis*, below.

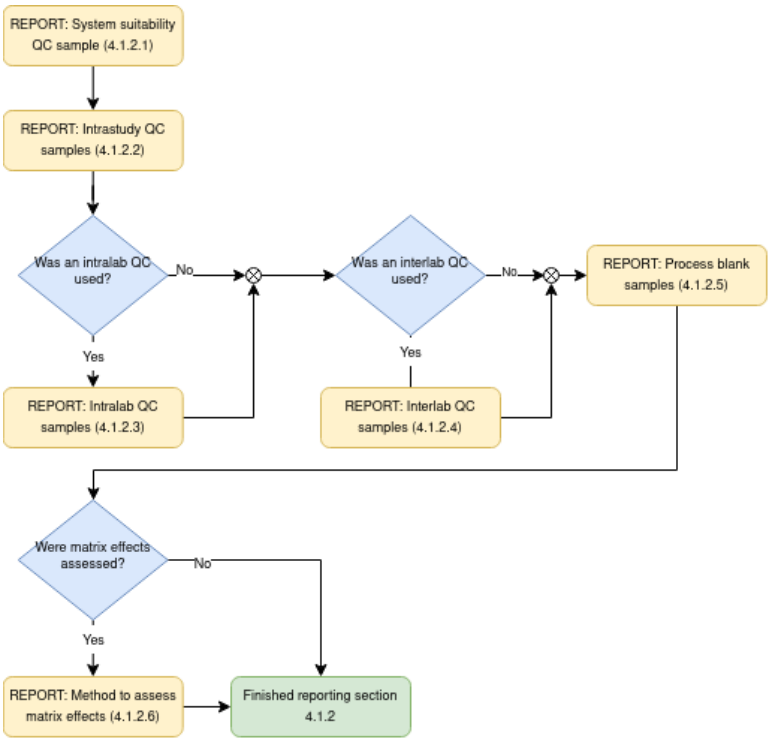

Figure 2: Workflow for mass spectrometry metabolomics reporting - Analytical quality assurance and preparation of quality control samples

4.1.2.1. System suitability QC sample

Used to ensure that LC or GC retention times (if chromatography is used), *m/z* measurements and feature intensities are within specification of the instrument. Should be consistent over long periods and potentially

usable across multiple laboratories. A synthetic sample comprising a mixture of metabolite standards or reference material can be used for this purpose.

REPORT:

1. Source
2. Preparation details
3. Storage conditions
4. Days since preparation

4.1.2.2. Intrastudy QC sample

Used to pre-condition the chromatography (if used) and mass spectrometry, and analysed repeatedly throughout an analytical batch to assess, and potentially correct, any drifts in measurement performance. It is essential that the intrastudy QC is highly representative of the biological samples in the study. Typically this type of QC is derived from a small aliquot of the biological samples within the study.

REPORT:

1. Source
2. Preparation details
3. Storage conditions
4. Days since preparation

4.1.2.3. Intralaboratory QC sample (optional)

Used to assess (and potentially correct for) any differences between separate studies within one laboratory. Should be representative of the biological sample type in the study and hence derived from a one-time pool of multiple extracted samples by a specific laboratory using a defined protocol, or a synthetic sample covering the relevant metabolite space, or a reference material of sufficiently similar metabolic composition to the biological sample type.

REPORT:

1. Source
2. Preparation details
3. Batch number
4. Certificate of analysis (COA) - if relevant
5. Storage conditions
6. Days since preparation

4.1.2.4. Interlaboratory QC sample (optional)

Used to assess (and potentially correct for) any differences between individual laboratories. Should be accessible to multiple laboratories, has known providence, is stable, characterised and available in controlled batch numbers. Ideally this type of QC has a similar metabolic composition or matrix to the biological samples in the study, although this is not always possible, in which case use as close to the same composition as possible.

REPORT:

1. Source
2. Preparation details
3. Batch number
4. Certificate of analysis (COA) - if relevant

5. Storage conditions
6. Days since preparation

4.1.2.5. Process blank sample

Used to provide a measure of background contamination arising from the extraction and LC-MS, GC-MS or DIMS analysis. It is study specific, prepared in the same manner as the biological samples except that no biological material is present. It is important to define the start and end points of the 'process' used to prepare this type of QC sample.

REPORT:

1. Type of process blank
2. Start and end points of the 'process' used to prepare
3. Storage conditions

4.1.2.6. Assessment of sample matrix effects (optional)

Matrix effects, mostly related to ESI sources, can be estimated by either post-column infusion for qualitative estimation [ref SR1] or by the method described by Matuszewski et al. for quantitative estimation [ref SR2], after selective sample preparation. The latter approach includes the estimation of analyte loss during the extraction step and the signal alteration (ion enhancement or suppression) due to the interfering compounds from the matrix. Additional care must be taken with this phenomenon, especially in quantitative analysis or relative quantification, as mentioned by the Food and Drug Administration (FDA) which recommends identifying any matrix effects [ref SR3].

REPORT:

1. Method to assess matrix effects.

4.1.3. Acquisition of mass spectrometry metabolomics and metabolite data

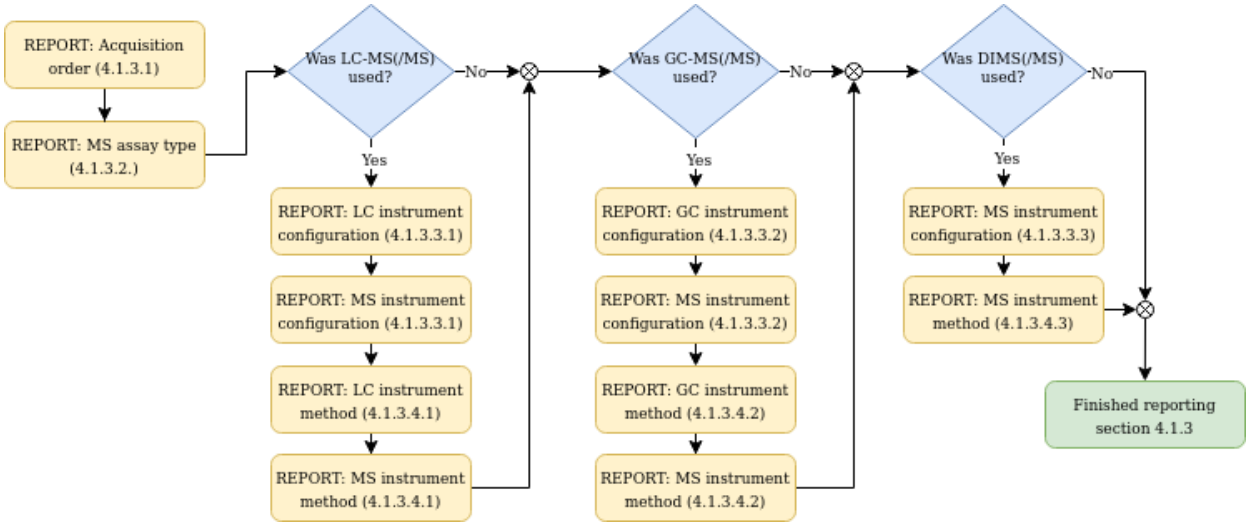

Figure 3: Workflow for mass spectrometry metabolomics reporting - Acquisition of mass spectrometry metabolomics and metabolite data

4.1.3.1. Acquisition order for QC samples, biological samples and reference standard samples

REPORT:

1. Acquisition order of all types of QC samples, biological samples and reference standard samples (if relevant), thereby indicating the number of pre-conditioning QC samples, the number of process blank samples, and the frequency of analysis of intrastudy QC samples. Columns of reported table should include:
    a. Run order
    b. File name
    c. Sample type (must be able to distinguish between QC samples, biological samples and process blanks)

4.1.3.2. Mass spectrometry assay type(s)

Mass spectrometry metabolomics assays can be untargeted, targeted, or a hybrid. Each assay can have one or more levels of annotation/identification and quantification. Given there are multiple subtypes of these assays, Table 2 describes the list of possible metabolomics/metabolite assays. A single toxicology study can comprise a combination of several assays.

REPORT:

1. Mass spectrometry assay type(s) used (see Table 2)
2. Mass spectrometry assay type description

Table 2: Summary of all relevant mass spectrometry assay types for metabolomics/metabolite analysis.

	Name	**Notes**
Untargeted assays (no features are pre-selected for data acquisition)		
1	Untargeted with relative quantification	No standards measured (for identification or quantification) at time of assay.
Targeted assays (all features are pre-selected for data acquisition)		
2	Targeted with relative quantification only	This includes targeting of 'known unknown' metabolites for which the metabolite identity is unknown.
3	Targeted with semi-quantification only	This includes measuring intensities when using a metabolite class standard, and/or measuring intensities when matrix effects have not been accounted for.
4	Targeted with absolute quantification only	Traditional targeted assay.

5	Targeted with combination of quantification methods	Any combination of the targeted approaches 2-4.
Hybrid assays (some features are pre-selected and some features are not pre-selected for data acquisition)		
6	Hybrid with relative quantification	Combination of untargeted approach (1) and targeted approaches (2).
7	Hybrid with combination of quantification methods	Combination of untargeted approach (1) and targeted approaches (2-5). Need ≥ 1 reference standard used for either semi or absolute quantification

4.1.3.3. Instrument configuration(s)

4.1.3.3.1. LC-MS(/MS)

REPORT: LC configuration

1. LC configuration description
2. LC manufacturer
3. LC model number/name
4. Software package(s) and version number(s)
5. Column and pre/guard column manufacturer
6. Column model number/name
7. Stationary phase composition and particle size
8. Column internal diameter and length
9. Injection vials or plate manufacturer and model number

REPORT: MS configuration

10. MS configuration description
11. MS manufacturer
12. MS model number/name
13. Software package(s) and version number(s)
14. Ionisation source (ESI, APCI, APPI, etc.)
15. Type of mass analyser (triple quadrupole MS, Orbitrap, time-of-flight, FT-ICR, ion-trap, etc.)

4.1.3.3.2. GC-MS(/MS)

REPORT: GC configuration

1. GC configuration description
2. GC manufacturer
3. GC model number/name
4. Software package and version number
5. Column manufacturer
6. Column model number/name
7. Stationary phase composition

8. Column internal diameter and length
9. Injection vials manufacturer and model number
10. Plates manufacturer and model number

REPORT: MS configuration

11. MS configuration description
12. MS manufacturer
13. MS model number/name
14. Software package(s) and version number(s)
15. Ionisation source (EI, CI, APCI, etc.)
16. Electron energy (for EI)
17. Source voltage
18. Source temperature
19. Gas flows
20. Type of mass analyser (quadrupole MS, Orbitrap, time-of-flight, FT-ICR, ion-trap, etc)

4.1.3.3.3. DIMS(/MS)

REPORT: MS configuration

1. MS configuration description
2. MS manufacturer
3. MS model number/name
4. Software package(s) and version number(s)
5. Ionisation source (ESI, APCI, APPI, etc.)
6. Source voltage (or target current, as relevant to the ion source)
7. Source temperature
8. Gas flows
9. Type of mass analyser (triple quadrupole MS, Orbitrap, time-of-flight, FT-ICR, ion-trap, etc.).

4.1.3.4. Instrument method(s)

4.1.3.4.1 LC-MS(/MS)

REPORT: LC method

1. LC method description
2. LC method name
3. Mobile phase composition
4. Mobile phase flow rate
5. Composition of the wash solvent
6. Column temperature and pressure
7. Gradient profile
8. Amount of sample injected

REPORT: MS method

9. MS method description
10. MS method name
11. MS acquisition mode(s) (full scan, MRM, SRM, DDA, DIA, MSn, etc.);
12. Polarity (positive or negative ion analysis)
13. m/z scan range

14. Mass resolution
15. Collision energies
16. Isolation width
17. Lock spray details
18. Source voltage
19. Source temperature
20. Gas flows

4.1.3.4.2. GC-MS(/MS)

REPORT: GC method

1. GC method description
2. GC method name
3. inlet system (e.g. split/splitless)
4. inlet temperature (including whether constant or ramped)
5. Transfer line temperature
6. Gas flows and pressure
7. Temperature gradient
8. Amount of sample injected

REPORT: MS method

9. MS method description
10. MS method name
11. MS acquisition mode(s) (full scan, MRM, SRM, DDA, DIA, MSn, etc.);
12. Polarity (positive or negative ion analysis)
13. m/z scan range
14. Mass resolution
15. Lock spray details
16. Source voltage
17. Source temperature
18. Gas flows

4.1.3.4.3. DIMS(/MS)

REPORT: MS method

1. MS method description
2. MS method name
3. MS acquisition mode(s) (full scan, MRM, SRM, DDA, DIA, MSn, etc.);
4. Polarity (positive or negative ion analysis)
5. m/z scan range
6. Mass resolution
7. Lock spray details
8. Source voltage
9. Source temperature
10. Gas flows

4.1.4 Processing of mass spectrometry metabolomics and metabolite data

This section covers the reporting of the data processing steps that are required to transform raw data into a form amenable to statistical analysis. This usually requires production of a 2-dimensional data table with each sample represented in one dimension (typically a row) and each metabolic feature (e.g. peak) in the other dimension (typically a column).

The entries in the data table will either be used for relative quantification, absolute quantification or "semi-quantification" of the specific feature in each sample. Each metabolic feature in the data table may also have different levels of metabolite annotation/identification (Ref MSI levels).

This section incorporates the reporting of both 'data processing' and 'data post-processing' as described by the MERIT guidelines [ref].

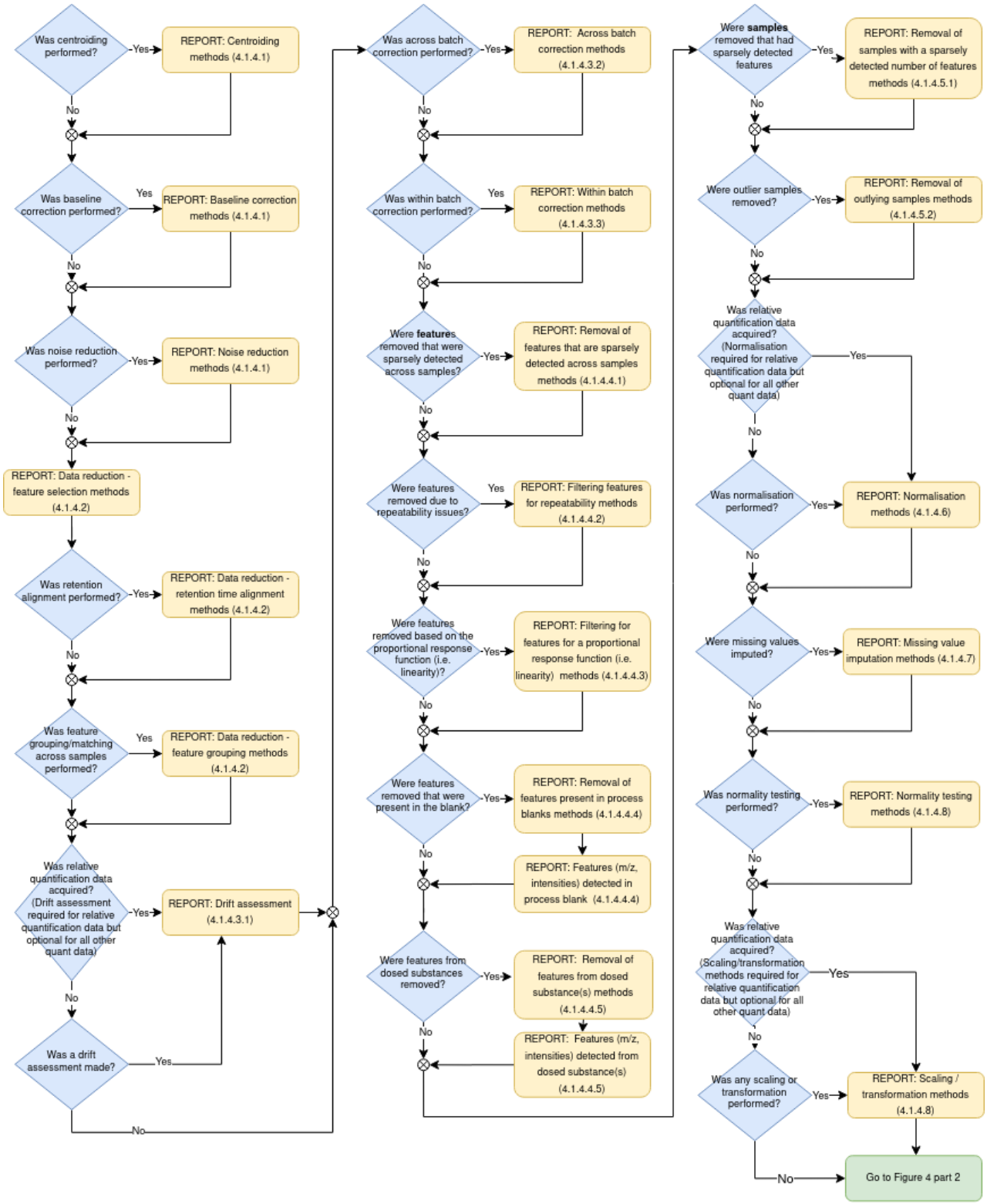

Figure 4: Workflow for mass spectrometry metabolomics reporting - Processing of metabolomics data (1 of 2)

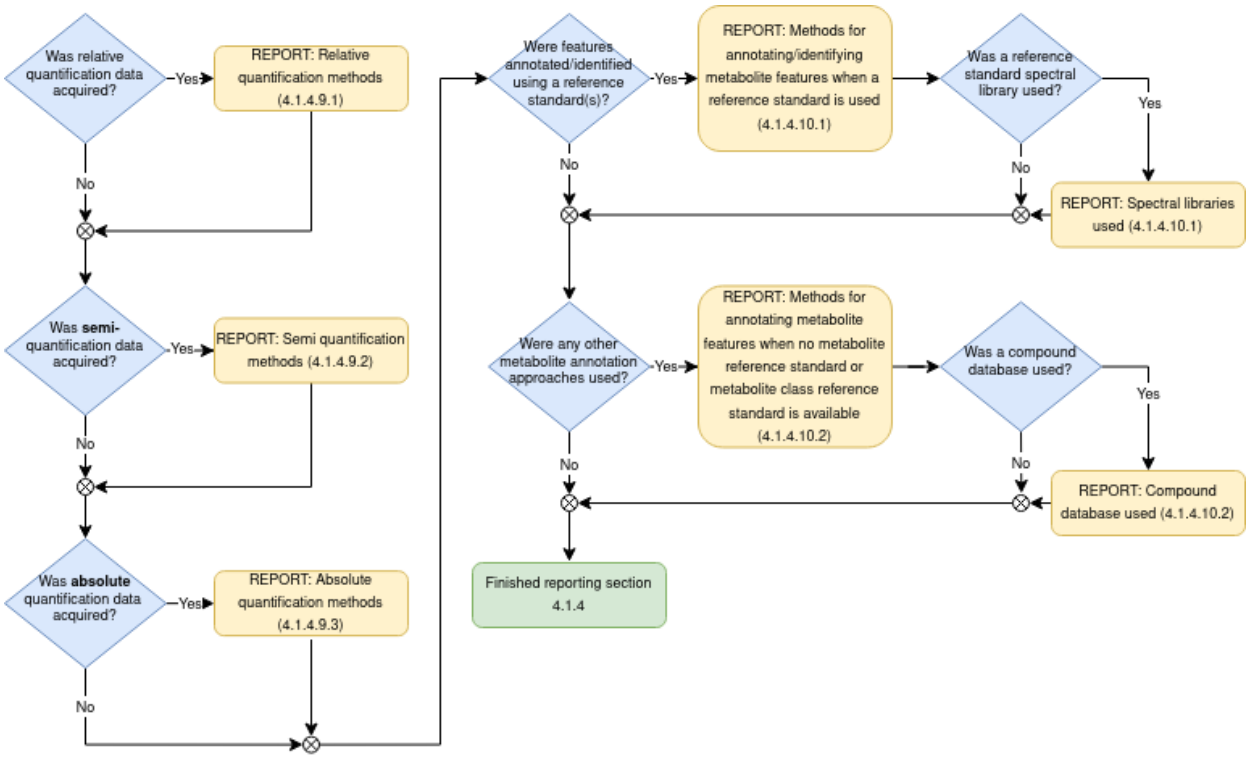

Figure 4: Workflow for mass spectrometry metabolomics reporting - Processing of metabolomics data (2 of 2)

4.1.4.1. Centroiding, baseline correction and noise reduction (optional)

Mass spectrometry data is typically measured in 'profile mode'. To reduce file sizes considerably the raw data files are often reduced to a form where each feature is represented as an individual *m/z* with zero line width, a process called centroiding.

REPORT: Centroiding

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)

Typically, raw data from the instrument software does not require further processing. Some further optional operations can be performed: baseline correction and noise reduction.

REPORT: Baseline correction

6. **S**oftware
7. Software version(s)
8. DOI or URL of the software / script(s) (optional)
9. Algorithm(s)/method(s)
10. Parameter(s)

REPORT: Noise reduction

11. **S**oftware
12. Software version(s)
13. DOI or URL of the software / script(s) (optional)
14. Algorithm(s)/method(s)
15. Parameter(s)

4.1.4.2. Data reduction

Data reduction involves four steps: feature detection/picking; retention time alignment to take into account shifts in retention time of the same analytes in different samples; grouping/matching of features from the same analyte across different samples (if relevant); and feature integration when estimating the abundance of a metabolite.

REPORT: Feature detection/feature picking

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)

REPORT: Retention time alignment

6. **S**oftware
7. Software version(s)
8. DOI or URL of the software / script(s) (optional)
9. Algorithm(s)/method(s)
10. Parameter(s)

REPORT: Grouping/matching

11. **S**oftware
12. Software version(s)
13. DOI or URL of the software / script(s) (optional)
14. Algorithm(s)/method(s)
15. Parameter(s)

4.1.4.3. Feature intensity drift and/or batch correction

While every effort should be made experimentally to minimise variations in mass spectrometry signal intensity within and between analytical batches, these can be (partially) corrected in the data processing step, typically using signals recorded in the intrastudy QC samples.

4.1.4.3.1 Assessment of the presence of within- and/or between-batch signal intensity drift (optional for absolute quantification/semi-quantification**)**

Evaluation of the presence of such signal intensity drift typically includes a PCA analysis, showing both biological and intrastudy QC samples and relative standard deviation measurements (RSD; also known as

coefficient of variation (CV)) for some metabolites present in QC samples (covering a wide range of physicochemical properties and concentrations). If the intrastudy QC samples show significant differences in scores or RSD values within any batch, or a drift in scores values or RSD values between batches, then drift and/or batch effects are present.

REPORT:

1. PCA scores plots showing biological and intrastudy QC samples before any correction applied
2. PCA scores plot of just the intrastudy QC samples (optional)

4.1.4.3.2 Batch correction (optional)

If used, the effects of batch correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the batch correction.

REPORT:

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)
6. Evidence that batch correction has been successful (e.g. PCA scores plots showing intrastudy QCs).

4.1.4.3.3 Signal intensity drift correction within a batch (optional)

If used, the effects of drift correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the correction is made.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)
6. Evidence that the correction has been successful (e.g. PCA scores plots showing intrastudy QCs).

4.1.4.4. Identification and removal ('filtering') of features (optional)

Particularly untargeted mass spectrometry analysis may detect a significant number of low intensity noisy features and/or systematically biased features, which should be considered for removal to improve the overall reliability of the data set. Several optional filtering processes may be conducted and reported as described here.

4.1.4.4.1 Removal of features that are sparsely detected across biological and/or intrastudy QC samples (optional)

Features can be removed if they are only present below a defined percentage of biological and/or intrastudy QC samples, or below a defined percentage of a biological sample group.

REPORT:

1. **S**oftware
2. Software version(s)

3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including threshold

4.1.4.4.2 Filtering features for repeatability (optional)

A widely used procedure in untargeted metabolomics is to remove features with high analytical variability in intensity. The RSD of the intensity of each feature can be estimated from the intrastudy QC samples and features with an RSD greater than a threshold are removed.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including RSD threshold

4.1.4.4.3 Filtering for features for a proportional response function (i.e. linearity) (optional)

Some analytical designs may include a dilution series, where an intrastudy QC is diluted by known factors. Within the instruments linear range, reliable features should exhibit intensities which correlate very strongly with the known dilution factors. A feasible strategy is therefore to remove features whose intensities do not correlate well to the dilution factors.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including correlation threshold

4.1.4.4.4 Removal of features present in process blanks (optional)

Features detected in process blanks are thought to result from the solvents or plasticware rather than the biological sample and therefore can be considered for removal from the final data set.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including threshold settings for their removal
6. Features (*m/z*, intensities and retention time (if available)) detected in process blank. Columns of reported table should include
    a. Feature ID
    b. mz
    c. Intensity
    d. Retention time (if available)

4.1.4.4.5 Removal of features from dosed substance(s) (optional)

In toxicology studies, in which the biological system is deliberately exposed to an exogenous chemical, it is common to observe the dosed parent substance and/or its biotransformation products in the resulting

data. For the purposes of analysing the endogenous metabolic effect of the exposure, it is important that these signals are removed from the data set. Identification of the relevant features to remove will typically involve comparison to control spectra from non-dosed animals, spectra from a chemical standard, and/or literature. Typically features identified for removal will simply be deleted from the data set.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including threshold settings for their removal
6. Features (*m/z*, intensities, retention time (if available)) detected from dosed substance(s). Columns of reported table should include
    a. Feature ID
    b. mz
    c. Intensity
    d. Retention time (if available)

4.1.4.5. Identification and removal ('filtering') of outlying samples

It is important to identify potential outlying samples and to remove then, if necessary, before applying statistical analysis. Reasons for removing each outlier must be clearly explained in terms of biological, analytical or data analytical aspects; e.g. sample removed due to limited volume, contamination was detected, results of an independent assay indicating abnormality, etc.

4.1.4.5.1 Removal of samples with a sparsely detected number of features (optional)

If multiple features are missing for a particular sample, above a defined threshold (e.g. 50%), then that sample should be considered for removal. A further option, before applying this method to filter samples, is to apply a univariate outlier analysis to flag outlying values of each feature as missing data. Then the filter is applied to remove any samples with a number of missing values above the defined threshold.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including threshold settings for their removal
6. Summary of number of feature intensity values excluded with justification (if relevant); list of outlying samples removed with justification for each. Columns of reported table should include
    a. Sample ID
    b. Justification for removal

4.1.4.5.2 Removal of outlying samples (optional)

Multivariate methods are recommended for outlier detection, e.g. PCA using Hotelling's T^2 distribution on the scores and/or F-tests on the residuals.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)

4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)
6. List of outlying samples removed with ju
7. stification for each. Columns of reported table should include
    a. Sample ID
    b. Justification for removal

4.1.4.6. Normalisation (optional for absolute quantification/semi-quantification)

Normalisation is the process of removing technical or otherwise irrelevant variation from the data on a sample by sample basis. Typically, for relative quantification data, the intensity for each sample is multiplied by a scalar factor, which is different for each sample. Normalisation is usually applied to take account of uncontrolled factors such as dilution or overall instrument response.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)

4.1.4.7. Missing value imputation (optional)

Missing values occur in mass spectrometry metabolomics datasets for a variety of reasons, such as loss of samples, failure to detect a feature in a given sample, or data processing effects. Their presence can significantly affect the performance of the statistical analysis and thus influence the results of the study.

REPORT:

1. Reasons for missing values (if known)
2. Any patterns in missing value distribution (e.g. in m/z, RT, intensity, presence/absence in control group, if known)
3. **S**oftware
4. Software version(s)
5. DOI or URL of the software / script(s) (optional)
6. Algorithm(s)/method(s)
7. Parameter(s)

4.1.4.8. Normality testing, scaling and/or transformations

These processes are applied to each metabolite feature and are particularly important for multivariate analysis. They are typically performed to allow all features to contribute more evenly to a model, or to bring distributions closer to normality. Normality testing is of particular importance for selecting the appropriate statistical approach to use. Types of scaling and transformations include unit variance, Pareto, log, generalised log, range, level and no scaling/transformation. The appropriate type will depend on the nature of the data.

REPORT: Normality testing

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)

REPORT: Scaling/transformation

6. **S**oftware
7. Software version(s)
8. Algorithm(s)/method(s)
9. DOI or URL of the software / script(s) (optional)
10. Parameter(s)

4.1.4.9. Processing methods for metabolite quantification

Due to the importance of the type of metabolite quantification used in a regulatory toxicology study, this section reports the methods used even though they may already have been listed in section 4.1.4.2. Data reduction.

4.1.4.9.1 - Relative quantification - no reference standard available

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)

4.1.4.9.2 - Semi-quantification

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)

4.1.4.9.3 - Absolute quantification

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)

4.1.4.10. Processing methods for metabolite annotation and/or identification

Metabolite annotation/identification for mass spectrometry datasets may involve the use of multiple approaches and/or standards, depending upon the nature of the data.

Metabolite identification (i.e. MSI level 1 identification) can only be achieved by matching the retention time and m/z value(s) of a reference standard (representing a single metabolite) with the retention time and m/z value(s) in a biological sample (representing a single metabolite) - where both the biological sample data and reference standard data were acquired in the same laboratory with the same analytical methods.

Metabolite annotation (i.e. MSI levels 2-3) can be achieved by a plethora of approaches [REFs], some of which use libraries of either in house or public reference standards, and other approaches that do not require libraries of standards at all. In cases where no reference standard is available, annotation approaches include searching either the experimental m/z, calculated neutral mass or calculated molecular formula of the unknown metabolite feature against public and/or commercial libraries of compounds. If fragmentation spectra have been collected for the feature of interest, features can also be annotated to *in silico* fragments and/or predict a metabolite structure using machine learning approaches. In some cases the annotation will only be to a metabolite class level rather than a single metabolite structure.

REPORT

1. Description of metabolite annotation/identification method
2. **S**oftware
3. Software version(s)
4. DOI or URL of the software / script(s) (optional)
5. Algorithm(s)/method(s) - including method for calculating similarity between reference standard and biological sample (e.g. dot-product-cosine) if spectral matching performed
6. Parameter(s)
7. If reference standard used from an in-house or external spectral library(ies) - spectral library(ies) used (and version(s) used - if known)
8. Source of compound library(ies) (if used) (and version(s) used - if known)

4.1.4.11. Data processing workflow(s) (optional)

Either a subset of the data processing or all of the data processing might have been performed in either a dedicated data analysis/processing workflow platform - either open source or proprietary (e.g. Galaxy, KNIME, Nextflow, Compound Discoverer™, Symphony™) - or a programming script(s) that combines multiple tools and/or methods together. Sharing sufficient details to be able to re-run the data processing workflows ultimately improves the reproducibility of data processing.

REPORT

1. Data analysis/processing workflow platform(s) used (e.g. Galaxy, KNIME, Nextflow, Compound Discoverer™, Symphony™)
2. Version of workflow platform(s) used
3. Data processing workflow(s) - either the script(s)/workflow file(s) or a reference to a URL or DOI of the script(s)/workflow file(s)
4. Data processing history(ies) or log(s) (optional)

4.1.5. Demonstration of quality of mass spectrometry metabolomics analysis

This section covers the reporting of data that will allow the quality of the mass spectrometry metabolomics dataset to be assessed by the regulator. Measures of quality are derived from the appropriate uses of different types of QC samples.

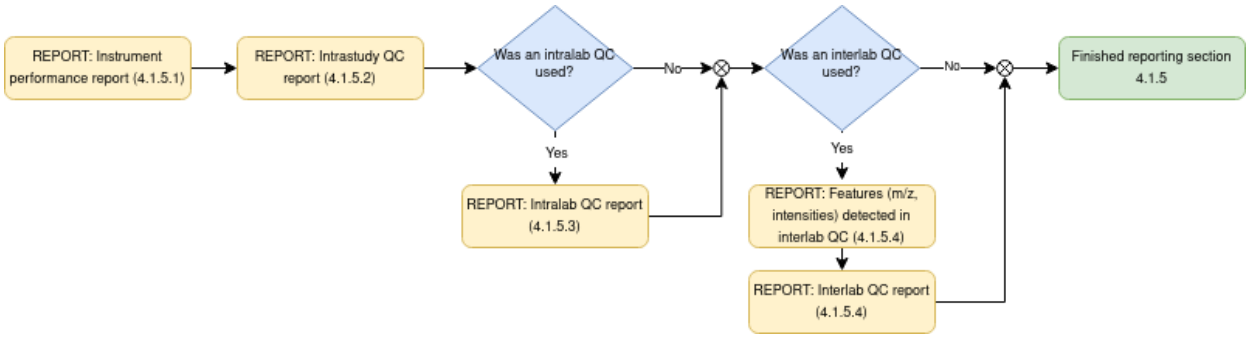

Figure 5: Workflow for mass spectrometry metabolomics reporting - Demonstration of quality of metabolomics analysis

4.1.5.1. Instrument performance report

REPORT:

1. Performance achieved using system suitability QC, relative to a laboratory's acceptance criteria, to confirm instrumentation is fit for purpose. Depending on the assay selected, relevant criteria can include: quantitative reporting of *m/z* shift; retention time shift; shape and/or intensity of selected peaks.

4.1.5.2. Intrastudy QC precision report

REPORT:

1. Measure of a study's analytical precision achieved using intrastudy QC samples, relative to a laboratory's acceptance criteria, to confirm analyses are of sufficient quality for regulatory purposes. To include quantitative reporting of the RSDs of feature intensity (mandatory), *m/z* (optional) and retention time (if relevant; optional), reported as the distribution and median of RSD values of all feature intensities across all intrastudy QC samples; and a qualitative assessment of the similarity of intrastudy QCs using PCA, reported as PCA scores, loadings, variance explained and cross-validation statistics (e.g. Q^2 goodness-of-fit for test samples) from a global analysis of all the biological and intrastudy QC samples.

4.1.5.3. Intralaboratory QC reproducibility report (optional)

REPORT:

1. Measure of intralaboratory (and interstudy) reproducibility, using intralaboratory QC samples, to assess any long term differences between separate studies within the laboratory.

4.1.5.4. Interlaboratory QC reproducibility report (optional)

REPORT:

1. Features (*m/z*, intensities) detected in interlaboratory QC (e.g. defined features in a (standard) reference material). This reporting is likely to evolve as the community improves its use of interlaboratory QC samples.
2. Measure of interlaboratory reproducibility, using interlaboratory QC samples, to assess any differences between separate laboratories; *i.e.* performance standard achieved relative to specified amounts of metabolites.

4.1.6. Outputs: Data matrices from metabolomics assays: sample list, metabolite annotation/identification and intensities

Reporting the *methods* used for metabolite annotation/identification and for determining metabolite intensities was addressed in Section 4.1.3 (analytical) and Section 4.1.4 (computational). Here we describe how to report the *results* of those procedures on a feature-by-feature and/or metabolite-by-metabolite basis. In addition, we describe how to report the level of confidence in annotation/identification, utilising the international criteria established by the Metabolomics Standards Initiative (MSI) in 2007 [ref] that are currently being reviewed by the Metabolite Identification task group of the International Metabolomics Society. We also describe how to report the level of confidence in the quantification of features/metabolites.

Each of these reporting sections (4.1.6.1 - 4.1.6.5) describe data derived from one or more mass spectrometry assays used in the toxicological study.

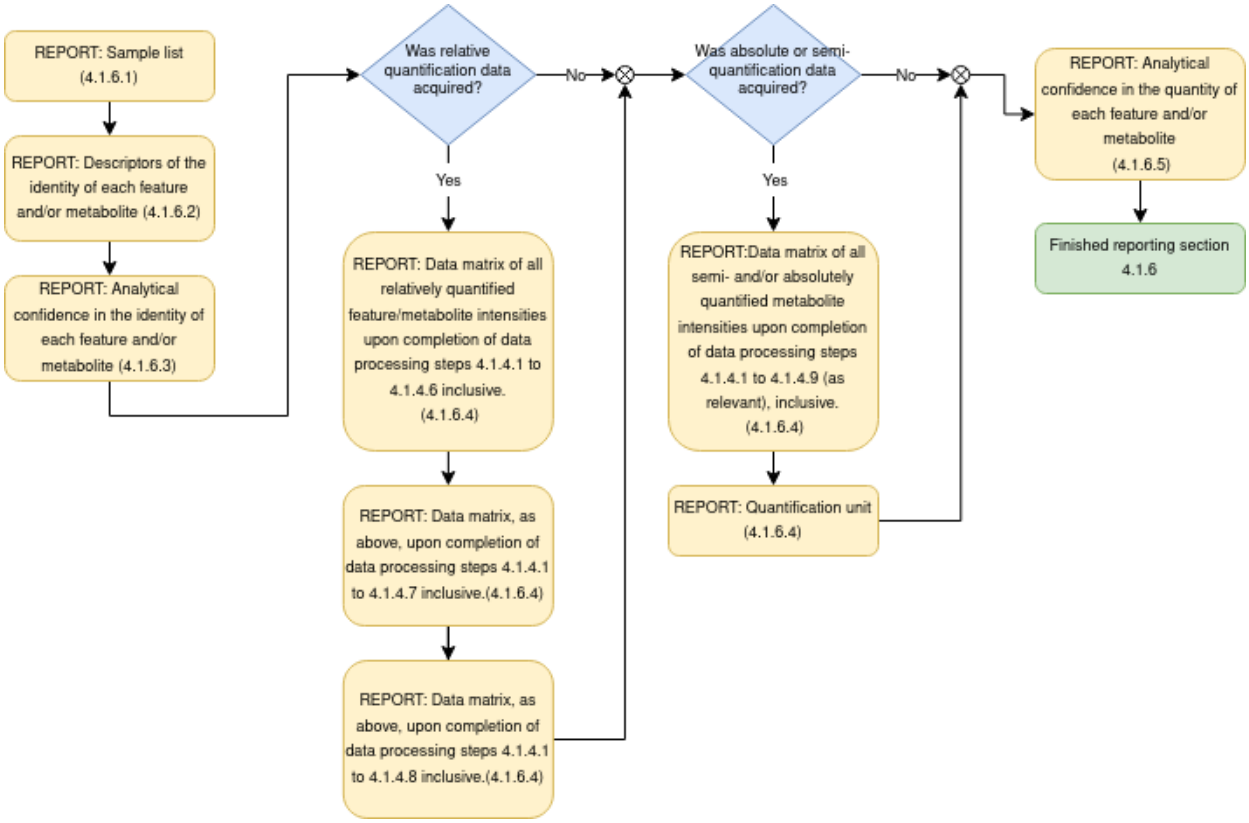

Figure 6: Workflow for mass spectrometry metabolomics reporting - Data matrices from metabolomics assays: sample list, metabolite annotation/identification and intensities

4.1.6.1. Sample list

REPORT

1. For each biological sample
    a. Unique biological sample identifier (e.g. a concatenation of selected parameters, such as: m(001)p(7d)DG0MOAXY where m = male, (001) = animal number, p = matrix (p=plasma), (7d) = day of sampling after study start, DG0 = dose group (0 = control group), MoAXY = study identifier of study XY);
    b. Order of extraction of biological samples;
    c. Order of mass spectrometric data acquisition;
    d. Whether any biological samples were removed from the study and justification for doing so (e.g., outlier detected using PCA due to limited sample volume)**.**

4.1.6.2. Descriptors used to annotate/identify each feature and/or metabolite

REPORT

1. For each feature and/or metabolite:
    a. Analytical identifiers (*m/z*, retention time (if relevant), fragmentation data (optional));
    b. Ion form (*i.e.* adduct, isotope) (if known);
    c. Molecular formula(e) of neutral metabolite (if known);
    d. Neutral mass (if known);
    e. Common metabolite name (if known);
    f. Structural code (e.g. standard InChI string or SMILES) (if known);
    g. Metabolite identifier(s) from relevant database(s) (e.g. PubChem, HMDB);
    h. Common metabolite class name (if relevant).

4.1.6.3. Analytical confidence in the identity of each feature and/or metabolite

REPORT:

1. For each feature and/or metabolite:
    a. MSI level of identification (levels 1-4; [ref MSI paper])
        i. Level 1 - Identified compounds;
        ii. Level 2 - Putatively annotated compounds (e.g. without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries);
        iii. Level 3 - Putatively characterized compound classes (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class);
        iv. Level 4 - Unknown compounds—although unidentified or unclassified these metabolites can still be differentiated and quantified based upon spectral data.
    b. For level 4 only, indicate whether the feature is a ‘known unknown’ or not. The MRF defines a ‘known unknown’ as meaning a consistently observed feature (i.e., consistent *m/z* and retention time) in the sample matrix under investigation that has been detected repeatedly in one or more laboratories.
    c. *m/z* experimental error from MS1 data (optional if threshold reported in 4.1.4.10).
    d. Retention time error (optional if threshold reported in 4.1.4.10).

e. Score for metabolite identification derived from fragmentation data (optional if threshold reported in 4.1.4.10).

4.1.6.4. Descriptors of the quantity of each feature and/or metabolite in each sample

Reporting the quantity of features/metabolites is structured into two parts, one matrix describing all relatively quantified data, and a second matrix describing semi- and/or absolutely quantified data.

Relative quantification:

For a number of reasons, including the potential commercial sensitivity associated with untargeted mass spectrometry measurements of the parent substance and its biotransformation products, the mandatory reporting of raw data is not currently possible. Instead, data should be reported after processing steps 4.1.4.1 to 4.1.4.6 inclusive, which includes removal of features from the dosed chemical.

Furthermore, different statistical analysis methods have differing requirements for the type of processing applied to the metabolomics data. For example, univariate statistical analysis is often applied to the normalised (but not missing value imputed or transformed) data matrix (processed up to and including section 4.1.4.6). Benchmark dosing typically requires the imputation of missing values (processing that includes section 4.1.4.7). Multivariate statistical analysis typically requires the data to be normalised, missing value imputed and transformed (processing that includes section 4.1.4.8). Therefore, multiple data matrices should be reported here.

REPORT:

1. Data matrix of all relatively quantified feature/metabolite intensities for all remaining samples (biological and intrastudy QCs), upon completion of data processing steps 4.1.4.1 to 4.1.4.6 inclusive.
2. Data matrix, as above, upon completion of data processing steps 4.1.4.1 to 4.1.4.7 inclusive.
3. Data matrix, as above, upon completion of data processing steps 4.1.4.1 to 4.1.4.8 inclusive.

Semi- and/or absolute quantification:

REPORT:

4. Data matrix of all semi- and/or absolutely quantified metabolite intensities for all remaining samples (biological and intrastudy QCs), upon completion of data processing steps 4.1.4.1 to 4.1.4.9 (as relevant), inclusive.
5. Quantification unit (e.g. ng/g wet weight).

4.1.6.5. Analytical confidence in the quantity of each feature and/or metabolite

REPORT

1. For each feature and/or metabolite:
    a. Level of quantification (levels 1-3, where 1 is the highest confidence, see Table 1)
        i. Level 1 = absolute quantification
        ii. Level 2 = semi-quantification
        iii. Level 3 = relative quantification

b. RSD of the technical variability of the feature/metabolite intensity derived from repeated measurements of a representative sample (e.g. intrastudy QC sample), where 'representative' means the same sample type and sample matrix as the biological samples.

4.2 NMR Spectroscopy Metabolomics Module

Nuclear magnetic resonance spectroscopy (NMR) is a non-destructive and highly precise analytical tool (albeit less sensitive than mass spectrometry) for the measurement of metabolic phenotypes [ref An early Nicholson paper + W.B. Dunn, N.J.C. Bailey, H.E. Johnson. 2005. Measuring the metabolome: current analytical technologies. The Analyst, 130: 606-625 + Smolinska et al., 2012. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Analytica Chimica Acta 750: 82-97.]. In this **Data Acquisition and Processing Reporting Module** we describe the reporting required for an **NMR-based metabolomics study**, including (1) sample processing, metabolite extraction and addition of standards, (2) analytical QA/QC, (3) acquisition and (4) processing of NMR data, (5) demonstrating the quality of the data, and (6) the data matrices produced by this technology, including metabolite annotation/identification and intensities.

4.2.1. Sample processing, metabolite extraction and addition of chemical reference standards

While the specific protocols for metabolite extraction will depend on the sample type (i.e., biofluid, cells, tissue, etc.) being measured, a reporting framework that can capture the most important information about a diverse range of methods is presented here.

4.2.1.1. Metabolite extraction from biofluids, cells and tissues

Extraction methods differ for liquid, cellular and tissue sample types, though typically require addition of an organic solvent to denature any proteins present and therefore stop any enzymatic activity that would otherwise change the metabolome.

REPORT:

1. Extraction method general description
2. Sample volume(s) used
3. Solvent(s) used
4. Means of agitation/maceration
5. Temperatures and times
6. Post extraction handling, e.g., storage temperature and duration of extracts

4.2.1.2. Extract concentration and reconstitution in solvent for NMR analysis

For many sample types, before analysis, the metabolite extract is evaporated to dryness and reconstituted in (1) deuterated versions of the relevant solvent (e.g., deuterated water), and (2) if necessary, provided with a pH buffer to minimize inter-sample and interstudy inconsistencies in chemical shift values.

REPORT: Evaporation and reconstitution method including reconstitution solvent(s) and volume (including pH buffer type and concentration, if used); storage temperature and duration of reconstituted extracts (if relevant).

1. Evaporation and reconstitution method general description
2. Reconstitution solvent(s) and final volume (if used)
3. pH buffer type and concentration (if used)
4. Storage temperature (if relevant)
5. Duration of reconstituted extracts (if relevant)

4.2.1.3. Addition of reference standard(s)

Reference standards can be added for a variety of purposes including standards for quality assessment, metabolite standards used for annotation/identification purposes only, metabolite standards used for annotation/identification and quantification purposes, and a chemical shift reference standard to facilitate metabolite annotation/identification and optionally for quantification.

4.2.1.3.1. Internal reference standards for assessing the quality of sample extraction (quality assessment reference standards)

Internal standards can provide information on the extraction efficiency for each sample. The internal standard is added at the start of the extraction procedure (pre-spiked) - and serves as an 'extraction standard'.

REPORT:

1. Quality assessment reference standard general preparation description
2. For each standard used - report the following:
    a. Standard use (extraction or injection standard)
    b. Name
    c. PubChem cid (optional)
    d. HMDB id (optional)
    e. KEGG id (optional)
    f. CAS (optional)
    g. InChi (optional)
    h. InChiKey (optional)
    i. SMILES (optional)
    j. Purity (if known)
    k. Supplier
    l. Concentration
    m. Point of addition in sample preparation procedure

4.2.1.3.2. Reference standards to aid metabolite annotation/identification **only**

This refers to metabolite reference standards prepared to accurately identify one or more individual metabolites in a biological sample. These standards are only used for annotation/identification purposes and not used for quantification. All the standards should ideally be acquired on the same instrument type and method as applied to measure the biological samples. The standards can be internal to the biological sample (internal reference standards). Otherwise the reference standard can be spiked into alternative solutions (e.g., buffer only) - these standards are referred to here as external reference standards.

REPORT:

1. Reference standards for metabolite annotation/identification general preparation description

2. For each standard each standard or each named panel of standards used - report the following:
    a. If internal or external standard
    b. Metabolite or metabolite class standard?
    c. Was surrogate used? If yes - then provide detail
    d. Name
    e. PubChem cid (optional)
    f. HMDB id (optional)
    g. KEGG id (optional)
    h. CAS (optional)
    i. InChi (optional)
    j. InChiKey (optional)
    k. SMILES (optional)
    l. Purity (if known)
    m. Supplier
    n. Concentration
    o. Point of addition in sample preparation procedure

Existing in-house as well as external (both public and commercial) NMR spectral libraries of standards can also be used for annotation/identification purposes. In these cases the full details regarding the reference standard preparation might not be known or available but the source of the NMR spectral library and version should be reported in Section 4.2.4.9.

4.2.1.3.3. Reference standards to aid metabolite annotation/identification **and** quantification

This refers to reference standards prepared to identify and quantify one or more metabolites. All of the standards need to be acquired on the same instrument type and method as the biological samples.

REPORT:

1. Reference standards to aid metabolite annotation/identification and quantification preparation details
2. For each standard each standard or each named panel of standards used - report the following:
    a. Type of standard (e.g. external quantification reference such as Bruker QuantRef)
    b. If internal or external standard
    c. Metabolite or metabolite class standard?
    d. Was surrogate used? If yes - then provide detail
    e. Name
    f. PubChem cid (optional)
    g. HMDB id (optional)
    h. KEGG id (optional)
    i. CAS (optional)
    j. InChi (optional)
    k. InChiKey (optional)
    l. SMILES (optional)
    m. Purity (if known)

n. Supplier
o. Concentration
p. Point of addition in sample preparation procedure

4.2.1.3.4. Chemical shift reference standard

An internal chemical shift reference standard is required for spectral alignment and identification of NMR resonances. It is also useful for assessing spectral quality, and may be used for the quantification of metabolites.

REPORT:

1. Chemical shift reference standard preparation details
2. For each standard used - report the following:
    a. Metabolite or metabolite class standard?
    b. Was surrogate used? If yes - then provide detail
    c. Name
    d. PubChem cid (optional)
    e. HMDB id (optional)
    f. KEGG id (optional)
    g. CAS (optional)
    h. InChi (optional)
    i. InChiKey (optional)
    j. SMILES (optional)
    k. Purity (if known)
    l. Supplier
    m. Concentration
    n. Point of addition in sample preparation procedure

4.2.2. Analytical quality assurance and preparation of quality control samples

When conducting an untargeted metabolomics study using NMR, it is essential to have a quality assurance (QA) framework and use quality control (QC) samples, as described in the MERIT best practice guidelines [ref]. Here, the reporting of the analytical QA/QC is described for NMR instrument set up and calibration, and for NMR analysis of a set of biological and QC samples. The QC results are reported in the section 4.2.5 - *Demonstration of quality of metabolomics analysis*, below.

4.2.2.1. System suitability QC sample to assess NMR instrument calibration

For a regulatory toxicology study, a system suitability QC must be used to ensure that sample temperature is properly calibrated, and that adequate water suppression has been achieved (for aqueous samples). Temperature calibration (using deuterated methanol (MeOD)) and water suppression (using a sucrose solution) should be conducted according to the methods described by Dona et al. (ref).

REPORT:

1. Source(s)
2. Preparation details
3. Storage conditions
4. Days since preparation

4.2.2.2. Intrastudy QC sample

Used to provide measures of intrastudy precision and monitor, assess and potentially correct for systematic errors in measurements (e.g., drift in chemical shift, baseline fluctuations, shimming problems). It is essential that the intrastudy QC is highly representative of the biological samples in the study. Typically this type of QC is derived from a small aliquot of the biological samples within the study.

REPORT:

1. Source
2. Preparation details
3. Storage conditions
4. Days since preparation

4.2.2.3. Intralaboratory QC sample (optional)

Used to assess (and potentially correct for) any differences between separate studies within one laboratory. Should be representative of the biological sample type in the study and hence derived from a one-time pool of multiple extracted samples by a specific laboratory using a defined protocol, or a synthetic sample covering the relevant metabolite space, or a reference material of sufficiently similar metabolic composition to the biological sample type.

REPORT:

1. Source
2. Preparation details
3. Batch number
4. Certificate of analysis (COA) - if relevant
5. Storage conditions
6. Days since preparation

4.2.2.4. Interlaboratory QC sample (optional)

Used to assess (and potentially correct for) any differences between individual laboratories. Should be accessible to multiple laboratories, has known providence, is stable, characterised and available in controlled batch numbers. Ideally this type of QC has a similar metabolic composition or matrix to the biological samples in the study.

REPORT:

1. Source
2. Preparation details
3. Batch number
4. Certificate of analysis (COA) - if relevant
5. Storage conditions
6. Days since preparation

4.2.2.5. Process blank sample(s)

There are two types of process blank that are commonly used in NMR metabolomics. The first is used to provide a measure of any background contamination arising from the reconstitution solvent/buffer itself, hence the 'process' includes making the solvent/buffer, adding it to an NMR tube, and the NMR analysis. This is often termed a 'buffer blank'. The second process blank is used to provide a measure of background contamination arising from the extraction of the biological sample as well as from the NMR solvent/buffer, hence the process includes extraction, making the solvent/buffer, adding that solution to the extracted sample, and the NMR analysis. This is often termed an 'extraction blank'.

REPORT:

1. Type of process blank
2. Start and end points of the 'process' used to prepare
3. Storage conditions

4.2.3. Acquisition of NMR metabolomics and metabolite data

Despite shortcomings in sensitivity, NMR metabolomics can reliably detect a variety of metabolites with outstanding precision and robustness (ref). Furthermore, NMR is capable of detecting a considerable range of nuclei (^{1}H, ^{13}C, ^{15}N, etc.) and collecting valuable chemical structural information using two dimensional (2D) experiments (both homonuclear and heteronuclear). However, NMR spectra collected from one dimensional (1D) ^{1}H-NMR experiments are the primary source of metabolomics and metabolite data. Thus, the reporting requirements detailed in this document will be restricted to instrument configuration and calibration, and the subsequent data acquisition and processing associated with 1D ^{1}H-NMR spectra [ref MERIT].

4.2.3.1. Acquisition order for QC samples, biological samples and reference standard samples

REPORT:

1. Acquisition order of all types of QC samples, biological samples and reference standard samples (if relevant), thereby indicating the number of process blank samples, and the frequency of analysis of intrastudy QC samples, etc. Columns of reported table should include:
    a. Run order
    b. File name
    c. Sample type (must be able to distinguish between QC samples, biological samples and process blanks)

4.2.3.2. NMR spectroscopy assay type(s)

NMR spectroscopy metabolomics assays can be untargeted (i.e. the traditional approach in which no particular metabolites are pre-selected for study) or targeted (e.g. Bruker B.I.Methods™ such as B.I.QUANT-UR that targets and quantifies up to 150 metabolites). Each assay can have one or more levels of metabolite annotation/identification and quantification. A single toxicology study can comprise a combination of several assays.

REPORT:

1. NMR spectroscopy assay type(s) used

2. NMR spectroscopy assay description

4.2.3.3. NMR instrument configuration

REPORT:

1. NMR magnet manufacturer
2. NMR magnet model number/name
3. Magnetic field strength as proton NMR frequency
4. NMR console manufacturer
5. NMR model number/name
6. Probe type (e.g. 5 mm RT probe)
7. Probe manufacturer
8. Probe model number/name
9. Software packages and version number(s)

4.2.3.4. NMR method

REPORT:

1. Sample temperature °C
2. Pulse sequence type (e.g., 1D ^1H NOESY presat)
3. Sweep width (ppm)
4. Pre-delay (i.e. relaxation delay) (seconds)
5. Carrier frequency (MHz)
6. 90-degree pulse width (microseconds)
7. Number of accumulated scans
8. Acquisition time (seconds)
9. confirm sample spinning not used.

4.2.4. Processing of NMR metabolomics and metabolite data

This section covers the reporting of established processes for converting raw NMR data acquired on an instrument to processed NMR data that is in a form amenable for statistical analysis. This processing typically produces a 2-dimensional (2D) data table with each sample represented in one dimension (typically a row) and its resonance intensity values in the other dimension (typically a column). For a traditional untargeted NMR metabolomics study (for which no metabolites were pre-selected for analysis and the NMR pulse sequence uses a short relaxation delay), the entries in the data table typically indicate the relative quantities of the metabolites in each sample. This can be in the form of 'binned' (or 'bucketed') data, for which there can be several bins per metabolite, or in the form of one relative quantity per metabolite, depending on how the data is processed. For a targeted NMR metabolomics study (targeting a defined list of metabolites), the entries in the data table typically describe the absolute concentrations of the metabolites. In general, only 1D ^{1}H-NMR data are acquired on each sample, with 2D data acquired primarily for metabolite identification purposes (with the potential exception of 2D *J*-resolved NMR spectroscopy which is gaining in popularity as a high throughput NMR metabolomics method (ref C. Ludwig and M.R. Viant. 2010. Two-dimensional J-resolved NMR spectroscopy: Review of a key methodology in the metabolomics toolbox. Phytochemical Analysis, 21(1): 22-32.)). Hence, here we focus on reporting for 1D ^{1}H-NMR data processing, based largely on the MERIT guidelines [ref].

4.2.4.1. Spectral pre-processing

In comparison to mass spectrometry metabolomics datasets, the pre-processing steps for NMR are relatively straightforward. At a minimum, these include application of an apodisation function prior to Fourier transformation, phase and baseline correction, and chemical shift calibration. These steps are essential for the pre-processing of raw NMR spectra prior to subsequent metabolomic data analyses and thus the parameters used should be reported as part of any NMR-based metabolomics study. Other aspects of pre-processing such as zero-filling and linear prediction are occasionally employed and should be reported when used.

REPORT:

1. **S**oftware used for NMR spectral pre-processing
2. Software version(s) used for NMR spectral pre-processing
3. DOI or URL of the software / script(s) (optional)
4. Forward and/or backward linear prediction and the number of points for each (if performed).
5. Window function type and magnitude used for apodisation.
6. Zero-filling original and final points count (if performed).
7. Phasing method (manual or automatic) and parameters.
8. Baseline correction method and parameters.
9. Chemical shift calibration method used (manual or automatic).

4.2.4.2. Data reduction

This section covers the process of converting pre-processed spectral data into a tabular form for statistical analysis. Steps to report in this process include: (1) peak alignment and matching, followed by either (2) use of full resolution NMR spectra, (3) binning, and/or (4) peak fitting.

REPORT: Peak alignment and matching (if performed).

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)

REPORT: Data reduction (i.e. binning)

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s) e.g. binning (fixed or variable width) using Bruker AMIX, Mnova NMR, etc.
5. Parameter(s)
6. Report the use of full resolution NMR spectra if data reduction was not performed.

REPORT: Peak fitting

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s) e.g. using Bruker B.I.QUANT-UR software, Chenomx NMR Suite, etc.
5. Parameter(s)

4.2.4.3. Resonance intensity drift and/or batch correction

While every effort should be made experimentally to minimise variations in signal intensity within and between batches, these can be (partially) corrected in the data processing step, typically using signals recorded in the intrastudy QC samples.

4.2.4.3.1 Assessment of the presence of within- and/or between-batch signal intensity drift (optional)

Evaluation of the presence of such signal intensity drift typically includes a PCA analysis, showing both biological and intrastudy QC samples and relative standard deviation measurements (RSD) for some metabolites present in QC samples (covering a wide range of physicochemical properties and concentrations). If the intrastudy QC samples show significant differences in scores or RSD values within any batch, or a drift in scores values or RSD values between batches, then drift and/or batch effects are present.

REPORT:

1. PCA scores plots showing biological and intrastudy QC samples before any correction applied
2. PCA scores plot of just the intrastudy QC samples (optional).

4.2.4.3.2 Batch correction (optional)

If used, the effects of batch correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the batch correction.

REPORT:

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)
6. Evidence that batch correction has been successful (e.g. PCA scores plots showing intrastudy QCs).

4.2.4.3.3 Signal intensity drift correction within a batch (optional)

If used, the effects of drift correction should be demonstrated, for example with a PCA analysis and by inspecting the intensities of representative analytes, both before and after the correction is made.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)

5. Parameter(s)
6. Evidence that the correction has been successful (e.g. PCA scores plots showing intrastudy QCs).

4.2.4.4. Identification and removal ('filtering') of NMR resonances

Untargeted NMR analysis may detect certain resonances that should be considered for removal to improve the overall reliability of the data set.

4.2.4.4.1 Removal of resonances present in process blanks (optional)

Resonances detected in process blanks are thought to result from the solvents or plasticware rather than the biological sample and therefore can be considered for removal from the final data set.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including threshold settings for their removal
6. Resonances (chemical shift range(s)) detected in process blank(s), e.g., in solvent/buffer blank and extraction blank

4.2.4.4.2 Removal of resonances from dosed substances (optional)

In toxicology studies, in which the biological system is deliberately exposed to an exogenous chemical, the dosed parent substance and/or its biotransformation products may be observed in the resulting data. For the purposes of determining the endogenous metabolic effect of the exposure, it is important that these resonances are removed from the data set. The appropriate resonances for removal are typically determined by comparison to control spectra, spectra from a chemical standard, and/or the literature.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including threshold settings for their removal
6. Resonances (chemical shift range(s)) detected from dosed substances.

4.2.4.5. Identification and removal ('filtering') of outlying samples

It is important to identify potential outlying samples and to remove them, if necessary, before applying statistical analysis. Reasons for removing each outlier must be clearly explained in terms of biological, analytical or data analytical aspects; e.g. sample removed due to limited volume, contamination was detected, results of an independent assay indicating abnormality, etc.

4.2.4.5.1 Removal of samples with a sparsely detected number of resonances (optional)

If multiple resonances are missing for a particular sample, above a defined threshold (e.g. 50%), then that sample may be considered for removal. A further option, before applying this method to filter samples, is to apply a univariate outlier analysis to flag outlying values of each resonance as missing data. Then the filter is applied to remove any samples with a number of missing values above the defined threshold.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s) - including threshold settings for their removal
6. Summary of number of feature intensity values excluded with justification (if relevant); list of outlying samples removed with justification for each. Columns of reported table should include
    a. Sample ID
    b. Justification for removal

4.2.4.5.2 Removal of outlying samples (optional)

Multivariate methods are recommended for outlier detection, e.g. PCA using Hotelling's T^2 distribution on the scores and/or F-tests on the residuals.

REPORT:

1. **S**oftware
2. Software version(s)
3. Algorithm(s)/method(s)
4. DOI or URL of the software / script(s) (optional)
5. Parameter(s)
6. List of outlying samples removed with justification for each. Columns of reported table should include
    a. Sample ID
    b. Justification for removal

4.2.4.6. Normalisation

Normalisation is the process of removing technical or otherwise irrelevant variation from the data on a sample by sample basis. Typically, intensity data for each sample is multiplied by a scalar factor, which is different for each sample. Normalisation is usually applied to take account of uncontrolled factors such as dilution or overall instrument response.

REPORT:

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)

4.2.4.7. Normality testing, scaling and/or transformations

These processes are applied to each metabolite resonance and are particularly important for multivariate analysis. They are typically performed to allow all resonances to contribute more evenly to a model, or to bring distributions closer to normality. Normality testing is of particular importance for selecting the appropriate statistical approach to use. Types of scaling and transformations include unit variance, Pareto,

log, generalised log, range, level and no scaling/transformation. The appropriate type will depend on the nature of the data.

REPORT: Normality testing

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)

REPORT: Scaling/transformation

6. **S**oftware
7. Software version(s)
8. DOI or URL of the software / script(s) (optional)
9. Algorithm(s)/method(s)
10. Parameter(s)

4.2.4.8. Processing methods for metabolite quantification

Due to the importance of the type of metabolite quantification used in a regulatory toxicology study, this section reports the methods used even though they may already have been listed in section 4.2.4.2. *Data reduction*.

4.2.4.8.1 - Relative quantification

REPORT:

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)

4.2.4.8.2 - Absolute quantification

REPORT:

1. **S**oftware
2. Software version(s)
3. DOI or URL of the software / script(s) (optional)
4. Algorithm(s)/method(s)
5. Parameter(s)

4.2.4.9. Processing methods for metabolite annotation and/or identification

Metabolite annotation/identification for NMR datasets may involve the use of multiple data types (e.g., 1D NMR spectra of reference compounds, 2D homonuclear and/or heteronuclear NMR spectra, etc.). As a

result, the processing method(s) used to facilitate metabolite annotation/identification will depend upon the nature of the NMR data.

4.2.4.9.1 - Method of annotating/identifying metabolite resonances when a metabolite reference standard is used

Metabolite identification based on the match of chemical shift value(s) of a reference standard with those of the biological sample.

REPORT:

1. Type(s) of NMR data (e.g., 2D, 1D, etc.) used for metabolite annotation/identification.
2. **S**oftware
3. Software version(s)
4. DOI or URL of the software / script(s) (optional)
5. Algorithm(s)/method(s) for assessing similarity between reference standard and biological sample, and parameters used.
6. Parameter(s)
7. Whether reference standard was obtained from in-house or external spectral library
8. If external spectral library(ies) were used - spectral library(ies) used (and version(s) - if known).

4.2.4.9.2 - Method of annotating metabolite resonances when no metabolite reference standard is available

NMR spectroscopy is a powerful analytical tool for the *de novo* determination of the structure of small molecule metabolites. While this requires specialist expertise, it is a viable strategy for reporting of an NMR metabolomics study.

REPORT:

1. Type(s) of NMR data (e.g., 2D, 1D, etc.) used for metabolite annotation/identification.
2. **S**oftware
3. Software version(s)
4. DOI or URL of the software / script(s) (optional)
5. Algorithm(s)/method(s)
6. Parameter(s)

4.2.4.10. Data processing workflow(s) (optional)

Either a subset of the data processing or all of the data processing might have been performed in either a dedicated data analysis/processing workflow platform - either open source or proprietary (e.g. Galaxy, KNIME, Nextflow) - or a programming script(s) that combines multiple tools and/or methods together. Sharing sufficient details to be able to re-run the data processing workflows greatly improves the reproducibility of the data processing.

REPORT

1. Data analysis/processing workflow platform(s) used (e.g. Galaxy, KNIME, Nextflow)
2. Version of workflow platform(s) used
3. Data processing workflow(s) - either the script(s)/workflow file(s) or a reference to a URL or DOI of the script(s)/workflow file(s)
4. Data processing history(ies) or log(s) (optional)

4.2.5. Demonstration of quality of NMR metabolomics analysis

This section covers the reporting of data that will allow the quality of the NMR metabolomics dataset to be assessed by the regulator. Measures of quality are derived from the appropriate uses of different types of QC samples.

4.2.5.1. NMR performance report

REPORT:

1. Performance achieved using system suitability QC, relative to a laboratory's acceptance criteria, to confirm instrumentation is fit-for-purpose. To include quantitative reporting of chemical shift reference peak full width at half maximum height (FWHM) without window function applied, 90 degree pulse width, and probe temperature variation (plus or minus value in ^{o}C). Reporting an image(s) of aligned and stacked spectra for visual inspection is recommended to facilitate an assessment of the general quality of all spectra.

4.2.5.2. Intrastudy QC precision report

REPORT:

1. Measure of a study's analytical precision achieved using intrastudy QC samples, relative to a laboratory's acceptance criteria, to confirm analyses are of sufficient quality for regulatory purposes. To include quantitative reporting of the RSDs of resonance intensities, reported as the distribution and median of RSD values of all resonance intensities across all intrastudy QC samples; and a qualitative assessment of the similarity of intrastudy QCs using PCA, reported as PCA scores, loadings, variance explained and cross-validation statistics (e.g. Q^2 goodness-of-fit for test samples) from a global analysis of all the biological and intrastudy QC samples.

4.2.5.3. Intralaboratory QC reproducibility report (optional)

REPORT:

1. Measure of intralaboratory (and interstudy) reproducibility, using intralaboratory QC samples, to assess any long-term differences between separate studies within the laboratory.

4.2.5.4. Interlaboratory QC reproducibility report (optional)

REPORT:

1. Resonances detected in interlaboratory QC (e.g. defined features in a (standard) reference material). This reporting is likely to evolve as the community improves its use of interlaboratory QC samples.
2. Measure of interlaboratory reproducibility (using interlaboratory QC samples) to assess any differences between separate laboratories; *i.e.* performance standard achieved relative to specified amounts of metabolites.

4.2.6. Data matrices from metabolomics assays: sample list, metabolite annotation/identification and intensities

4.2.6.1. Sample list

REPORT

1. For each biological sample:
    a. Unique biological sample identifier (e.g. a concatenation of selected parameters, such as: m(001)p(7d)DG0MOAXY where m = male, (001) = animal number, p = matrix (p=plasma), (7d) = day of sampling after study start, DG0 = dose group (0 = control group), MoAXY = study identifier of study XY);
    b. Order of extraction of biological samples;
    c. Order of NMR data acquisition;
    d. Whether any biological samples were removed from the study and justification for doing so (e.g., outlier detected using PCA due to limited sample volume)**.**

4.2.6.2. Descriptors used to annotate/identify each resonance and/or metabolite

REPORT

1. For each resonance and/or metabolite:
    a. Chemical shift value(s) (ppm) for all relevant nuclei;
    b. Multiplicity/splitting pattern (e.g., doublet) for relevant ^{1}H peak;
    c. Molecular formula (if known);
    d. Common metabolite name (if known);
    e. Structural code (e.g. standard InChI string or SMILES) (if known);
    f. Metabolite identifier(s) from relevant database(s) (e.g. PubChem, HMDB);
    g. Common metabolite class name (if relevant).

4.2.6.3. Analytical confidence in the identity of each feature and/or metabolite

REPORT

1. For each feature and/or metabolite:
    a. MSI level of identification (levels 1-4; [ref MSI paper])
        i. Level 1 - Identified compounds;
        ii. Level 2 - Putatively annotated compounds (e.g. without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries);
        iii. Level 3 - Putatively characterized compound classes (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class);
        iv. Level 4 - Unknown compounds—although unidentified or unclassified these metabolites can still be differentiated and quantified based upon spectral data.
    b. For level 4 only, indicate whether the feature is a ‘known unknown’ or not. The MRF defines a ‘known unknown’ as meaning a consistency observed feature in the sample matrix under investigation that has been detected repeatedly in one or more laboratories.

4.2.6.4. Descriptors of the quantity of each feature and/or metabolite in each sample

Reporting the quantity of features/metabolites is structured into two parts, one matrix describing all relatively quantified data, and a second matrix describing absolutely quantified data.

Relative quantification:

For a number of reasons, including the potential commercial sensitivity associated with untargeted NMR measurements of the parent substance and its biotransformation products, the mandatory reporting of raw NMR data is not currently possible. Instead, data should be reported after processing steps 4.2.4.1 to 4.2.4.6 inclusive, which includes removal of resonances from the dosed chemical.

Furthermore, different statistical analysis methods have differing requirements for the type of processing applied to the metabolomics data. For example, univariate statistical analysis is often applied to the normalised (but not missing value imputed or transformed) data matrix (processed up to and including section 4.2.4.6). Multivariate statistical analysis typically requires the data to be normalised and transformed (processing that includes section 4.2.4.7). Therefore, multiple data matrices should be reported here.

REPORT:

1. Data matrix of all relatively quantified resonance/metabolite intensities for all remaining samples (biological and intrastudy QCs), upon completion of data processing steps 4.2.4.1 to 4.2.4.6 inclusive.
2. Data matrix, as above, upon completion of data processing steps 4.2.4.1 to 4.2.4.7 inclusive.

Absolute quantification:

REPORT:

3. Data matrix of all absolutely quantified metabolite intensities for all remaining samples (biological and intrastudy QCs), upon completion of data processing steps 4.2.4.1 to 4.2.4.8 (as relevant), inclusive.
4. Quantification unit (e.g. ng/g wet weight).

4.2.6.5. Analytical confidence in the quantity of each resonance and/or metabolite

REPORT

1. For each resonance and/or metabolite:
    a. Level of quantification (levels 1-3, where 1 is the highest confidence)
        i. Level 1 = absolute quantification
        ii. (Level 2 not used)
        iii. Level 3 = relative quantification
    b. RSD of the technical variability of the resonance/metabolite intensity derived from repeated measurements of a representative sample (e.g. intrastudy QC sample), where ‘representative’ means the same sample type and sample matrix as the biological samples.

5. Data Analysis Reporting Modules

Statistical analysis will depend heavily on the objective and design of the study. For example, a single treatment versus control study to discover a chemical's MoA requires a two-group analysis, while an experiment to derive molecular points-of-departure requires dose-response modelling. Best practices for the statistical analysis of metabolomics datasets are not yet available for all applications of this technology in regulatory toxicology. However, some statistical analyses are generic and hence applicable to many study designs, such as univariate and multivariate statistical analyses. Consequently, Section 5 describes two Data Analysis Reporting Modules, one for the application of univariate statistical analysis to discover 'differentially abundant metabolites' (Section 5.1), and the second describing multivariate statistical analysis of metabolomics data (Section 5.2).

5.1 Differentially Abundant Molecules (using univariate methods) Module

This multi-omics compliant module specifies the information needed by a regulatory scientist in order to assess univariate statistical analysis as used to discover differentially abundant molecules, i.e., for metabolomics, to discover differentially abundant features and/or metabolites (DAMs).

The goals of this module are to specify what needs to be reported while not being prescriptive about the analysis performed. We leave judgment about the appropriateness of the analysis plan to the operator and the regulatory scientist.

5.1.1. Statistical software

The requirements in this section are intended to facilitate reproducibility of analyses and make quality assessment possible. Operators are free to choose freely available open source, commercially available, or proprietary software. Operators must specify the name and version of all software that is used to discover DAMs. Operators must also specify all add-ons, plug-ins, packages, or libraries (hereafter "libraries") that are called, used, required, loaded, or otherwise brought into the software for use in discovering DAMs. This specification shall include the name and version of such libraries, and a hyperlink to download the libraries. In those cases where the libraries are locally developed and controlled by the operator, such that a hyperlink to download does not exist, the operator must specify the following: "this library is not available for download" and should then make the code available as specified below..

REPORT:

REQUIRED:

1) Name(s) and version(s) of the software used.
2) Name(s), version(s), and hyperlink(s) for download of all add-ons, plug-ins, packages, or libraries. If download hyperlink is not available, the operator shall state "this library is not available for download".

OPTIONAL:

3) If download hyperlink is not available, the operator should make source code and binary (if applicable) available as specified later in this module

5.1.2. Contrasts for which differentially abundant molecules were identified

To ensure clarity, operators must specify the contrasts (or group comparisons) that were performed to discover the DAMs. Generally, for simple control vs treated study designs, the contrast of interest is treated vs control. This can become more complicated when the number of treatment levels increase, or if a time-course design is being used. In cases where the term "control" can be confusing, the operator must use a more specific term. For instance, if the study design uses a laboratory control and a field control then the operator must specify whether the laboratory control or the field control is being used, or if it is some transformation based on these controls. For instance, "treated time 12-hr vs field control 12-hr" is clearer than stating "treated time 12-hr vs control" in the case where control could refer to laboratory control or field control. For the purpose of clarity, the *factor* (alternatively called an independent variable) is an explanatory variable manipulated by the experimenter. Each factor has two or more *levels* (i.e., different values of the factor). Combinations of factor levels are often called treatments.

REPORT:

REQUIRED:

1) A table or listing that must include:
    a) Each factor, and level within each factor.

5.1.3. Assay experimental design

In order to ensure that the omics data were analysed properly, the regulatory scientist must have a clear description of which samples were used in the DAM analysis, how many samples there are per experimental group, and how they may be assayed together/measured simultaneously or not. In addition, the regulatory scientist can use this information to perform post hoc power analyses, as well as estimates for Type M (magnitude) and Type S (sign) error, which are equally, if not more, informative in assessing data quality and the likelihood of drawing erroneous conclusions. For more on Type M and Type S errors, see Gelman and Carlin (2014).

REPORT:

REQUIRED:

1) A description or table that must include:
    a) Each group or factor listed per section 5.1.2 in MRF and:
        i) The number of samples used for discovering DAMs in each group or factor at a minimum. Should include the sample IDs that can be used to match across the report.
        ii) Identification of which samples may exhibit covariance (examples include: assayed on the same day, assayed in parallel, assayed on the same physical platform (e.g., chip/glass slide, cartridge, etc. for transcriptomics; LC-MS analytical batch, etc. for metabolomics), processed using the same wash solutions or reagents, processed using the same master mixes (transcriptomics only), littermates in the exposure study, animals which are housed together, etc.).

iii) An explicit statement if there is no reason to believe covariance exists between samples, and a justification to support this assertion.

b) Identification if any samples are technical replicates of each other, and how those samples are technical replicates (i.e., define what makes the samples technical replicates).

5.1.4. Statistical analysis to identify differentially abundant molecules

Today there are a myriad ways for discovering DAMs, all with their own strengths and weaknesses. It is also recognised that each experimental design may require its own set of assumptions and caveats in the analysis that makes it difficult to prescribe a universal standard. Thus, it is critical that the operator clearly communicates to the regulatory scientist exactly how the statistical analysis was performed so that they can understand what was done and establish if the approach taken was sound. To accomplish this, the operator needs to supply the statistical analysis plan. The following specifies the minimum requirements for a sound statistical analysis plan.

REPORT:

REQUIRED:

1) The name and description of the statistical approach
    a) Operators must clearly state all data transformations (e.g., log2 transformation) that are performed in the course of analysis. Operators must also declare if no data transformations were performed to ensure clarity.
    b) If using a general linear model, general linear mixed model, ANOVA, or something similar, the operator must specify the effects being modeled including all fixed and random effects, as well as their interactions, if any, and any nesting.
    c) If using a pairwise comparison approach, such as a Wald's test, Student's or Welch's t-test, or non-parametric analogues, the operator must specify what values are being used (model-based values from methods in section 5.1.4.1.b for MRF or the transformed/non-transformed values without adjustment from a model).
    d) If using a Bayesian approach, then standard reporting requirements are required, including specification of any priors, explicit specification of the posterior, specification of what is being modelled to discover the DAMs.
2) Specify the decision criteria for discovering the DAMs
    a) If using a p-value criteria
        i) Is multiple-testing correction being performed?
            (1) Yes: specify the nominal alpha value, the multiple-testing correction method (exact type, e.g. Bonferroni for family wise error rate correction or Benjamini & Hochberg for false discovery rate control), and any adjusted threshold value,
            (2) No: specify the nominal alpha value and the p-value threshold. In addition, justify why a multiple-testing correction has not been performed.

b) If using a fold-change or log fold-change criterion, specify the level (e.g., 2x change).
c) If using another approach for ranking and prioritization, specify the procedure clearly such that anyone could replicate the work should it be necessary.
d) If using a mixture of approaches (e.g., fold-change followed by p-value) then specify the exact order of operations and how the decision criterion is applied – this should be written clearly such that anyone could replicate this work should it be necessary.

5.1.5. Outputs

Because there are numerous approaches for the analysis and modelling of DAM data, there are also numerous formats for outputting this information. It would be impossible to enumerate all of the potential output types, styles, and other information here. Instead, we list general guidance regarding outputs that may be submitted in support of a regulatory application.

REPORT:

REQUIRED:

1. Operator must submit a file manifest (either within a report, or as a separate file titled MANIFEST.xyz where .xyz is an appropriate file extension (.txt if a text file, .md if a markdown file, .html if an html file, .pdf if a pdf file)
    a. The manifest must include the listing of all files included in the regulatory application package specific to the DAM data.
    b. Each file in the manifest must be accompanied by a description of the file
        i. If the file being described is a tabular file, then the rows and columns must be described so that anyone can understand the contents of the file.
    c. The manifest file must be in a standard format that is easily readable in non-proprietary software.
    d. The manifest file must not be in a binary format, such as Word, Excel, or other formats that require special software that is not freely available, or where the format might change significantly and not be backwards compatible (e.g., Excel and Word files are binary formats and although freeware exists that can read these files, they may not be able to be read easily in the future).

5.2. Multivariate Statistical Analysis Module

In multivariate analysis, the metabolic variables (e.g. peak intensities or metabolite abundances) are analysed simultaneously.

5.2.1. Software, method and parameters

The requirements in this section are intended to facilitate reproducibility of analyses and make quality assessment possible. Operators are free to choose freely available open source, commercially available, or proprietary software. Operators must specify the name and version of all software that is used. Operators must also specify all add-ons, plug-ins, packages, or libraries (hereafter "libraries") that are called, used, required, loaded, or otherwise brought into the software. This specification shall include the name and version of such libraries, and a hyperlink to download the libraries (or if download not available - hyperlink for further information on the software and/or library). In those cases where the libraries are locally developed and controlled by the operator, such that a hyperlink to download does not exist, the operator must specify the following: "this library is not available for download" and should then make the code available as specified below.

REPORT:

REQUIRED:

(1) Name(s) and version(s) of software used.
(2) Name(s), version(s), and hyperlink(s) for download of all add-ons, plug-ins, packages, or libraries. If download hyperlink is not available, the operator shall state "this library is not available for download".

OPTIONAL:

(3) If download hyperlink is not available, the operator should make source code and binary (if applicable) available. If software, add-ons, plug-ins, packages or libraries are not open source and/or the source code cannot be shared then a relevant hyperlink to the software website should be provided.

5.2.2. Experimental conditions for which differentially abundant molecules were identified

To ensure clarity, operators must specify the conditions which were included or compared) in the multivariate analysis. For unsupervised analysis, specific contrasts may not be defined (e.g. one may just model the control group). For supervised analysis, the operator will define specific conditions which are compared or specific factors which are examined (e.g. dose). For simple control vs treated study designs, the contrast of interest is treated vs control. This can become more complicated when the number of treatment levels increase, or if a time-course design is being used. In cases where the term "control" can be confusing, the operator must use a more specific term. For instance, if the study design uses a laboratory control and a field control then the operator must specify whether the laboratory control or the field control is being used, or if it is some transformation based on these controls. For instance, "treated time 12-hr vs field control 12-hr" is clearer than stating "treated time 12-hr vs control" in the case where control could refer to laboratory control or field control. For the purpose of clarity, the *factor* (alternatively called an independent variable) is an explanatory variable usually manipulated by the experimenter. Discrete factors have two or more *levels* (i.e., different values of the factor). Continuous factors (e.g. time) may take a range of values. Combinations of factor levels are often called treatments.

REPORT:

REQUIRED:

1) A table or listing that must include:
    a) Each factor, and for discrete factors, the levels within each factor.

5.2.3. Assay experimental design

In order to ensure that the omics data were analysed properly, the regulatory scientist must have a clear description of which samples were used in the multivariate analysis, how many samples there are per experimental group, and how they may be assayed together/measured simultaneously or not. In addition, the regulatory scientist can use this information to perform post hoc power analyses, as well as estimates for Type M (magnitude) and Type S (sign) error, which are equally, if not more, informative in assessing data quality and the likelihood of drawing erroneous conclusions. For more on Type M and Type S errors, see Gelman and Carlin (2014).

REPORT:

REQUIRED:

1) A description or table that must include:
    a) Each group or factor listed per section 5.2.2 in MRF and:
        i) The number of samples used in each group or factor at a minimum. Should include the sample IDs that can be used to match across the report.
        ii) Identification of which samples may exhibit covariance (examples include: assayed on the same day, assayed in parallel, assayed on the same physical platform (e.g., chip/glass slide, cartridge, etc. for transcriptomics; LC-MS analytical batch, etc. for metabolomics), processed using the same wash solutions or reagents, processed using the same master mixes (transcriptomics only), littermates in the exposure study, animals which are housed together, etc.).
        iii) An explicit statement if there is no reason to believe covariance exists between samples, and a justification to support this assertion.
    b) Identification if any samples are technical replicates of each other, and how those samples are technical replicates (i.e., define what makes the samples technical replicates).

5.2.4. Multivariate statistical analysis – unsupervised

Many different types of unsupervised multivariate analysis can be applied. The goal of the unsupervised analysis is to provide an overview of the data to explore structures such as the major sources of variance, clustering or trends. These structures may result from, but are not limited to, the following: the biological effects being studied (e.g. dose effect), uncontrolled natural biological variation, or residual variance in the analytical procedure.

REPORT:

REQUIRED:

1) Name and description of the multivariate method
2) The groups or conditions included in the analysis
3) If variable selection used
    a) The name and description of the variable selection method
    b) Selected variables

4) Parameter settings (e.g. scaling method)
5) Output of the multivariate method including plots (e.g. PCA scores, loadings, proportion of variance explained).
6) Methods should include clear citations to literature describing them

5.2.5. Multivariate statistical analysis – supervised

Supervised methods (usually classification or regression methods) are commonly used to focus the analysis on specific questions, e.g. whether the metabolic data can classify a sample into a chemical MoA, and to find which metabolic variables are most responsible for this classification. These methods are able to model data in cases where the treatment effect is small compared to other sources of variation. Many methods exist, but the chosen method should exhibit an ability to handle data with a) many variables, b) high degree of correlation between variables, c) high levels of noise, and d) missing data (if any). Common methods include partial least squares (PLS) or Orthogonal PLS (OPLS) regression when the outcome is continuous, or the equivalent discriminant analysis (PLS-DA and OPLS-DA) when the outcome is discrete (e.g. classification).

REPORT:

REQUIRED:

1) Name and description of the multivariate method used
2) The groups, conditions or factors contrasted or examined in the analysis
3) If variable selection used
    a) The name and description of the variable selection method used
    b) Selected variables
4) Parameter settings (e.g. scaling method)
5) Output of the multivariate method including plots (e.g. PLS scores, loadings/weights, regression coefficients, proportion of variance in outcome explained by the model).
6) Methods should include clear citations to literature describing them

5.2.6. Multivariate statistical analysis - validation

All models, both supervised and unsupervised, must be statistically validated to show that they are robust and predictive. This should be done by either a) separating the metabolomics data into independent training (typically a maximum of 70% of dataset) and test sets (remaining 30% minimum of dataset), or b) internal cross-validation. In both cases, summary statistics such as Q^2 or misclassification rate should be calculated. Model complexity (e.g. number of principal components) should be chosen based on predictivity of the model.

REPORT:

REQUIRED:

1) Name and description of the model validation method used
2) Parameter settings
3) Validation statistics (Q2, error rate, etc.)
4) Methods should include clear citations to literature describing them

5.2.7. Multivariate statistical analysis - variable importance for feature selection

One of the main objectives of multivariate analysis in metabolomics studies is usually to determine which variables (e.g. metabolites) are important in the observed effect, e.g. those which show differential regulation between a chemical treatment and control. Many approaches are based on assessing the weight of each variable in the developed model; additionally some methods can assess the statistical significance of these weights. For example, bootstrap procedures may be used to estimate confidence intervals on PCA loadings, allowing selection of variables where confidence intervals do not contain zero. Methods such as Variable Importance in Projection (VIP) or S-plot are able to rank variables by importance in PLS models. They can be used as long as a statistically sound approach to determining the significance threshold is used (e.g. bootstrap resampling).

REPORT:

REQUIRED:

1) Name and description of method used for variable importance/selection
2) Details of how significance threshold determined
3) List of features selected by the method including
    a) Feature importance (e.g. VIP value)
    b) Feature significance (e.g. p-value, if available).
4) Methods should include clear citations to literature describing them.

5.2.8. Outputs

Because there are numerous approaches for the analysis and modeling of multivariate data, there are also numerous formats for outputting this information. It would be impossible to enumerate all of the potential output types, styles, and other information here. Instead, we list general guidance regarding outputs that may be submitted in support of a regulatory application.

REPORT:

REQUIRED:

2. Operator must submit a file manifest (either within a report, or as a separate file titled MANIFEST.xyz where .xyz is an appropriate file extension (.txt if a text file, .md if a markdown file, .html if an html file, .pdf if a pdf file)
    a. The manifest must include the listing of all files included in the regulatory application package specific to the DAM data.
    b. Each file in the manifest must be accompanied by a description of the file
        i. If the file being described is a tabular file, then the rows and columns must be described so that anyone can understand the contents of the file.
    c. The manifest file must be in a standard format that is easily readable in non-proprietary software.
    d. The manifest file must not be in a binary format, such as Word, Excel, or other formats that require special software that is not freely available, or where the format might change significantly and not be backwards compatible (e.g., Excel and Word files are binary formats and although freeware exists that can read these files, they may not be able to be read easily in the future).

6. References – TO DO