Systematic Review of NMR-Based Metabolomics Practices in Human Disease Research

Katherine Huang, Natalie Thomas, Paul R. Gooley and Christopher W. Armstrong

Abstract: Nuclear magnetic resonance (NMR) spectroscopy is one of the principal analytical techniques for metabolomics. It has the advantages of minimal sample preparation and high reproducibility, making it an ideal technique for generating large amounts of metabolomics data for biobanks and large-scale studies. Metabolomics is a popular “omics” technology and has established itself as a comprehensive exploratory biomarker tool; however, it has yet to reach its collaborative potential in data collation due to the lack of standardisation of the metabolomics workflow seen across small-scale studies. This systematic review compiles the different NMR metabolomics methods used for serum, plasma, and urine studies, from sample collection to data analysis, that were most popularly employed over a two-year period in 2019 and 2020. It also outlines how these methods influence the raw data and the downstream interpretations, and the importance of reporting for reproducibility and result validation. This review can act as a valuable summary of NMR metabolomic workflows that are actively used in human biofluid research and will help guide the workflow choice for future research.

Keywords: NMR spectroscopy; metabolomics; workflow; variation; standardisation

1. Introduction

Metabolomics aims to identify and measure metabolite snapshots of biospecimens that are representative of the biological condition of a subject, inclusive of internal and external factors. The dynamic metabolome is sensitive to perturbation of the subject’s state, and therefore requires the processes of metabolomics research to be well considered, precise, and rapid. Furthermore, for multiple snapshots within or between studies to be compared, the metabolomics process requires reproducible analytical instruments, standardised collection and acquisition procedures, robust statistical workflows, and minimal manual sample handling (automation).

The two main analytical instruments used for metabolomics are mass spectrometry (MS), coupled to gas or liquid chromatography, and nuclear magnetic resonance (NMR) spectroscopy. The advantages and disadvantages for each technique have been detailed previously [1]. MS is more widespread and commonly used, likely due to the improved sensitivity that enables more metabolites to be measured using both targeted and untargeted approaches. However, NMR offers key advantages in reproducibility for data acquisition, requires less sample preparation steps and allows for absolute quantitation that makes it favourable for data comparison across studies. It is also faster, and the methods are more economical than MS-based metabolomics experiments, making it ideal for large sample sizes.

Metabolomics has become an attractive method for exploring human health, especially disease states, however, challenges remain in the direct comparison and replication of results between studies. Previous reviews have reported best practices for NMR metabolomics [2–5], and emphasized the importance of standardised reporting among metabolomic studies [6–8]. However, it is unknown how common these practices are.
employed in research settings and whether the reporting guidelines have been followed. Real-world evidence of the variation and popularity of workflows being used would provide useful guidance for researchers to decide on the right NMR-based workflow to apply to their disease of interest.

To this end, we systematically summarised NMR-based metabolomics workflows of all studies reported on human disease patients in 2019 and 2020. In this work we summarised 131 articles with the aim to reveal common NMR metabolomics practices employed in the literature, the variations at each workflow step and their impacts, comparability of data across studies, and adherence to reporting standards.

2. Materials and Methods

2.1. Protocol and Registration

For this research we could not preregister the study in PROSPERO as it did not measure any health-related outcomes. The concept of this research was to identify variations and inconsistencies in the steps of NMR metabolomics workflow. We first critically analysed the first ten eligible articles in our literature search and extracted topics regarding different steps in the NMR metabolomics workflow that were to be examined. We then iteratively read and entered the corresponding information into columns for the rest of the articles. From our knowledge no such review exists that relates NMR-based metabolomics pitfalls supported by evidence from actual studies.

2.2. Eligibility Criteria, Information Sources and Search Parameters

The search terms were “nuclear magnetic resonance”, “metabolomics” and “patients” in the title/abstract of papers published in 2019 and 2020 and included all studies that analysed human biofluids with NMR metabolomics. The search was conducted on the 8 August 2021 in a single database: PUBMED database.

2.3. Study Selection

We exported the search results into EndNote. Abstracts were assessed independently by KH and NT to remove articles that were non-human, reviews, method papers or involved in vivo NMR experiments. Full texts of publications were checked by KH to confirm eligibility criteria. After the identification of eligible articles, if articles investigated multiple biofluids, they were further separated into their own study. In this research an “article” refers to the entire publication and “studies” refers to the individual biofluid workflows. The final literature search was filtered for biofluids that were serum, plasma, or urine.

2.4. Data Collection Process and Data Items

For each included study KH extracted the data in Excel. If an article referenced a protocol paper, the details reported in the protocol paper was compiled and referenced. For all included studies CA checked if the data extraction was performed correctly. Additionally, the following variables were also extracted but were not included into the results synthesis: (1) demographics of cohorts, (2) storage conditions for samples, (3) univariate analyses performed for clinical data, (4) pathways analysis, (5) software used for data analysis.

2.5. Synthesis of Results

No meta-analyses were predefined at the study conception stage, as the number of topics analysed, and the factors explored as determinants of articles addressing a topic could only be assessed after conducting the systematic review. The data extracted from this systematic review were all qualitative; bivariate (presence/absence) or non-ordinal categorical variables. Analysis began with generating frequency tables for each individual topic to identify initial trends. Topics were then analysed based on the flow of a typical NMR metabolomics workflow. Finally, topics were analysed holistically to visualise the different points of diversions that can occur across the NMR workflow. All data visualisations,
frequency tables and percentages were constructed and calculated using the “R” statistical environment [9].

2.6. Risk of Bias

We are not aware of any tools that measure the risk of bias for this type of systematic review. All eligible studies gathered under the inclusion criteria were used in the result synthesis. Extracted information were validated by one or more researcher.

3. Results

3.1. Study Selection and Data Collected

We identified 172 articles published over a two-year period (2019 and 2020) that used NMR metabolomics to analyse biospecimens from human patients [10–180]. The screening process is shown in Figure 1. The 172 articles were divided into separate studies for each biospecimen type. Twenty-eight articles had investigated two or more biospecimens, producing 208 studies after breaking down biospecimens into their own individual workflow. These were then filtered for biofluids that were either serum, plasma, or urine. The final literature sample consisted of 131 articles with 147 studies. The studies of other biospecimens (n = 61) were excluded due to inadequate data for comparison.

Figure 1. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram. After screening 204 articles, we synthesised 147 studies analysing NMR metabolomics workflows. N = number of articles, n = number of individual study arms.

The core steps in an NMR metabolomics workflow include the pre-analytical phase, data generation, data analysis and biological interpretation (Figure 2). Briefly, the pre-analytical phase includes sample collection and sample preparation, which may include metabolite extraction or removal of macromolecules. Once the sample is prepared data generation is conducted by data acquisition, with the appropriate experimental parameters,
spectral processing and then data can be presented as discrete bins that represent the area under the spectrum or individual metabolite concentrations. Data analysis includes data pre-treatment, analysis, and the identification of significant metabolites. Finally, biological interpretation provides biological context to the significant metabolites.

Figure 2. A typical workflow for untargeted NMR metabolomics. The NMR metabolomics workflow is divided into four main phases: (a) pre-analytical including sample collection and sample preparation, (b) data generation which involves NMR spectra acquisition, spectral processing, and the generation of spectral bins and/or metabolite concentrations. (c) Data analysis is performed for both data types including pre-treatment, multivariate and univariate analysis. The dashed arrow shows the optional integration of multivariate and univariate analyses. The fourth phase is (d) biological interpretation.

For the purposes of this review, we focused on all the main factors that were involved in the pre-analytical phase: (1) biofluid type, (2) collection method, (3) sample preparation
In the data generation phase, we considered (5) the pulse sequence(s) applied for data acquisition, (6) references for data acquisition protocols (7) whether the study generated binned data, metabolite concentrations or both, (8) uniform or variable width binning, (9) method used for metabolite profiling, as the chosen method impacts the data produced and data sharing capabilities. In the data analysis phase, we extracted (10) pre-treatment strategies, (11) tests for normality, (12) unsupervised multivariate analyses, (13) supervised multivariate analyses, (14) univariate analyses, (15) multiple testing correction, as it reflects the way the data are packaged for consumption through publication. The compiled studies used a diverse range of procedures in their metabolomics workflow summarised in Figure 3. We discuss the nature of this diversity in the following sections.

Figure 3. Sunburst charts showing the variation in workflows for NMR metabolomics. Each ring shows a step in the NMR metabolomics workflow and the segments show the percentage of studies that had used each method represented in (a) the whole literature search, (b) serum studies, (c) plasma studies and (d) urine studies only. The flavours of sunbursts can be explored interactively at https://rpubs.com/sunburstNMRMetabolomics accessed on 25 May 2022. One plasma study (Article ID: 99) that used sodium-citrate collection tube was removed in the visualization to keep the segments at a perceivable size (see Supplementary Data for the detailed workflow).
3.2. Pre-Analytical Phase

There were 25 different biospecimens identified in our systematic review, including serum, urine, tissue, cerebral spinal fluid, follicular fluid, stool, exhaled breath condensate, saliva, aqueous humour, bile, platelets, seminal fluid, sputum, synovial fluid, amniotic fluid, acute promyelocytic blasts, interstitial fluid, mouth wash out, pancreatic juice, peritoneal effusion, red blood cells, tears, tongue swab, and whole blood. The most popular were serum (32.2%), urine (22.1%), and plasma (16.3%). These biofluids are generally minimally invasive and relatively simple to collect.

3.3. Blood Collection

Whole blood from patients is often altered upon collection by the tubes and processing steps that come after. Serum and plasma are the most commonly analysed blood extracts but analysis can be extended to whole blood [181], platelets [103], red blood cells [56], peripheral blood mononuclear cells [182] and dried blood spots [183].

Plasma is obtained after the removal of cells and platelets via centrifugation. It is the liquid proportion of blood that contains clotting factors and protein. Plasma is collected in tubes containing an anticoagulant. Overall, 41.2% of plasma studies used ethylenediaminetetraacetic acid (EDTA) tubes, 14.7% used heparin followed by 2.9% using sodium citrate and 41.2% not reporting the nature of the collection tube. The anticoagulant should be chosen carefully as it may inhibit other biological processes and influence the metabolic profile. Heparin inhibits coagulation activator (thrombin) whereas EDTA and citrate chelate divalent metal ions, such as calcium and magnesium, thus inhibiting magnesium-dependent coagulation enzymes [183]. EDTA produces intense peaks in NMR spectra, obscuring signals from choline, dimethylamine, and citrate; and samples may contain endogenous citrate. Therefore, it is recommended to avoid EDTA and sodium citrate anti-coagulants [184]. Conversely, it has also been suggested to avoid heparin as it causes broad peaks in the spectrum complicating lipid quantitation [185]. Furthermore, 38.2% plasma studies reported centrifugation parameters (rotor speed, time, temperature) for processing at collection, 41.2% prior to sample preparation and 14.7% reported at both steps. For each of the biofluids (including urine), there were a diverse range of rotor speeds, the centrifugation time ranged from 5 to 20 min and temperature reported included 4 °C or at room temperature. Not all studies reported all three parameters; a detailed breakdown can be found in the Supplementary Data.

An alternative to plasma that does not require addition of anticoagulants that interfere with resonances of the NMR spectra is serum. Serum consists of similar constituents as plasma with the absence of clotting factors. It is prepared by letting whole blood sit at room temperature for clot formation, usually for 30 min. Only 31.3% of serum studies reported a clot time, which ranged from 15 to 240 min. During clotting time, enzymatic reactions and degradation can still occur. For example, the activation of platelets can release additional compounds such as hypoxanthine, xanthine, and amino acids [186]. Overall, 19.4% of serum studies used tubes containing no additives, 7.5% a gel separator, 4.5% silica-coated tubes and 68.7% did not report the type of tube. Centrifugation parameters were reported for 50.7% serum studies at collection, 23.9% before sample preparation and 7.5% reported at both steps.

3.4. Urine Collection

Urine samples can be collected under different conditions. We identified four different urine types; first morning void (41.3%), random urine (6.52%), 24 h urine (4.35%), spot urine (2.17%), and 45.7% did not report the condition of urine collection. First morning void occurs following an overnight fast and is least likely to be affected by daily routine. Random urine can be collected at any time of the day and may induce the most unwanted variability from different collection times and conditions. We considered a sample to be random urine if it was collected after appointments or treatments occurring at unspecified times during the day. In contrast, spot urine is collected at specified times. We have defined
second morning void as spot urine, as it is collected under pre-defined conditions without fasting. The 24 h urine is a pooled sample of all voids within a 24 h period [187]. This sample type can average out fluctuations from the circadian cycle.

Another consideration for urine collection is bacterial contamination. Overall, 21.7% of urine studies reported midstream collection which reduces the risk of collecting a sample contaminated from the urinary tract [188]. Further, removing cells and bacteria from the sample is a standard practice that is achieved with mild centrifugation and/or filtration with a 0.20 \( \mu \text{m} \) syringe filter [185]. In all, 63.0% urine studies reported performing centrifugation and/or filtration, with the remainder unknown. Meanwhile, 33% studies reported centrifugation parameters at collection, 43.5% before sample preparation for the thawed sample and 15.2% reported for both steps.

3.5. Sample Preparation

There are additional sample preparation steps that may be included in the workflow to isolate metabolites from macromolecules that remain in the samples after treatment. These steps include ultrafiltration or metabolite extraction. We found that 82.2% of blood (serum and plasma) studies did not perform additional sample preparation steps, 9.5% performed ultrafiltration and 7.9% performed metabolite extraction. Overall, 4.3% of urine studies performed ultrafiltration, and the remainder conducted no preparation.

Part of sample preparation is converting raw biofluid into an optimal medium that is suitable for the analytical instrument. Deuterated phosphate buffer is often added to the sample prior to NMR data acquisition. Adding buffer maintains a constant pH (physiological pH 7.4) across all the samples which minimises metabolite chemical shift variations and allows for more accurate metabolite identification against library standards. Sodium or potassium phosphate buffer was used in 79.6% studies, 6.1% used deuterated water (D\(_2\)O) only, 2.7% used saline solution with two deuteration levels (100% or 10%) and 0.7% (one urine study) used Chenomx Internal Standard Solution consisting of D\(_2\)O, sodium trimethylsilylpropanesulfonate (DSS) and sodium azide (Table 1). Sodium azide (NaN\(_3\)) can also be added in the buffer to prevent bacterial growth [189], which 30.6% studies had reported. Overall, 68.7% studies reported a chemical shift reference, with trimethylsilylpropanoic acid (TSP) (54.5%) and DSS (10.2%) being the most popular.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Collection Tube</th>
<th>Sample Prep</th>
<th>Sample: Buffer Ratio</th>
<th>Buffer</th>
<th>% D(_2)O</th>
<th>Chemical Shift Reference (mM)</th>
<th>NaN(_3) (mM)</th>
<th>pH</th>
<th>NMR Experiments</th>
<th>Temp (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beckonert [184]</td>
<td>Li-heparin</td>
<td>1:2</td>
<td>103 mM NaCl</td>
<td>6.66</td>
<td></td>
<td></td>
<td></td>
<td>noesy, cpmg, (jres, diff)</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>Bernini [185]</td>
<td>EDTA or citrate</td>
<td>1:1</td>
<td>35 mM Na(_2)HPO(_4)</td>
<td>10</td>
<td>TSP (27.5)</td>
<td>19</td>
<td>7.4</td>
<td>noesy, cpmg</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>Dona [3]</td>
<td>Li-heparin or EDTA</td>
<td>1:1</td>
<td>37.5 mM Na(_2)HPO(_4)</td>
<td>10</td>
<td>TSP (2.73)</td>
<td>3.08</td>
<td>7.4</td>
<td>noesy, cpmg, jres</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>Soininen [190]</td>
<td>Ultrafiltration 3 kDa</td>
<td>1:1</td>
<td>37.5 mM Na(_2)HPO(_4)</td>
<td>10</td>
<td>TSP-d(_4) (2.32)</td>
<td>3.08</td>
<td>7.4</td>
<td>noesy, cpmg</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>Chenomx</td>
<td></td>
<td>9:1</td>
<td>D(_2)O</td>
<td>10</td>
<td>DSS-d(_6) (0.5)</td>
<td>1.54</td>
<td>noesy</td>
<td>298</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Collection Tube</th>
<th>Sample Prep</th>
<th>Sample Buffer Ratio</th>
<th>Buffer</th>
<th>% D2O</th>
<th>Chemical Shift Reference (mM)</th>
<th>NaN3 (mM)</th>
<th>pH</th>
<th>NMR Experiments</th>
<th>Temp (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beckonert</td>
<td>0.05% wt/vol NaN3</td>
<td>2:1</td>
<td>82.3 mM Na2HPO4</td>
<td>6.66</td>
<td>TSP (0.33)</td>
<td>1</td>
<td>7.4</td>
<td></td>
<td>noesy</td>
<td>300</td>
</tr>
<tr>
<td>Bernini</td>
<td>3 mM NaN3</td>
<td>9:1</td>
<td>150 mM K2HPO4</td>
<td>10</td>
<td>TSP (1.0)</td>
<td>7.4</td>
<td></td>
<td>noesy, jres</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Dona [3]</td>
<td>0.05% wt/vol NaN3</td>
<td>9:1</td>
<td>150 mM KH2PO4</td>
<td>1</td>
<td>TSP (0.68)</td>
<td>0.2</td>
<td>7.4</td>
<td>noesy, jres</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Chenomx</td>
<td></td>
<td>9:1</td>
<td>D2O</td>
<td>10</td>
<td>DSS-d6 (0.5)</td>
<td>1.54</td>
<td></td>
<td>noesy</td>
<td>298</td>
<td></td>
</tr>
</tbody>
</table>

Serum follows the same parameters as plasma minus the collection tube. NMR experiments listed in brackets were optional recommendations. TSP-d4: 3-(Trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt. DSS-d6: 3-(Trimethylsilyl)-1-propanesulfonic acid-d6 sodium salt.

3.6. Data Generation Phase

3.6.1. NMR Introduction

NMR is a powerful analytical technique known for its ability to characterise molecular structures and dynamics. The majority of NMR applications take advantage of NMR-active nuclei from isotopes such as $^1$H, $^{13}$C, $^{15}$N, $^{31}$P. $^1$H has 99% natural abundance and is present in most metabolites, including amino acids, sugars and fatty acids making it particularly useful and popular for the identification of known metabolites. A single 1D $^1$H NMR spectrum can capture hundreds to thousands of signals from molecules that may be low or high in molecular weight [191]. There are different NMR experiments with various pulse sequences, which are series of microsecond radio frequency pulses and magnetic gradients that can be manipulated to excite the active nuclei to produce characteristic NMR spectra.

3.6.2. NMR Experiments

The non-destructive nature of NMR allows multiple NMR experiments to be applied to a single sample. Overall, 69.4% of studies performed a single experiment, 29.4% performed multiple experiments and 0.7% did not report the type of experiment. We identified two main experiments which were the 1D Carr-Purcell-Meibom-Gill (CPMG) experiment with presaturation for solvent suppression and $T_2$ relaxation filtering and the 1D nuclear Overhauser enhancement spectroscopy (NOESY) experiment also with presaturation [184]. Overall, 48.3% of studies performed a CPMG experiment, 46.3% performed a NOESY experiment, 9.5% studies sent their samples to the Nightingale Health metabolomics platform for data acquisition and generation [192], 8.8% performed a 2D J-resolved experiment [193], 8.2% performed a diffusion-edited experiment [194] and 8.8% applied other NMR experiments (Supplementary Table S1).

Different NMR pulse sequences can exploit the behaviour of nuclei to produce a characteristic spectrum. The 1D presaturation NOESY pulse sequence suppresses the water signal, without sacrificing the signal intensity of the majority of metabolite peaks, capturing high and low molecular weight compounds [195]. The CPMG experiment attenuates signals that have short transverse relaxation times, such as large proteins and lipoproteins, leaving only slowly relaxing small metabolites and those signals from slow relaxing protons, such as methyl groups of lipid signals in the spectrum. Since serum, plasma and urine consist of different molecular constituents and NMR experiments can filter or select for molecules, their relationship is shown in Figure 4.
Figure 4. Bar graph showing how data are acquired on NMR. Different NMR experiments and the Nightingale Health platform used to generate NMR data from different sample preparation steps in serum, plasma, and urine studies.

Briefly, Nightingale Health has created a high-throughput and automated NMR metabolomics platform for serum and plasma samples; able to detect lipids, lipoprotein particles and subclasses (LIPO), and low molecular weight metabolites (LMWM) [190,192]. It uses robotics to prepare the blood samples for the LIPO and LMWM spectra, applying the $^1$H NOESY and CPMG pulse sequences, respectively. A manual lipid extraction is performed, and a third lipid spectrum is acquired using $^1$H NOESY pulse sequence. All the spectra are automatically processed (including phase correction, baseline correction, spectrum alignment) and metabolites are automatically identified and quantified using in-house software based on Bayesian modelling [196]. The 2D $^1$H NMR J-resolved experiment separates the scalar coupling and chemical shifts of resonant peaks into two dimensions which spread out overlapping signals, aiding in the identification of metabolites [193]. Diffusion-edited experiments can produce a spectrum only containing metabolites by subtracting the macromolecule spectrum from the whole spectrum based on the diffusion coefficient of the nuclei [194,197].

There were a few research groups that were frequently referenced for sample preparation and data acquisition: Beckonert et al. (2007) [184], Bernini et al. (2011) [185] Dona et al. (2014) [3], and Soininen et al. (2015) [192] which describe the sample preparation and data acquisition parameters used in the Nightingale Health NMR metabolomics platform. Their parameters are described in Table 1.

3.6.3. Spectral Binning

NMR spectra of biological mixtures yield information-rich data that can be quantified by spectral binning or metabolite profiling where for the latter concentrations are determined. We found 39.5% of studies generated binned data only, 35.4% analysed metabolite
concentrations only, and 25.2% investigated both. There was a higher proportion of urine studies (80.4%) compared to blood that generated binning data, with 48.0% also generating metabolite concentrations. More blood studies generated metabolite concentrations only (42.6%), while 32.7% employed binning only and 24.8% for both.

Spectral binning offers a rapid and consistent method in identifying global trends in spectral peak patterns without the initial need for metabolite identification [198]. The simplest method is uniform binning. It divides the spectrum into equal widths (for example 0.01, 0.04 ppm). Software from Bruker including AMIX and AssureNMR, Mnova, Chenomx, MATLAB, ACD/Labs, NMRProcFlow [199], dataChord Spectrum Miner [200], KnowItAll Software (John Wiley & Sons, Inc., Hoboken, NJ, USA) were used to generate uniform binning data (Supplementary Table S1). Due to the small spectral widths (10–12 ppm) of $^1$H NMR spectra, metabolite peaks can overlap and summate therefore a bin may not always contain a single peak corresponding to a single $^1$H moiety of a metabolite. Further, peak-shift problems and the presence of noise can influence downstream analyses. This has prompted the development of various intelligent, adaptive, and dynamic binning algorithms to account for intensity variation and multiple metabolite peaks [201–204]. The basic concept behind these algorithms is to isolate every peak by determining local maxima and minima, form bin boundaries with variable widths and subsequently remove noise. Out of the studies that generated binned data, 16.3% used variable width bins generated from algorithms or manual integration.

Bins that show significant differences between groups are identified in the downstream data analysis. Peaks of the metabolites within these bins are assigned based on their chemical shift by referencing to various databases such as the Human Metabolome Database (HMDB), Biological Magnetic Resonance Bank (BMRB), BBIOREFCODE (Bruker) and assignments reported in the literature.

### 3.6.4. Metabolite Profiling

Metabolite profiling involves fitting the sample spectrum to pure known metabolite spectra. Through this process, metabolites are both identified and quantitated. Fully automated metabolite profiling remains a challenge in the NMR-based metabolomics workflow due to overlapping metabolite signals and slight chemical shift differences from pH, ionic strength, temperature, and biological matrix discrepancies. Many studies resort to manual or semi-automatic deconvolution methods and commercial software which can be slow, subjective, and error prone.

Our findings suggest that commercial metabolite profiling tools or platforms were preferred over open source. The most frequently used tools or platforms included Chenomx (42.7%), Nightingale Health (15.7%) and Bruker (AMIX, IVDr, PERCH Solutions) (14.4%). To maximise accuracy of automatic fitting and quantitation, samples should be acquired with the same experimental parameters as used for the reference library. Chenomx provides standard operating protocols (SOPs) (Table 1) for sample preparation and their own NMR acquisition parameters. Despite their popularity as a metabolite profiling tool, only one profiling study used the Chenomx SOP (1.1%). Open source web-servers including MAGMET [205], MetaboHunter [206] and Metabominer [207] were each employed once (1.1%) by profiling studies. Various deconvolution algorithms exist [208], however only BATMAN [209] appeared (also once, 1.1%).

### 3.7. Data Analysis Phase
#### 3.7.1. Data Pre-Treatment

Metabolomic data may be subjected to confounding biological and experimental variations. Therefore, it is necessary to perform pre-treatment such as normalisation, scaling, and transformation to produce a “clean” dataset that make samples (participants) and variables (bin integrals or metabolite concentrations) more comparable and suitable for specific analyses. The application of various pre-treatment methods emphasize different aspects of the data and can profoundly affect biological interpretation [210].
We found two main normalisation techniques: 32.0% used total sum; 12.2% used probabilistic quotient normalisation (PQN); and 41.5% did not report a normalization technique. Total sum represents each variable relative to all the number of variables [211]. PQN divides each sample spectrum by a reference spectrum that is representative of the median [212]. The advantage of PQN is that regions of the spectrum are normalised against itself, therefore areas of interest are not influenced by the rest of the spectrum. Both total sum and PQN are global normalisation approaches [213]. Other approaches include referencing the spectra to a single entity that can either be the chemical shift reference or an endogenous metabolite such as creatinine and formate (Supplementary Table S1).

Unit variance scaling, also known as autoscaling was performed by 30.6% of the studies, 19.3% performed pareto scaling, 3.4% performed mean centring only and 46.7% did not report. Unit variance scaling and pareto scaling initially involves mean centring the data. All the metabolite concentrations therefore fluctuate around zero, accounting for any bias that may favour abundant metabolites. Autoscaling subsequently divides each variable by their standard deviation so that all metabolites are treated with equal importance in downstream analyses and pareto scaling divides the mean-centred data by the square root of the standard deviation to simulate the relative abundance of metabolites from the original dataset in the scaled dataset [210].

3.7.2. Multivariate Analyses

Multivariate analysis was performed by 87.1% of the studies. The majority began with unsupervised principal components analysis (PCA), followed by a supervised classification model and the identification of key metabolites through variable importance scores. PCA was a common analysis and was applied by 68.0% of the studies that performed multivariate analysis. PCA is a dimensionality reduction technique which reconstructs high-dimensional data into linear combinations called principal components that preserve the maximum variation observed in the original dataset [214]. The variation can separate the data into different clusters. Being an unsupervised analysis, PCA does this without the knowledge of class labels, which is the attribute of interest that defines a group; in metabolomics studies class labels are often the presence or absence of a disease.

Projection to Latent Structures Discriminant Analysis (PLS-DA) is a classification machine learning algorithm [215]. Like PCA, it also performs dimensionality reduction by generating linear combinations, however with the knowledge of class labels. Overall, 42.2% of multivariate studies performed PLS-DA, which has two additional variations: orthogonal PLS-DA (OPLS-DA) and sparse PLS-DA (sPLS-DA). The best performing PLS-DA variant is debated in the literature [216]. Both variations are aimed to identify important variables and remove those that do not contribute to the prediction of the class label. OPLS-DA was performed by 48.4% of multivariate studies and 0.5% performed sPLS-DA. The weakness of PLS-DA is that it is prone to overfitting, which means they rely heavily on the initial training dataset for accurate prediction.

Other unsupervised and supervised machine learning algorithms that appeared in our literature search included random forest (11.7%), hierarchical clustering analysis (5.5%), support vector machine (4.7%), k-means clustering (2.3%), and t-SNE (0.8%). Although outside the scope of this review, many reviews exist that explore the broad array of statistical objectives, misconceptions, and pitfalls implemented [217,218]. At this stage, the studies from our systematic review showed that machine learning models are still at the training stage and are yet to become inferential.

3.7.3. Univariate Analyses

Univariate analysis was performed by 85.7% of studies. It uses one variable to describe or infer a conclusion on a sample population using hypothesis testing or univariate generalised linear modelling for testing the strength of association between the metabolite and class label. Here, we will mainly discuss how hypothesis testing is used in metabolomic studies. Choosing the correct hypothesis test for the data can yield more reliable results.
which can be achieved by meeting all assumptions the test relies upon, e.g., normality, independence, and equal variance [219,220].

A normal (gaussian) distribution is when observations appear most frequently around the mean value, which can be visualized as a symmetrical bell-shaped plot. To assess for normality, tests reported in our studies included Kolmogorov-Smirnoff, Shapiro–Wilk, D’Agostino-Pearson, and histograms and Q-Q plots for visual inspection. Overall, 90.5% of studies that performed hypothesis testing did not perform a normality test prior.

Simultaneously performing multiple hypothesis tests for the large number of metabolite features increases the possibility of false positives (Type I error). Significance levels and \( p \)-values need to be adjusted which is referred as multiple testing correction [221]. Studies reported controlling the false discovery rate inclusive of Benjamini–Hochberg and Benjamini–Yekutieli procedures, and the family-wise error rate including the Bonferroni correction (Supplementary Table S1). Out of the studies that performed hypothesis testing, 36.5% also adjusted for multiple testing. Some may argue that multiple testing correction is not necessary as it may be more detrimental to the research if potential significant metabolites were missed. Considering this, the multiple testing corrections have different thresholds which can be applied based on the exploratory nature of the research question.

4. Discussion

The main aim of this systematic review of NMR-based metabolomics research on human biofluids was to determine how much variation in workflow existed between different studies. Although various tissue, cells and compartmentalised fluids may provide more localised significance and relevance to metabolic perturbations observed in a disease, serum, plasma, and urine accounted for more than 70% of the studies in our literature search. The variation of workflows on these three biofluids (visualised in Figure 3) is quite substantial, especially for serum and plasma.

Serum and plasma are extracts of whole blood. Previous studies found that the quantitation of metabolites in plasma was more reproducible due to reduced handling [222], however higher concentrations of amino acids were found in serum which may be explained by volume displacement effects [223]. A recent study investigating the impact of collection tubes between serum and plasma revealed that heparin plasma, followed by EDTA plasma had a closer metabolic profile to serum collected in tubes with no-additives [224]. Further, serum separator tubes containing polymeric gel has been shown to alter metabolite levels therefore it is recommended to use additive-free glass or plastic tubes [184]. Another pre-analytical factor that needs to be considered are the centrifugation parameters. The force applied may affect platelet count [225] or cause hemolysis which alter metabolite levels in blood [2]. Various studies have investigated the influence of different centrifugation parameters have on serum, plasma and urine metabolomes and developed optimal protocols for their respective processing [226,227].

The goal of sample collection and sample preparation steps is to preserve the metabolic composition of the sample to ensure an accurate representation of the metabolome at the time of collection. Therefore, SOPs specific for metabolomics studies are crucial as slight variations in sample collection, processing and storage conditions can significantly affect metabolite stability and abundance [183]. However, the issue of having numerous best practices remains and it is up to the researcher to be well-versed and transparent in their decisions. To achieve this, complete reporting of collection method including collection tubes and durations and processing parameters are mandatory for reproducibility and to reduce analytical bias.

Sample preparation steps are often used to enhance the metabolite peak signals of the NMR spectra. Macromolecules, such as protein and lipids, give rise to intense and broad signals in NMR spectra, obscuring metabolite signals, and the chemical shift reference may bind to protein causing difficulties and errors in metabolite quantification. Ultrafiltration (with 3 kDa or 3.5 kDa cut-off) is the simplest and fastest method for removing macromolecules. However, metabolites can be lost in the filter membrane and protein-bound
metabolites are filtered out along with the protein [228]. Liquid-liquid extraction (LLE) is a metabolite extraction method that uses organic solvents to precipitate protein and separate polar metabolites and lipids into hydrophilic and hydrophobic phases, respectively. Using deuterated methanol and chloroform has been shown to prevent further enzymatic activity [229] and yield efficient recovery of metabolites [230]. LLE appears to enhance metabolite peaks compared to ultrafiltration [230], most likely due to the capture of protein-bound metabolites after precipitation. Solvent peaks that are introduced into the NMR sample may be removed via lyophilisation, however this process will remove some of the volatile metabolites. The chemical shift reference is used for spectra alignment and absolute metabolite quantitation. Without a chemical shift reference, spectra can be aligned using isolated metabolite peaks that are unlikely to be affected by pH, however, with this approach metabolites cannot be absolutely quantitated.

The NOESY pulse sequence can consistently generate high-quality spectra with short acquisition time, and it is relatively simple to set up with few optimisation parameters, suitable for the non-spectroscopist. The disadvantage of the NOESY pulse sequence is that the solvent suppression can leave the baseline distorted, or alter signal intensities near the suppressed solvent peak [231]. Distorted baselines will cause difficulties for automatic processing and may introduce inaccurate or subjective corrections. When a sample contains high molecular weight compounds such as protein and lipoproteins, they create broad peaks in the NMR spectrum which may dominate metabolites signals, therefore a CPMG pulse sequence is more ideal for quantitating metabolites in biospecimens with high protein content. Another consideration of using the CPMG pulse sequence is that protein-bounded metabolites will not be observed in the NMR spectrum as they share an effective relaxation time with the protein and are filtered out together with the protein signals. Furthermore, lipids in the sample can obscure upfield metabolite signals. As part of the MSI guidelines [6], all instrument parameters should be fully reported so that experimental procedures can be replicated; all studies except for one followed this requirement either briefly describing the parameters or referencing the protocol followed.

For the sake of comparing NMR data between separate studies it is most important that sample collection, sample preparation and data generation techniques are as consistent as possible. Urine was shown to have a very consistent workflow of neat collection, no preparation, and NOESY pulse sequence being applied in 71.7% of studies. Serum and plasma were both inconsistent and this was confounded by a lack of reported information on collection tubes used to produce the serum or plasma. Removing the issue of the collection tube from the equation, the most common combination for serum and plasma was no preparation and CPMG pulse sequence (56.4% of blood studies). The Nightingale Health NMR metabolomics platform provides a consistent combination of collection factors; however, the limitation of this platform is that spectra are not provided, so binning analysis cannot be performed. Minimal handling of samples is a sensible way to reduce human error and manipulation of in vivo measurements of blood. Such minimal handling may explain why no sample preparation is the most common method for NMR-based metabolomics.

The generation of metabolite data can be considered one side of the workflow, the other side is the analysis and interpretation of that data. The data analysis workflow and the presentation of data in research papers was highly varied between projects. There are many ways to look at data, but the standardisation of the steps appears to be limited. Beyond variation, there appeared to be minimal justification for the data analysis steps taken by researchers presenting metabolomics data. Prior to the analysis of data, the data needs to be converted from the NMR spectral format to a numerical format. This conversion of data can be performed by profiling metabolite concentrations or by producing bins that quantitate the area under the spectrum.

Datasets continue to grow and become more complex therefore requiring proper pre-treatment. Specific scaling strategies should be applied for different statistical analyses as the aim of the scaling method should match what the analysis is measuring. There was a large proportion of studies that did not report a normalisation or scaling strategy, we
strongly encourage researchers to provide a detailed reasoning for their choice, or lack of pre-treatment.

Although hypothesis testing procedures are routine tests seen in metabolomics studies, we found that they were misused. Both parametric hypothesis tests, which are tests that meet the conditions of a normal distribution, and non-parametric tests, which are tests for non-normally distributed data, were often conducted without first assessing normality [232]. Parametric tests are more powerful than non-parametric, therefore significant metabolites may be inadvertently deemed important or missed if either parametric or non-parametric tests are applied with incorrect normality assumptions [233]. Once an initial assessment for normality is conducted, non-normal data may be scaled or transformed so that a parametric test can be performed. An alternative is algorithmic modelling or machine learning which uses functions to identify patterns within the data and is measured based on prediction accuracy [234]. It is important to recognise that there are several models that may be applicable to the data [234], therefore it is crucial to describe the biological interpretation in context with the analyses performed, whether it is univariate, multivariate or algorithmic-based. The drawback for machine learning algorithms is that large sample sizes are required for accurate prediction, whereas most sample sizes for metabolomics studies are still relatively small. We believe that the standardisation in data generation for data collation will pave the way for the next advancements of metabolomics and machine learning applications.

While this systematic review aims to provide an update on current methods commonly employed in the NMR metabolomics workflow and point out the extent of their variation between studies, we still expect to see variation in the future as improvements are made to existing methods or new methodologies are developed. The necessity for standardisation is often debated as there are continued advancements made in the NMR metabolomics field such as sample collection devices [235,236], data processing software [237], NMR experiments and pulse sequences [238], improved instrument sensitivity [239,240], and metabolite deconvolution software [241–243]. Therefore, when using new methods, we encourage careful considerations in their implementation and detailed reporting.

5. Conclusions

Our analysis of metabolite disease studies published in 2019 and 2020 has highlighted that significant variation exists in NMR-based metabolomics data generation and data analysis workflows. The variation in data analysis workflows is expected but more justification of steps taken should be reported. Given that NMR reproducibility is a real strength to using this platform for metabolomics, we recommend using data generation workflows that are consistent within the field while leaning towards minimal sample handling steps. For urine, there is a consistent workflow of neat collection, no preparation and using the NOESY pulse sequence to acquire NMR data (this data generation workflow is likely most suitable for any biofluid that lacks macromolecules). For serum and plasma, there are inconsistencies between studies, but the use of glass tubes without additives for serum and heparin for plasma appear to be the most common; this is followed by no preparation step and using both NOESY and CPMG pulse sequences for acquiring NMR data. Overall, this review can act as a valuable starting resource for any research group wanting to standardise their research and make it suitable for data collation and comparison.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/metabo12100963/s1, Table S1: Summary table for serum, plasma, and urine.

Author Contributions: K.H., N.T. and C.W.A. designed the review concept; K.H. performed the data curation, investigation (with N.T.), data analysis, visualisation and the original draft preparation; and C.W.A. and P.R.G. guided the project. All authors contributed to the review process and editing. All authors have read and agreed to the published version of the manuscript.
Funding: This salary of the researchers that conducted this work was supported by grants from Open Medicine Foundation and Mason Foundation.

Data Availability Statement: Compiled dataset in an excel spreadsheet available in Supplementary Information.

Acknowledgments: This work was supported by grants from The Mason Foundation and Open Medicine Foundation.

Conflicts of Interest: The authors declare no conflict of interest.

References


23. Castaldo, G.; Pagano, I.; Grimaldi, M.; Marino, C.; Molettieri, P.; Santoro, A.; Stillitano, I.; Romano, R.; Montoro, P.; D’Ursi, A.M.; et al. Screening for Early Gastric Cancer Using a Noninvasive Urine Metabolomic Signature. J. Proteome Res. 2020, 2, e0272. [CrossRef]


54. Lima, A.R.; Pinto, J.; Barros-Silva, D.; Jerónimo, C.; Henrique, R.; Bastos, M.L.; Carvalho, M.; Guedes Pinho, P. New findings on urinary prostate cancer metabolome through combined GC-MS and 1H NMR analytical platforms. Metabolomics 2020, 16, 70. [CrossRef]


57. Metere, A.; Graves, C.E.; Chirico, M.; Caramujo, M.J.; Pisanu, M.E.; Iorio, E. Metabolomic Reprogramming Detected by 1H-NMR Spectroscopy in Human Thyroid Cancer Tissues. Biology 2020, 9, 112. [CrossRef]


95. Righi, V.; Cavallini, N.; Valenti, A.; Pinna, G.; Pavese, G.; Rossi, M.C.; Puzzolante, A.;ucci, A.; Cocchi, M. A metabolomic data fusion approach to support glosma grading. *NMR Biomed.* 2020, 33, e4234. [CrossRef]


131. Molinero, N.; Ruiz, L.; Milani, C.; Gutiérrez-Díaz, I.; Sánchez, B.; Mangifesta, M.; Segura, J.; Cambero, I.; Campelo, A.B.; García-Bernardo, C.M.; et al. The human gallbladder microbiome is related to the physiological state and the biliary metabolic profile. *Microbiome* 2019, 7, 100. [CrossRef]


146. Faitot, F.; Ruhland, E.; Oncioiu, C.; Besch, C.; Addeo, P.; Ciccek, A.E.; Bachellier, P.; Namer, I.J. Metabolomic profiling highlights the metabolic bases of acute-on-chronic and post-hepatocytic liver failure. *HPB (Oxford)* 2019, 21, 1354–1361. [CrossRef]


156. Yanshole, V.V.; Yanshole, L.V.; Snytnikova, O.A.; Tsentalovich, Y.P. Quantitative metabolomic analysis of changes in the lens and aqueous humor under development of age-related nuclear cataract. *Metabolomics* 2019, 15, 29. [CrossRef]


