Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Cross-sectional Plasma Metabolomic Profiles across Acute, Recovery-Stage, and Recurrent Ischemic Stroke: A Quality-Control and Multiplicity-Aware Exploratory Study

**shiyang liu**
The Second Affiliated Hospital of Anhui University of Chinese Medicine
**Meixiang Sun**
The Second Affiliated Hospital of Anhui University of Chinese Medicine
**Hui Liu**
The Second Affiliated Hospital of Anhui University of Chinese Medicine
**Yu Wang**
The Second Affiliated Hospital of Anhui University of Chinese Medicine
**Fang Zhang**
The Second Affiliated Hospital of Anhui University of Chinese Medicine
**Nan Li**
The Second Affiliated Hospital of Anhui University of Chinese Medicine
**Tingting Qian**
The Second Affiliated Hospital of Anhui University of Chinese Medicine
**Ling Zou**
Yangzhou Hospital of TCM
**Peiyang Sun**
sunpeiyang2008@aliyun.com

The Second Affiliated Hospital of Anhui University of Chinese Medicine

**Research Article**

**Keywords:** ischemic stroke, recurrence, metabolomics, plasma, quality control, exploratory study

**Posted Date:** November 5th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11063881/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Cross-sectional Plasma Metabolomic Profiles across Acute, Recovery-Stage, and Recurrent Ischemic Stroke: A Quality-Control and Multiplicity-Aware Exploratory Study

Running title: Plasma metabolomics across stroke states

Shiyang Liu[1,2] Meixiang Sun[1] Hui Liu[1] Yu Wang[1] Fang Zhang[1] Nan Li[1] Tingting Qian[1] Ling Zou[3] Peiyang Sun[1*]

Corresponding author: Peiyang Sun (sunpeiyang2008@aliyun.com)

[1] Department of Neurology, The Second Affiliated Hospital of Anhui University of Traditional Chinese Medicine, Hefei, Anhui, P.R. China

[2] Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

[3] Rehabilitation Department,Yangzhou Hospital of TCM,Yangzhou,Jiangsu,China.

## Abstract

Background and Purpose: Plasma metabolomics may capture biological heterogeneity across clinically defined ischemic stroke states, but small discovery datasets require careful control of analytical variability, metabolite-annotation uncertainty, and multiple testing. We assessed whether a four-group cross-sectional plasma dataset contained reproducible global or feature-level differences after explicit quality-control filtering and multiplicity correction. Materials and Methods: Plasma samples from 40 participants were compared across healthy control (HC), acute ischemic stroke (AIS), recovery-stage stroke, and recurrent ischemic stroke (RIS) groups (n = 10 per group). Positive- and negative-ion untargeted liquid chromatography-mass spectrometry (LC-MS) feature tables were analyzed separately using the documented detection, transformation, normalization, pooled-quality-control (QC), and imputation workflow. Global structure was assessed by principal component analysis and one-factor PERMANOVA with 1,999 permutations. Feature-wise differences were evaluated by Kruskal-Wallis tests with Benjamini-Hochberg false-discovery-rate (FDR) correction within ion mode. Results: After QC filtering, 1,605 positive-mode and 872 negative-mode feature rows remained. PCA showed extensive overlap among groups. PERMANOVA did not support a global group effect in positive mode (pseudo-F = 0.988, $R^2$ = 0.076, P = 0.486) or negative mode (pseudo-F = 1.168, $R^2$ = 0.089, P = 0.117). Although 68 positive-mode and 77 negative-mode features had nominal omnibus P < 0.05, none survived FDR correction. Several QC-qualified, putatively annotated lipid-related signals showed descriptive cross-sectional patterns but remained analytically unconfirmed. Conclusions: No global or feature-level difference met multiplicity-controlled criteria in this pilot dataset. The nominal signals should therefore be regarded as candidates for targeted chemical confirmation and longitudinal validation rather than as stage-specific or recurrence-associated biomarkers.

Keywords: ischemic stroke; recurrence; metabolomics; plasma; quality control; exploratory study

## 1. Introduction

Recurrent ischemic stroke remains an important cause of cumulative neurological disability and mortality despite established secondary-prevention strategies. Recurrence risk is heterogeneous because it arises from multiple vascular mechanisms, including large-artery atherosclerosis, cardioembolism, small-vessel disease, and

other or undetermined causes, while also being modified by blood pressure, glycemic status, lipid burden, smoking, medication adherence, renal function, inflammation, and the timing of secondary-prevention treatment [1,2]. This heterogeneity creates an important translational problem: patients with similar conventional risk profiles may follow different biological courses after an index event. Molecular phenotyping may therefore be useful not as a replacement for established clinical assessment, but as a complementary way to interrogate residual biological vulnerability.

Metabolomics provides a systems-level readout of small molecules influenced by genomic background, diet, medication, organ function, vascular biology, and ongoing pathophysiology. Across ischemic-stroke studies, reported changes have involved amino acids, organic acids, ketone bodies, fatty acids, lysophospholipids, phosphatidylcholines, sphingolipids, acylcarnitines, and oxidized lipid mediators [3–5,11–17,22–31]. Several clinical studies have identified alterations in glutamate, branched-chain amino acids, arginine-related metabolism, glycerophospholipids, sphingolipids, and fatty-acid pathways, while large validation efforts have also reported changes in phenylalanine, glutamine, histidine, HDL-related measures, phosphatidylcholine, and ketone bodies [24–31]. These findings support the concept that circulating metabolism is perturbed after ischemic stroke, but they also demonstrate that the exact molecular signature depends heavily on cohort definition, sampling time, analytical platform, feature space, and statistical strategy.

The problem is especially important for lipid-related metabolites. Phospholipids and sphingolipids are major components of neuronal and vascular membranes, whereas acylcarnitines reflect mitochondrial substrate handling and oxylipins act as bioactive mediators of inflammation, thrombosis, vascular tone, and oxidative injury. Recent reviews have highlighted oxylipin biology as a potentially informative link between ischemic injury and circulating lipid signals, but also note that different oxylipin subclasses may exert opposing biological effects and that analytical confirmation is essential before mechanistic interpretation [17]. Thus, the biological attractiveness of lipid metabolites increases rather than decreases the need for rigorous analytical and statistical control.

A cross-sectional comparison of clinically distinct stroke states can still be useful when its inferential boundaries are explicit. Acute ischemia may reflect immediate tissue injury, stress responses, reperfusion status, and acute treatment exposure; recovery-stage samples may reflect partial resolution, remodeling, or persistent systemic abnormalities; recurrent-stroke samples may represent either unresolved vulnerability or a biologically distinct recurrent event. However, because different individuals contribute to each group, a cross-sectional design cannot establish temporal trajectories. Recent work combining cross-sectional and longitudinal metabolomics in large-vessel occlusion illustrates why repeated sampling is valuable: metabolite directions observed between patients and controls can be tested against within-person evolution across days and months [15]. In the absence of such repeated sampling, stage labels should be treated as clinically defined comparison states rather than as a temporal sequence.

Published clinical metabolomics studies also differ substantially in their inferential goals. Some are designed for acute diagnosis or stroke-subtype discrimination [24,25,27,28], whereas others examine functional recovery, future stroke risk, or external validation in independent cohorts [26,29–31]. For example, targeted metabolomics has discriminated large-artery atherosclerotic from cardioembolic stroke after adjustment for clinical variables [25], while recovery-oriented studies have linked amino-acid patterns to rehabilitation outcomes [26]. Large-scale NMR work has validated circulating stroke-associated metabolites in the UK Biobank [29], and

prospective serum metabolomics has identified metabolic changes preceding thrombotic stroke [30]. These designs provide useful context but are not directly comparable with the present four-group cross-sectional analysis.

The statistical setting is equally important. Untargeted LC-MS datasets may contain hundreds to thousands of correlated features, so dozens of nominal P values below 0.05 can arise even when no feature meets a multiplicity-adjusted threshold. Small studies are additionally vulnerable to unstable effect estimates, optimistic supervised separation, and annotation errors. Best-practice recommendations therefore emphasize pooled-QC monitoring, transparent preprocessing, unsupervised visualization, explicit multiplicity control, conservative metabolite identification, and independent validation [8–10]. These considerations informed the design of the present reanalysis.

We therefore reanalyzed an existing four-group plasma untargeted metabolomics dataset using explicit QC filtering, unsupervised multivariate assessment, feature-wise multiplicity correction, and conservative metabolite interpretation. The primary objective was to test whether overall metabolomic structure or individual feature rows differed across HC, AIS, recovery-stage stroke, and RIS. A secondary, post hoc objective was to describe cross-sectional configurations among nominally associated, QC-qualified features for prioritization in future targeted assays. Predictive classification was not attempted because the sample size and cross-sectional design were not adequate for reliable model development or validation.

## 2. Materials and methods

### 2.1 Study design and participants

This exploratory cross-sectional study analyzed plasma samples from 40 participants assigned to four comparison groups: HC, AIS, recovery-stage stroke, and RIS (n = 10 per group). The source sample register documented sex and collection date but did not include age, vascular risk factors, stroke mechanism, neurological severity, medication exposure, fasting status, renal or hepatic function, timing from the index stroke, or recurrence interval. These variables were therefore unavailable for matching or covariate adjustment. Operational diagnostic criteria and exact sampling-time definitions were not available in the analytic source files; accordingly, the group labels were treated as source-defined clinical categories, and finer temporal or recurrence-interval stratification was not undertaken.

Collection dates ranged from 8 April to 10 May 2025 for HC, 15 April to 17 May 2025 for AIS, 5 April to 21 April 2025 for recovery-stage stroke, and 17 April to 8 May 2025 for RIS. One sample per participant was identified. No participant-level repeated or longitudinal sampling was identified; therefore, all references to stroke 'states', 'patterns', or 'configurations' in this manuscript denote cross-sectional between-person contrasts rather than within-person metabolic trajectories.

### 2.2 Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of The Second Affiliated Hospital of Anhui University of Chinese Medicine (Approval No. 2023-zj-26). Written informed consent was obtained from all participants or their legal representatives.

### 2.3 Plasma extraction and LC–MS acquisition

Plasma was thawed for 1–2 min. A 300-μL aliquot was mixed with 600 μL gas-chromatography-grade dichloromethane and vortexed for 5 min at 1,400 rpm, followed by addition of 600 μL LC–MS-grade methanol

and a further 5-min vortex. After adding 500 µL 0.25 mol/L sodium chloride, samples were vortexed for 10 min and centrifuged at 3,000 rpm for 15 min. The lower dichloromethane and upper methanol–water phases were separated from the intervening protein layer, dried under nitrogen, and stored at −80 °C. Dried extracts were reconstituted in 150 µL LC–MS-grade methanol, vortexed for 5 min, sonicated for 5 min, vortexed for another 5 min, and centrifuged at 6,000 rpm for 20 min before supernatants were transferred to vial inserts.

The available laboratory protocol specified a Dionex U3000 system coupled to a Q Exactive Plus Orbitrap mass spectrometer. Chromatographic separation used either a Kinetex F5 column (2.1 × 50 mm, 1.7 µm) for the dichloromethane fraction or an Acquity UPLC HSS T3 column (150 × 2.1 mm, 1.8 µm) for the aqueous fraction. Mobile phase A was water containing 0.1% formic acid and 1% ammonium acetate; mobile phase B was acetonitrile:methanol (6:4, v/v) containing 0.1% formic acid and 1% ammonium acetate. The gradient was 2% B at 0–1.5 min, 99% B at 9–13.5 min, and 2% B at 14–15 min. Injection volume was 10 µL and column temperature was 45 °C. Data were acquired separately in positive and negative electrospray-ionization modes across m/z 75–1,125 using full-scan/data-dependent $MS^2$ acquisition, with MS1 resolution 70,000, MS2 resolution 17,500, and normalized collision energies of 20, 40, and 60. The protocol stated source voltages of +4.0 kV and −3.5 kV, a capillary temperature of 350 °C, and an auxiliary-heater temperature of 250 °C. A separate project presentation listed a Shimadzu LCMS-8060 system and MS-DIAL software, creating a documentation inconsistency. Because primary instrument records were not available for independent reconciliation in this reanalysis, platform-specific interpretation was avoided and the discrepancy was treated as an analytical-documentation limitation.

### 2.4 Data preprocessing and quality control

The preprocessing strategy was designed to remove poorly detected or analytically unstable feature rows before biological comparison. The ≥80% detection rule and pooled-QC RSD threshold were applied as pragmatic discovery-stage quality filters, and positive- and negative-ion modes were analyzed separately because they represent distinct analytical feature spaces. Run-order and batch metadata were not available in the analytic files; therefore, no empirical drift correction or batch adjustment was fitted. This choice avoided imposing an unverifiable correction model, while residual run-order effects remain a limitation of the dataset.

Positive- and negative-ion feature tables were processed independently. The first 40 sample columns were mapped to the four biological groups using the sample register, and five pooled-QC columns were retained for analytical-precision filtering. Feature names and database annotations were imported as supplied. According to the documented workflow, features were retained when a non-zero signal was present in at least 80% of all biological samples and at least 80% of pooled-QC samples. Intensities were log2-transformed and sample-median normalized within ion mode; features with pooled-QC relative standard deviation (RSD) >30% were excluded, and remaining missing values were replaced by the within-feature median. No data-driven batch correction was applied because run-order and batch metadata were unavailable. The analytic records did not contain sufficient detail to reconstruct the exact computational order, the intensity scale used for QC-RSD calculation, zero handling before log transformation, or stage-specific missingness; these documentation constraints were considered when interpreting the results.

Reported metabolite names are putative library annotations rather than confirmed chemical identities. An annotation was considered database-supported for descriptive prioritization when an HMDB or KEGG identifier was supplied in the source table; this designation does not establish identification confidence. Drug-, dye-, pesticide-, plant-, microbial-, or otherwise implausible matches were flagged as possible exogenous compounds or library mismatches and excluded from the descriptive pattern summary. Authentic-standard confirmation,

retention-time matching, review of raw MS/MS spectra, and blank-aware peak inspection are required before any feature is presented as a definitive metabolite or biomarker [8,9].

### 2.5 Statistical analysis

The statistical analysis was structured hierarchically. First, PCA and PERMANOVA were used to ask whether the multivariate data contained an overall group-level signal. Second, individual feature rows were screened with non-parametric omnibus tests and multiplicity correction. Third, only after these primary analyses were completed were nominally associated features summarized using descriptive rule-based patterns. This hierarchy was chosen to separate confirmatory evidence from hypothesis generation and to avoid allowing visually interesting or pairwise findings to override a negative multiplicity-adjusted omnibus analysis.

After preprocessing, features were standardized to zero mean and unit variance for principal component analysis (PCA). A one-factor PERMANOVA based on Euclidean distances and 1,999 label permutations assessed whether multivariate feature profiles differed by group. Because PERMANOVA can be affected by unequal within-group dispersion, PCA was interpreted descriptively and the PERMANOVA result was not treated as evidence of classification. Formal multivariate-dispersion testing was not available in this reanalysis; accordingly, the PERMANOVA findings were interpreted cautiously and were not used to support classification or biomarker claims.

For each retained feature, the four groups were compared using a Kruskal-Wallis test. P values were adjusted separately within positive- and negative-ion modes using the Benjamini-Hochberg procedure, with FDR-adjusted P (q) $<0.05$ as the threshold for feature-level evidence. The two ion modes were therefore treated as separate multiple-testing families. Six exploratory pairwise Mann-Whitney U tests were also calculated per feature and FDR-adjusted within each pairwise contrast; these analyses were secondary and were not used to override the omnibus result. Median differences are reported as group median minus HC median on the log2-normalized scale. No covariate-adjusted models were fitted because the clinical variables required for defensible adjustment were not available.

For hypothesis generation only, a post hoc descriptive pattern analysis was applied after the primary feature-wise tests. Features first had to show nominal omnibus $P < 0.05$ and an absolute between-group median shift of at least 0.5 log2 units. They were then assigned to rule-based descriptive categories: persistent stroke-associated displacement, RIS-associated divergence, recovery-stage reversal, or acute-associated displacement with recovery-stage normalization. Because this procedure conditions on nominal significance and effect magnitude, the resulting category counts are selection-conditioned descriptive summaries rather than independent statistical evidence. These categories describe cross-sectional between-person configurations only. No prospective sample-size calculation was performed because all currently available samples were analyzed; with $n = 10$ per group, the study was treated as a pilot exploratory analysis. Analyses were conducted in Python 3 using SciPy, statsmodels, scikit-learn, pandas, NumPy, matplotlib, and seaborn.

### 2.6 Quality assurance and reporting framework

The documented analytical workflow was interpreted against published quality-assurance and quality-control guidance for untargeted LC-MS metabolomics. Community studies and mQACC recommendations emphasize system suitability, pooled-QC samples, internal standards, randomized acquisition, blanks, analytical-batch monitoring, transparent feature-filtering criteria, and reproducible preprocessing reporting [18–20,32–36,40]. These recommendations were used as a reporting framework and were not taken as evidence that procedures absent from the available records had been performed.

Metabolite identification was interpreted with the same conservative approach. Accurate-mass or library matches alone do not establish unambiguous chemical identity, particularly in the presence of isomers, adducts, fragments, and co-eluting compounds [8,9,21,37–39]. Accordingly, HMDB or KEGG identifiers in the supplied feature tables were treated as database support rather than confirmation. The analytical documentation available for this reanalysis, and its implications for interpretation, are summarized in Table 1.

**Table 1. Analytical quality-control documentation and implications for interpretation.**

| QA/QC domain | Documented information | Analytical status | Interpretation for this study |
|---|---|---|---|
| Biological samples | 40 plasma samples; 4 groups; n = 10 per group | Documented | Supports cross-sectional group comparison only. |
| Pooled-QC samples | Five pooled-QC columns | Available but limited | Used for precision filtering; batch-wide stability could not be fully assessed. |
| Detection filtering | Non-zero in ≥80% of biological samples and ≥80% of pooled-QC samples | Documented rule | Applied uniformly within each ion mode. |
| QC precision criterion | Pooled-QC RSD ≤30% | Threshold documented | Used as a discovery-stage precision criterion; exact calculation stage was unavailable. |
| Run order / drift | Run-order metadata unavailable | Not assessable | No empirical drift correction was applied; residual run-order effects cannot be excluded. |
| Blank handling | Blank records not available in the analytic files | Not assessable | Carry-over or background contamination could not be evaluated from the available records. |
| Internal standards | Internal-standard records not available in the analytic files | Not assessable | Performance could not be benchmarked against internal-standard behavior. |
| Batch correction | No data-driven correction applied | Documented | No correction model was imposed in the absence of run-order and batch metadata. |

| QA/QC domain | Documented information | Analytical status | Interpretation for this study |
|---|---|---|---|
| Metabolite annotation | Database names and some HMDB/KEGG identifiers | Putative | Names were used only for descriptive prioritization, not as confirmed identities. |
| Identification confidence | No formal MSI level documented | Putative | Orthogonal analytical confirmation is required before molecular-level claims. |
| Software versions / seed | Python 3 packages named; versions and permutation seed not documented | Partially documented | Limits exact computational reproducibility but does not change the reported analysis hierarchy. |
| Independent validation | No validation cohort | Absent | No biomarker or clinical-performance inference was made. |

## 3. Results

### 3.1 Sample characteristics

All 40 registered biological samples and five pooled-QC samples were available for analysis. Each biological group contained ten participants. Sex counts were broadly balanced, although the recovery-stage group included six men and the RIS group included six women (Table 2). No missingness occurred for group, sex, or collection date. Because age and clinical variables were unavailable, baseline comparability and confounding could not be evaluated.

**Table 2. Available group-level sample characteristics**

| Group | n | Women/men | Collection-date range | Clinical variables |
|---|---|---|---|---|
| HC | 10 | 5/5 | 8 Apr–10 May 2025 | Not available |
| AIS | 10 | 5/5 | 15 Apr–17 May 2025 | Not available |
| Recovery-stage | 10 | 4/6 | 5–21 Apr 2025 | Not available |
| RIS | 10 | 6/4 | 17 Apr–8 May 2025 | Not available |

AIS, acute ischemic stroke; HC, healthy control; RIS, recurrent ischemic stroke; TOAST, Trial of Org 10172 in Acute Stroke Treatment; NIHSS, National Institutes of Health Stroke Scale; mRS, modified Rankin Scale. Age,

vascular risk factors, TOAST subtype, NIHSS/mRS scores, medication, fasting status, and timing variables were absent from the available register.

### 3.2 Feature retention and global metabolic structure

The supplied feature tables contained 1,607 positive-mode and 873 negative-mode annotated feature rows. After the documented detection and QC-precision filtering, 1,605 positive-mode and 872 negative-mode feature rows remained (Table 3). PCA showed substantial overlap among the four groups in both ionization modes (Figure 1). Consistently, PERMANOVA did not detect a statistically supported global group effect: positive mode pseudo-F = 0.988, $R^2$ = 0.076, P = 0.486; negative mode pseudo-F = 1.168, $R^2$ = 0.089, P = 0.117.

**Table 3. Quality-control retention and multivariate group tests**

| **Ion mode** | **Raw named features** | **QC-retained features** | Nominal P < 0.05 | FDR q < 0.05 | **PERMANOVA pseudo-F; $R^2$; P** |
|---|---|---|---|---|---|
| Positive | 1,607 | 1,605 | 68 | 0 | 0.988; 0.076; 0.486 |
| Negative | 873 | 872 | 77 | 0 | 1.168; 0.089; 0.117 |

QC retention required a non-zero signal in ≥80% of all biological samples and ≥80% of pooled-QC samples, together with pooled-QC RSD ≤30%. FDR values were Benjamini-Hochberg adjusted separately by ion mode. PERMANOVA used 1,999 permutations.

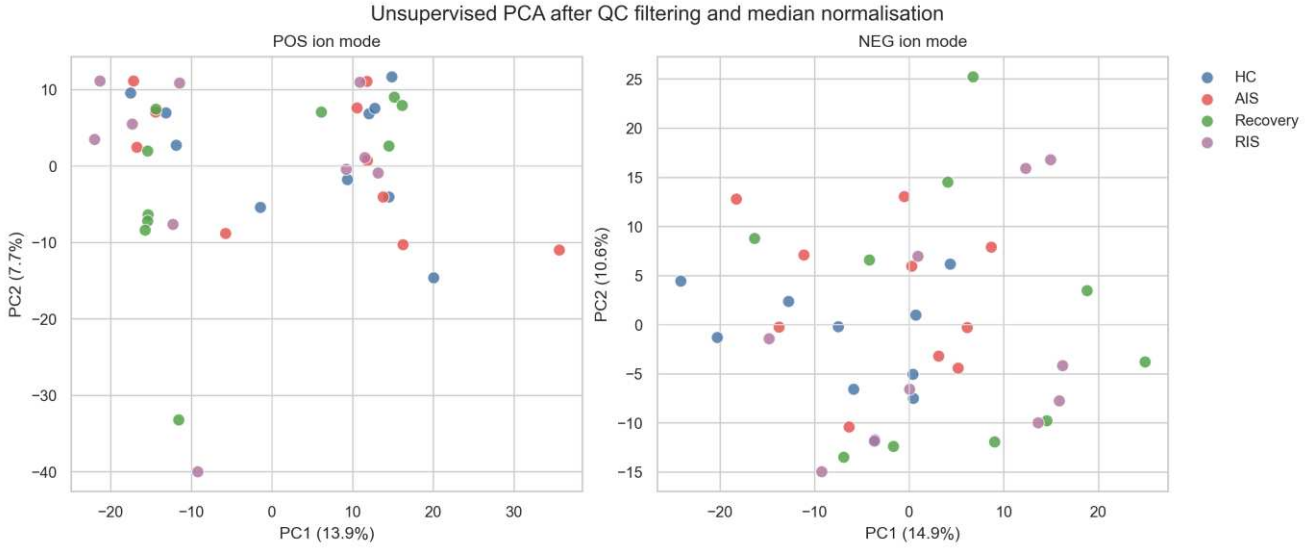



Figure 1. Principal component analysis of QC-qualified plasma metabolomic feature rows. Scores are shown separately for positive and negative ionization modes. The four cross-sectional groups overlap extensively; ellipses are descriptive and do not establish classification performance. AIS, acute ischemic stroke; HC, healthy control; RIS, recurrent ischemic stroke.

### 3.3 Feature-wise evidence after multiplicity control

Kruskal-Wallis screening yielded 68 positive-mode and 77 negative-mode features with nominal omnibus $P < 0.05$. None remained significant at FDR $q < 0.05$, and none reached $q < 0.10$. This multiplicity-adjusted result is the primary feature-level finding and indicates that the nominal signals should not be interpreted as confirmed between-group metabolomic differences. Exploratory pairwise comparisons were therefore retained only as secondary descriptive analyses. The lowest omnibus q values were 0.240 in positive mode and 0.376 in negative mode, illustrating the multiple-testing burden relative to the available sample size.

Figure 2 contrasts nominal and multiplicity-adjusted evidence. Across both ion modes, 145 feature rows met nominal omnibus $P < 0.05$ (68 positive mode and 77 negative mode), but none remained significant after FDR correction. The figure is therefore intended to distinguish hypothesis-generating signals from statistically supported discoveries in this high-dimensional pilot dataset.

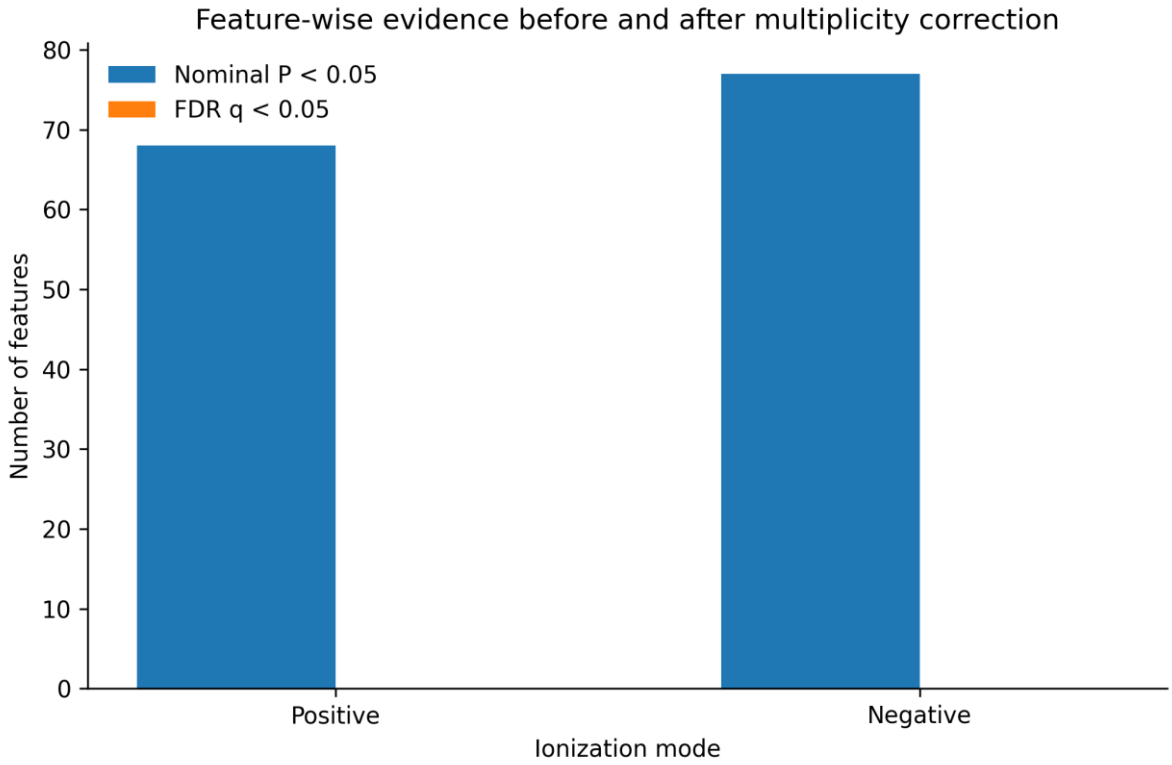



Figure 2. Feature-wise evidence before and after multiplicity correction. Bars show the number of feature rows with nominal omnibus $P < 0.05$ and the number remaining significant at Benjamini-Hochberg FDR $q < 0.05$ in each ionization mode. No feature survived FDR correction.

### 3.4 Exploratory cross-sectional patterns

Among features already selected by nominal omnibus $P < 0.05$, application of the post hoc effect-size criterion and exclusion of annotations flagged as possible exogenous compounds or library mismatches yielded 28 positive-mode and 30 negative-mode candidates assigned to descriptive cross-sectional patterns. Only 8 positive-mode and 12 negative-mode candidates had supplied HMDB or KEGG identifiers. Persistent stroke-associated displacement was the most common configuration (18 positive and 20 negative), followed by RIS-associated divergence (8 positive and 9 negative), recovery-stage reversal (2 positive), and acute-associated displacement with recovery-stage normalization (1 negative) (Figure 3). These counts are conditional

descriptive summaries, not inferential results, because the source features were selected using nominal significance and no feature survived FDR control.

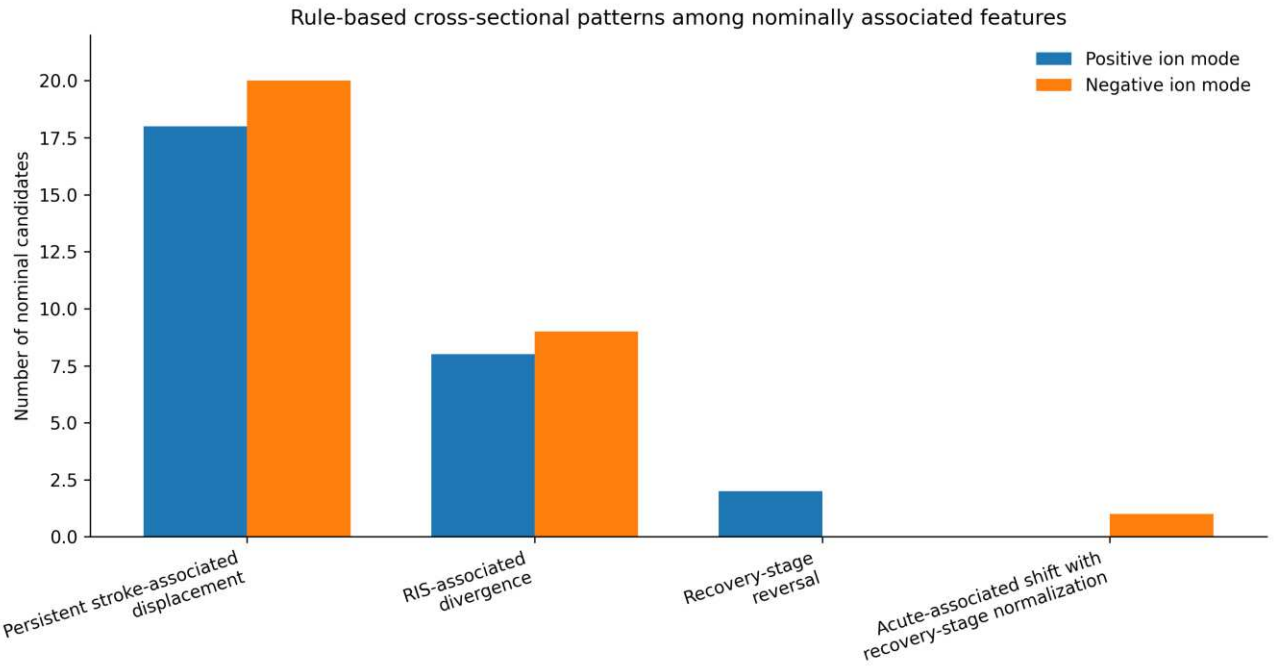



Figure 3. Counts of rule-based cross-sectional configurations among nominally associated, QC-qualified features. Features were required to have nominal omnibus $P < 0.05$, an absolute group-median shift ≥0.5 log2 units, and no flag for a possible exogenous compound or library-mismatch annotation. Counts are selection-conditioned descriptive summaries and do not imply within-person trajectories or independent inferential evidence. POS, positive ionization mode; NEG, negative ionization mode.

Selected database-supported candidates are summarized in Table 4 to support hypothesis generation rather than statistical inference. Bicyclo prostaglandin E2 and 3-hydroxylinoleoylcarnitine showed nominal patterns consistent with persistent stroke-associated elevation relative to HC, whereas cerebronic acid and MG(24:0/0:0/0:0) showed nominal decreases across stroke groups. PE-NMe2(16:0/18:3), 9(10)-EpOME, glycolic acid, and undecanedioic acid showed nominal RIS-associated divergence. All listed features had $q > 0.05$, and all chemical assignments remain putative; consequently, Figures 4 and 5 should be read as descriptive visual summaries of heterogeneous group medians rather than as a validated molecular signature.

**Table 4. Selected database-supported exploratory candidates**

| Mode | Putative annotation | Pattern | QC RSD % | P | q | Δ AIS | Δ Recovery | Δ RIS |
|---|---|---|---|---|---|---|---|---|
| POS | Bicyclo prostaglandin E2 | Persistent displacement | 17.0 | 0.0011 | 0.755 | +2.76 | +3.06 | +1.89 |
| POS | 3-Hydroxylinoleoylcarnitine | Persistent displacement | 4.5 | 0.0131 | 0.973 | +0.66 | +0.81 | +0.48 |
| NEG | Cerebronic acid | Persistent displacement | 3.9 | 0.0037 | 0.376 | −0.54 | −1.06 | −0.65 |
| NEG | MG(24:0/0:0/0:0) | Persistent displacement | 3.2 | 0.0048 | 0.417 | −0.67 | −1.35 | −0.68 |
| NEG | PE-NMe2(16:0/18:3) | RIS divergence | 2.2 | 0.0156 | 0.555 | −0.05 | +0.93 | −1.94 |
| NEG | 9(10)-EpOME | RIS divergence | 2.1 | 0.0160 | 0.555 | −0.86 | −0.70 | −1.55 |
| NEG | Glycolic acid | RIS divergence | 19.8 | 0.0223 | 0.555 | −0.06 | +0.09 | +0.89 |
| NEG | Undecanedioic acid | RIS divergence | 8.8 | 0.0294 | 0.561 | +0.04 | +0.15 | +0.84 |

P is the four-group Kruskal-Wallis P value; q is the Benjamini-Hochberg adjusted P value within ion mode. Δ values are group median minus HC median on the log2-normalized scale. 'Persistent displacement' denotes persistent stroke-associated displacement; 'RIS divergence' denotes RIS-associated divergence. No listed feature survived FDR correction. Metabolite names are putative library annotations and should not be interpreted as confirmed chemical identities without orthogonal analytical verification.

Figure 4 displays the direction and magnitude of median shifts for the eight selected putative candidates. The figure is intentionally effect-direction focused rather than significance focused: it shows that some candidates display broadly persistent displacement across stroke groups, whereas others show larger divergence in the RIS group. These patterns are heterogeneous and should not be interpreted as a coherent multivariate signature.

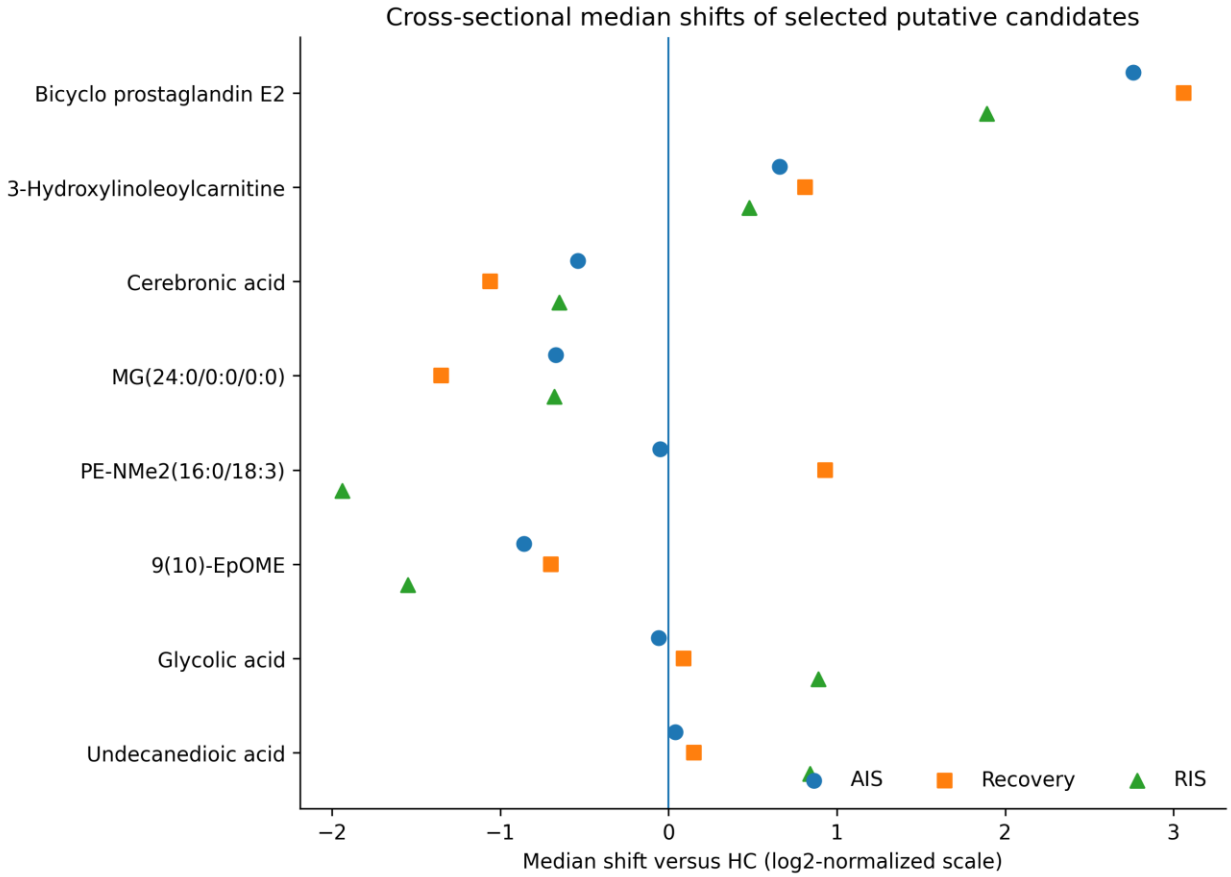



Figure 4. Cross-sectional median shifts of selected putative candidates relative to healthy controls. Points show group median minus HC median on the log2-normalized scale for AIS, recovery-stage stroke, and RIS. All candidates had FDR-adjusted $q > 0.05$; the figure is descriptive and does not establish biomarker performance.

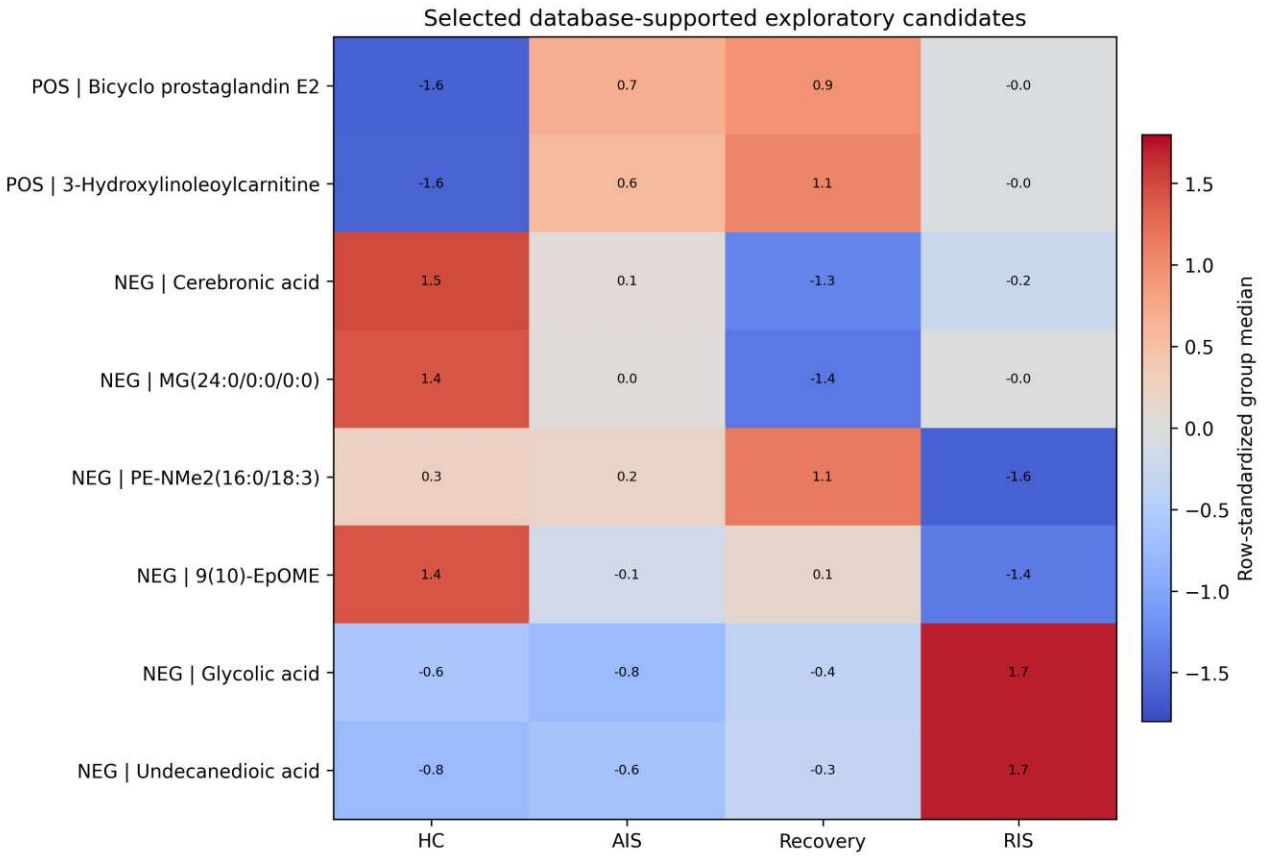



Figure 5. Exploratory heatmap of selected database-supported candidates. Values are row-standardized group medians after log2 transformation and within-mode sample-median normalization. Candidate ranking was exploratory; none of the displayed features survived false-discovery-rate correction. Metabolite names are putative library annotations requiring analytical confirmation. POS, positive ionization mode; NEG, negative ionization mode.

## 4. Discussion

### 4.1 Principal findings

The principal finding of this pilot study is the absence of statistically robust metabolomic separation across the four cross-sectional groups. Positive- and negative-ion PCA showed extensive overlap, PERMANOVA did not support a global group effect, and no individual feature survived FDR correction. These concordant results argue against presenting the current dataset as evidence for a stage-specific or recurrence-associated biomarker signature. Instead, the study defines the boundary between exploratory molecular signals and evidence that is sufficiently robust for clinical interpretation.

This negative result is scientifically informative because untargeted metabolomics combines a high-dimensional feature space with correlated measurements and, in pilot studies, small sample sizes. Under these conditions, nominal P values, visually appealing heatmaps, or apparent subgroup separation can emerge without reproducible feature-level evidence. The present analysis deliberately prioritized unsupervised structure, QC qualification, and multiplicity control over supervised classification. That strategy reduces the likelihood that an apparently high-performing signature would simply reflect overfitting, resubstitution bias, or instability [10].

### 4.2 Biological interpretation of exploratory signals

Although no candidate survived FDR correction, the composition of the nominal list provides biologically plausible directions for targeted follow-up. Several annotations fall within lipid-related classes, including fatty-acyl, glycerolipid, glycerophospholipid, acylcarnitine, and oxidized fatty-acid species. These classes are consistent at a broad level with previous stroke metabolomics literature implicating membrane remodeling, inflammatory lipid signaling, mitochondrial substrate utilization, and oxidative stress [3–7,11,12]. Importantly, this concordance is qualitative rather than inferential: it is not a formal enrichment analysis and does not validate any individual chemical assignment.

Among the selected candidates, bicyclo prostaglandin E2 was nominally higher across the stroke groups, whereas 9(10)-EpOME and PE-NMe2(16:0/18:3) showed nominal RIS-associated divergence; cerebronic acid and MG(24:0/0:0/0:0) were nominally lower. These observations may help define a targeted validation panel spanning oxylipins, acylcarnitines, phospholipids, monoacylglycerols, and very-long-chain fatty-acid species. However, mechanistic interpretation would be premature until the corresponding peaks are supported by raw MS/MS review, retention-time evidence, blank assessment, and preferably authentic standards.

To make the biological interpretation explicit without overstating chemical identity, Table 5 summarizes the broad biochemical context, observed cross-sectional direction, and minimum validation requirement for each selected candidate. The table should be read as a prioritization framework rather than as a mechanistic assignment.

**Table 5. Biological context and analytical validation priorities for selected exploratory candidates.**

| Putative candidate | Broad biochemical context | Observed pattern | Interpretive hypothesis | Validation priority |
|---|---|---|---|---|
| Bicyclo prostaglandin E2 | Prostaglandin/oxylipin-related lipid mediator | Persistent positive shift | May reflect inflammatory or vascular lipid signaling, but exact assignment is uncertain. | Raw MS/MS, RT match, blanks, authentic standard |
| 3-Hydroxylinoleoylcarnitine | Acylcarnitine / fatty-acid oxidation | Persistent positive shift | May reflect altered mitochondrial fatty-acid handling or incomplete metabolic recovery. | Targeted acylcarnitine assay and identity confirmation |
| Cerebronic acid | Very-long-chain 2-hydroxy fatty-acid / membrane-lipid context | Persistent negative shift | Potentially related to membrane or myelin-associated lipid remodeling; plasma origin is uncertain. | Orthogonal identification and targeted quantitation |
| MG(24:0/0:0/0:0) | Monoacylglycerol / neutral-lipid remodeling | Persistent negative shift | May indicate altered glycerolipid turnover; exact positional/isomeric identity is unresolved. | MS/MS/isomer confirmation and targeted lipidomics |

| Putative candidate | Broad biochemical context | Observed pattern | Interpretive hypothesis | Validation priority |
|---|---|---|---|---|
| PE-NMe2(16:0/18:3) | Methylated phosphatidylethanolamine-related annotation | RIS-associated divergence | Potential membrane-remodeling signal; library assignment requires strong confirmation. | Phospholipid-targeted MS/MS and RT verification |
| 9(10)-EpOME | Linoleic-acid-derived epoxide oxylipin | RIS-associated negative divergence | Could relate to oxidized fatty-acid signaling, vascular responses, or inflammation [17]. | Targeted oxylipin assay with stable-isotope standards |
| Glycolic acid | Small organic acid / systemic intermediary metabolism | RIS-associated positive divergence | Nonspecific systemic metabolic signal; susceptibility to preanalytical and exogenous influences. | Confirm chromatographic identity and clinical confounders |
| Undecanedioic acid | Dicarboxylic acid / omega-oxidation context | RIS-associated positive divergence | May reflect alternative fatty-acid oxidation, but biological specificity is low. | Targeted identity confirmation and metabolic-context analysis |

### 4.3 Relationship to previous recurrence studies

The present findings should be interpreted in the context of major differences among published stroke metabolomics designs. Jové et al. analyzed plasma obtained within 24 hours of transient ischemic attack in a 131-patient discovery cohort and confirmed key recurrence findings in an independent cohort of 162 participants; low LysoPC(16:0) and LysoPC(20:4) were among the recurrence-related signals [6]. Purroy et al. subsequently performed untargeted lipidomics in 460 consecutive TIA patients, 37 of whom experienced stroke recurrence within 90 days, and identified seven differential lipid species spanning triacylglycerols, a diacylglycerol, and a plasmalogen phosphatidylethanolamine [7]. These studies were prospective and outcome-based, which is fundamentally different from the present cross-sectional RIS comparison.

Other clinical studies reinforce the importance of cohort definition and analytical design. Non-targeted LC-MS studies have reported differences between acute ischemic stroke and healthy controls and between etiologic subtypes, including changes in fatty-acid and amino-acid metabolism [24]. Targeted studies have identified amino-acid panels for distinguishing ischemic stroke from stroke mimics [11] and for differentiating large-artery atherosclerotic from cardioembolic stroke [25]. Recovery-focused work has linked proline, glutamate, arginine, and related amino-acid patterns to functional recovery [26]. Earlier GC-MS and UPLC/Q-TOF studies likewise reported amino-acid, glycerophospholipid, sphingolipid, acylcarnitine, and purine disturbances [27,28]. At larger scale, NMR-based profiling in 420 stroke participants with validation in 117,988 UK Biobank participants identified reproducible perturbations in amino acids, HDL-related measures, essential fatty acids, phosphatidylcholine, and ketone bodies [29]. Prospective metabolomics has also identified serum metabolic differences in individuals at risk of thrombotic stroke before diagnosis [30]. The diversity of these results

underscores why a small four-group untargeted study should not be expected to reproduce a single universal stroke metabolomic signature.

Table 6 summarizes these representative design contrasts. The comparison is contextual rather than systematic; it is included to show why apparently inconsistent metabolomic results can arise from materially different clinical and analytical questions.

**Table 6. Representative published stroke metabolomics/lipidomics studies and their relation to the present design.**

| Study | Design / sample | Platform / analysis | Key reported result | Relevance to present study |
|---|---|---|---|---|
| Jové et al., 2015 [6] | TIA; 131 discovery + 162 validation; prospective recurrence | Plasma LC-MS; multivariate recurrence analysis | LysoPC(16:0) and LysoPC(20:4) associated with recurrence; independent confirmation | Outcome-based prospective design; stronger basis for prognostic inference |
| Purroy et al., 2023 [7] | 460 TIA patients; 37 recurrences at 90 days | Untargeted plasma lipidomics | Seven recurrence-associated lipid species, mainly glycerolipids plus plasmalogen PE | Shows recurrence-related lipid signals can be subtle even in much larger cohorts |
| Sidorov et al., 2025 [15] | 84 AIS + 82 controls cross-sectional; 15 AIS longitudinal | NMR plus LC-MS | Corroborated directions for several amino-acid and lipid classes across designs | Demonstrates value of repeated sampling for trajectory claims |
| Shi et al., 2025 [16] | 20 ischemic stroke + 20 matched controls | LC-MS lipidomics; 294 assayed lipids; FDR + VIP | 56 FDR-significant lipids; glycerophospholipid metabolism enriched | Illustrates a pilot study with a narrower lipid feature space and positive FDR findings |
| Present study | 40 participants; HC/AIS/recovery-stage/RIS, n=10/group | Untargeted LC-MS feature tables; PCA, PERMANOVA, Kruskal-Wallis, BH-FDR | No global group effect; 68/77 nominal features; 0 FDR-significant features | Cross-sectional, hypothesis-generating comparison without prognostic validation |

## 4.4 Strengths and limitations

The study has two principal strengths. First, the analysis used pooled-QC filtering, separate processing by ionization mode, unsupervised multivariate assessment, and formal multiplicity correction, while avoiding supervised prediction modeling in an underpowered dataset. Second, statistical evidence was kept separate from biological plausibility: database annotation, effect direction, nominal P values, and consistency with prior lipid biology were not treated as substitutes for FDR-controlled support. These choices reduce the risk of overinterpreting discovery-stage patterns.

The principal limitations relate to statistical power, clinical characterization, analytical documentation, and metabolite identity. With n = 10 per group, effect estimates are imprecise and the multiple-testing burden is substantial. Age, vascular risk factors, stroke etiology and severity, renal and hepatic function, medication and dietary exposure, fasting status, and exact sampling intervals were not available for covariate adjustment. The source files also did not permit

verification of detailed stage definitions or recurrence adjudication, so the four groups should be interpreted as source-defined cross-sectional clinical states rather than a temporal sequence.

Analytical limitations further constrain interpretation. Run order, batch structure, blank injections, internal-standard records, and several peak-processing details were unavailable; formal multivariate-dispersion testing was not included; and project documents contained a conflicting acquisition-platform label. Metabolite identities remain putative, the exact computational stage used for QC-RSD calculation was not documented, and only five pooled-QC columns were available. In addition, the ≥80% detection rule across all biological samples may attenuate strong group-specific presence/absence signals, and the post hoc pattern analysis is selection-conditioned. Together with the absence of an independent validation cohort, these constraints preclude temporal, causal, prognostic, or biomarker claims.

These limitations are interdependent. For example, incomplete timing and clinical metadata make it difficult to distinguish a recurrence-associated signal from differences in sampling context, while uncertain annotation and missing run-order information reduce the specificity of subtle molecular patterns. The conservative analytical strategy was therefore intended to define what the present dataset can support reliably and to avoid increasing apparent complexity through unstable predictive modeling.

### 4.5 Analytical quality and annotation confidence

Analytical quality is inseparable from biological interpretation in untargeted metabolomics. Current quality-assurance frameworks emphasize system-suitability testing, pooled QC, internal standards, randomized acquisition, blanks, and transparent peak-processing decisions [19,20,32–36,40]. In the present dataset, five pooled-QC columns and a 30% RSD threshold were documented, whereas run-order coverage, blank and internal-standard records, and the exact RSD implementation were not available. The reported results should therefore be understood as a conservative reanalysis of the documented feature tables rather than as a fully reconstructable raw-data workflow.

Annotation confidence is a second constraint. A database or library match is not equivalent to a confirmed metabolite when isomers, adducts, in-source fragments, co-eluting compounds, or implausible matches are possible. Reviews of LC-MS annotation recommend combining accurate mass with orthogonal evidence such as retention time, curated MS/MS spectra, and authentic standards [37–39]. The candidate names in this manuscript are therefore retained as putative annotations and are used only to define targets for subsequent analytical confirmation.

Figure 6 summarizes the inferential hierarchy adopted in this manuscript. The key distinction is that primary inference ends with multiplicity-controlled testing. Post hoc candidate patterns belong to a separate exploratory layer and acquire biological or clinical meaning only after targeted chemical identification and validation in a clinically characterized longitudinal cohort.

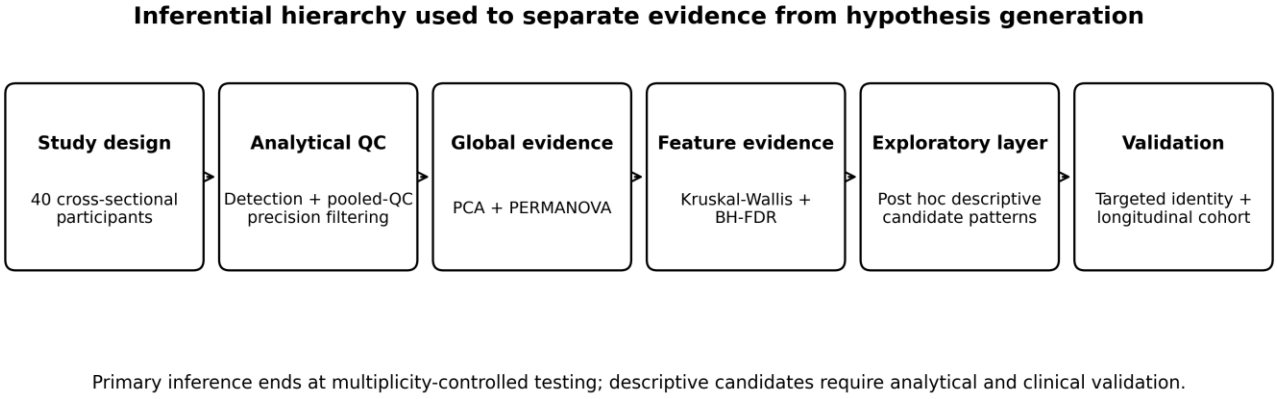



Figure 6. Inferential hierarchy from cross-sectional discovery to validation. The present study supports evidence through QC-qualified global and feature-level testing; post hoc patterns are descriptive. Biomarker interpretation requires targeted metabolite confirmation and independent longitudinal validation.

### 4.6 Implications for validation

The present dataset is not intended to alter recurrence-risk assessment or secondary-prevention decisions. Its practical contribution is to identify a small, QC-qualified set of analytical hypotheses and to clarify the evidence required before any recurrence-related molecular claim can be tested clinically.

A staged validation pathway is appropriate. The eight selected candidates should first be treated as an assay-development shortlist rather than as a biomarker panel. Their identities should then be evaluated by targeted MS/MS, retention-time matching, blank review, and authentic or isotope-labeled standards. Confirmed analytes can subsequently be tested in a clinically characterized longitudinal cohort with prespecified sampling windows and adjudicated recurrent events, followed by multivariable comparison with established clinical predictors.

The next study should therefore be designed as a prospective longitudinal validation program rather than as an extension of the present cross-sectional screening exercise. Sampling windows should be prespecified across the acute, early-recovery, and late-recovery phases, with rigorous adjudication of recurrent events. Clinical metadata should include age, vascular risk factors, TOAST subtype, imaging phenotype, NIHSS, mRS, renal and hepatic function, reperfusion treatment, secondary-prevention medication, fasting status, and exact sampling intervals. Analytical runs should use randomized order, repeated pooled QC, process blanks, and isotope-labeled internal standards. Putative candidates from the present study should first undergo targeted identity and quantitation confirmation; only after this step should prespecified modeling, internal resampling, and locked independent validation be used to evaluate discrimination, calibration, decision-curve utility, or incremental value beyond established clinical predictors.

## 5. Conclusions

In this 40-sample cross-sectional study, QC-aware and multiplicity-aware analysis did not identify a robust global metabolomic difference or any FDR-significant feature across HC, AIS, recovery-stage stroke, and RIS. The remaining nominal, putatively annotated signals are best regarded as hypothesis-generating candidates. Their value

lies in prioritizing targeted chemical confirmation and informing a larger, clinically characterized longitudinal validation study rather than supporting immediate stage-specific or recurrence-associated biomarker claims.

## Supplementary material

Supplementary Tables S1–S4 accompany this manuscript and provide candidate-level numerical results, the statistical-analysis hierarchy, an analytical-documentation summary, and the proposed targeted-validation roadmap.

## Statements and Declarations

### Ethics approval

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of The Second Affiliated Hospital of Anhui University of Chinese Medicine (Approval No. 2023-zj-26).

### Consent to participate

Written informed consent was obtained from all participants or their legal representatives.

### Consent for publication

Not applicable.

### Data availability

De-identified feature-level data are available from the corresponding author upon reasonable request, subject to ethics approval and institutional data-sharing requirements. No public repository or accession number has been assigned.

### Funding

This work was supported in part by the 2025 Key Teaching Research Project of Anhui University of Chinese Medicine (Grant No. 2025xjjy-zd071).

### References

1. Kolmos M, Christoffersen L, Kruuse C. Recurrent ischemic stroke—a systematic review and meta-analysis. J Stroke Cerebrovasc Dis. 2021;30(8):105935. doi:10.1016/j.jstrokecerebrovasdis.2021.105935.

2. Kleindorfer DO, Towfighi A, Chaturvedi S, et al. 2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack. Stroke. 2021;52(7):e364–e467. doi:10.1161/STR.0000000000000375.

3. Li W, Shao C, Li C, et al. Metabolomics: a useful tool for ischemic stroke research. J Pharm Anal. 2023;13(9):968–983. doi:10.1016/j.jpha.2023.05.015.

4. Karmelić I, Rubić I, Starčević K, et al. Comparative targeted metabolomics of ischemic stroke: thrombi and serum profiling for the identification of stroke-related metabolites. Biomedicines. 2024;12(8):1731. doi:10.3390/biomedicines12081731.

5. Ding L, Zhang M, Fan B, et al. Metabolomics reveals key biomarkers for ischemic stroke: a systematic review of emerging evidence. Front Neurol. 2025;16:1630390. doi:10.3389/fneur.2025.1630390.

6. Jové M, Mauri-Capdevila G, Suárez I, et al. Metabolomics predicts stroke recurrence after transient ischemic attack. Neurology. 2015;84(1):36–45. doi:10.1212/WNL.0000000000001093.

7. Purroy F, Ois A, Jové M, et al. Lipidomic signature of stroke recurrence after transient ischemic attack. Sci Rep. 2023;13:13706. doi:10.1038/s41598-023-40838-7.

8. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group Metabolomics Standards Initiative. Metabolomics. 2007;3(3):211–221. doi:10.1007/s11306-007-0082-2.

9. Alseekh S, Aharoni A, Brotman Y, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat Methods. 2021;18:747–756. doi:10.1038/s41592-021-01197-1.

10. Broadhurst D, Goodacre R, Reinke SN, et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics. 2018;14:72. doi:10.1007/s11306-018-1367-3.

11. Tao S, Xiao X, Li X, et al. Targeted metabolomics reveals serum changes of amino acids in mild to moderate ischemic stroke and stroke mimics. Front Neurol. 2023;14:1153193. doi:10.3389/fneur.2023.1153193.

12. Zhao T, Yan Q, Wang C, et al. Identification of serum biomarkers of ischemic stroke in a hypertensive population based on metabolomics and lipidomics. Neuroscience. 2023;533:22–35. doi:10.1016/j.neuroscience.2023.09.017.

13. Ke C, Pan CW, Zhang Y, Zhu X, Zhang Y. Metabolomics facilitates the discovery of metabolic biomarkers and pathways for ischemic stroke: a systematic review. Metabolomics. 2019;15(12):152. doi:10.1007/s11306-019-1615-1.

14. Wei L, Chen S, Deng X, et al. Metabolomic discoveries for early diagnosis and traditional Chinese medicine efficacy in ischemic stroke. Biomark Res. 2024;12(1):63. doi:10.1186/s40364-024-00608-7.

15. Sidorov EV, Smith K, Xu C, Route M, Sanghera DK. Corroborated evidence on change of metabolome after ischemic stroke due to large vessel occlusion. Metabolomics. 2025;21(2):41. doi:10.1007/s11306-025-02235-1.

16. Shi JJ, Nie ZJ, Wang SY, et al. Serum lipidomics profiling to identify potential biomarkers of ischemic stroke: a pilot study in Chinese adults. Biomed Environ Sci. 2025;38(8):918–925. doi:10.3967/bes2025.095.

17. Cheng L, Xia J. Impacts of oxylipins on ischemic stroke: from pathophysiology to clinical phenotypes. Metabolomics. 2025;21(6):148. doi:10.1007/s11306-025-02324-1.

18. Broeckling CD, Beger RD, Cheng LL, et al. Current practices in LC-MS untargeted metabolomics: a scoping review on the use of pooled quality control samples. Anal Chem. 2023;95(51):18645–18654. doi:10.1021/acs.analchem.3c02924.

19. Kirwan JA, Gika H, Beger RD, et al.; metabolomics Quality Assurance and Quality Control Consortium (mQACC). Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management. Metabolomics. 2022;18(9):70. doi:10.1007/s11306-022-01926-3.

20. Mosley JD, Schock TB, Beecher C, et al. Establishing a framework for best practices for quality assurance and quality control in untargeted metabolomics. Metabolomics. 2024;20(2):20. doi:10.1007/s11306-023-02080-0.

21. Kodra D, Pousinis P, Vorkas PA, et al. Is current practice adhering to guidelines proposed for metabolite identification in LC-MS untargeted metabolomics? A meta-analysis of the literature. J Proteome Res. 2022;21(3):590–598. doi:10.1021/acs.jproteome.1c00841.

22. Au A. Metabolomics and lipidomics of ischemic stroke. Adv Clin Chem. 2018;85:31–69. doi:10.1016/bs.acc.2018.02.002.

23. Zhang R, Meng J, Wang X, et al. Metabolomics of ischemic stroke: insights into risk prediction and mechanisms. Metab Brain Dis. 2022;37(7):2163–2180. doi:10.1007/s11011-022-01011-7.

24. Wang X, Zhang L, Sun W, et al. Changes of metabolites in acute ischemic stroke and its subtypes. Front Neurosci. 2021;14:580929. doi:10.3389/fnins.2020.580929.

25. Lee EJ, Kim DJ, Kang DW, et al. Targeted metabolomic biomarkers for stroke subtyping. Transl Stroke Res. 2024;15(2):422–432. doi:10.1007/s12975-023-01137-5.

26. Wang X, Liu T, Song H, et al. Targeted metabolomic profiling reveals association between altered amino acids and poor functional recovery after stroke. Front Neurol. 2020;10:1425. doi:10.3389/fneur.2019.01425.

27. Wang D, Kong J, Wu J, Wang X, Lai M. GC-MS-based metabolomics identifies an amino acid signature of acute ischemic stroke. Neurosci Lett. 2017;642:7–13. doi:10.1016/j.neulet.2017.01.039.

28. Sun H, Zhao J, Zhong D, Li G. Potential serum biomarkers and metabonomic profiling of serum in ischemic stroke patients using UPLC/Q-TOF MS/MS. PLoS One. 2017;12(12):e0189009. doi:10.1371/journal.pone.0189009.

29. Rout M, Vaughan A, Blair A, Stavrakis S, Sidorov EV, Sanghera DK. Discovery and validation of circulating stroke metabolites by NMR-based analyses using patients from the MISS and UK Biobank. Neurochem Int. 2023;169:105588. doi:10.1016/j.neuint.2023.105588.

30. Khan A, Shin MS, Jee SH, Park YH. Global metabolomics analysis of serum from humans at risk of thrombotic stroke. Analyst. 2020;145(5):1695–1705. doi:10.1039/C9AN02032B.

31. Evans AM, O'Donovan C, Playdon M, et al.; Metabolomics Quality Assurance, Quality Control Consortium (mQACC). Dissemination and analysis of the quality assurance and quality control practices of LC-MS based untargeted metabolomics practitioners. Metabolomics. 2020;16(10):113. doi:10.1007/s11306-020-01728-5.

32. Beger RD, Dunn WB, Bandukwala A, et al. Towards quality assurance and quality control in untargeted metabolomics studies. Metabolomics. 2019;15(1):4. doi:10.1007/s11306-018-1460-7.

33. Dunn WB, Kuligowski J, Lewis M, et al.; Metabolomics Quality Assurance and Quality Control Consortium (mQACC). Metabolomics 2022 workshop report: state of QA/QC best practices in LC-MS-based untargeted metabolomics, informed through mQACC community engagement initiatives. Metabolomics. 2023;19(11):93. doi:10.1007/s11306-023-02060-4.

34. Mosley JD, Dunn WB, Kuligowski J, et al.; Metabolomics Quality Assurance and Quality Control Consortium (mQACC). Metabolomics 2023 workshop report: moving toward consensus on best QA/QC practices in LC-MS-based untargeted metabolomics. Metabolomics. 2024;20(4):73. doi:10.1007/s11306-024-02135-w.

35. Lippa KA, Aristizabal-Henao JJ, Beger RD, et al. Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the Metabolomics Quality Assurance and Quality Control Consortium (mQACC). Metabolomics. 2022;18(4):24. doi:10.1007/s11306-021-01848-6.

36. Chaleckis R, Meister I, Zhang P, Wheelock CE. Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. Curr Opin Biotechnol. 2019;55:44–50. doi:10.1016/j.copbio.2018.07.010.

37. Folberth J, Begemann K, Jöhren O, Schwaninger M, Othman A. MS2 and LC libraries for untargeted metabolomics: enhancing method development and identification confidence. J Chromatogr B Analyt Technol Biomed Life Sci. 2020;1145:122105. doi:10.1016/j.jchromb.2020.122105.

38. Blaženović I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites. 2018;8(2):31. doi:10.3390/metabo8020031.

39. Metz TO, Chang CH, Gautam V, et al. Introducing identification probability for automated and transferable assessment of metabolite identification confidence in metabolomics and related studies. Anal Chem. 2025;97(1):1–11. doi:10.1021/acs.analchem.4c04060.

40. Dudzik D, Barbas-Bernardos C, García A, Barbas C. Quality assurance procedures for mass spectrometry untargeted metabolomics: a review. J Pharm Biomed Anal. 2018;147:149–173. doi:10.1016/j.jpba.2017.07.044.