

Metabolomic reconstruction fidelity between people and between visits in COVID-19 cohorts

Roberto Navarro Quiroz

Center for Research in Critical Dynamics

Eloina Zarate Peñata

Universidad Simón Bolívar

Cecilia Fernández-Ponce

Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA)

Lisandro Pacheco Lugo

Universidad Simón Bolívar

Leonardo Pacheco Londoño

Universidad Simón Bolívar

Antonio Acosta-Hoyos

Universidad Simón Bolívar

Nataly Galan Freyle

Universidad Simón Bolívar

Andrea Jaruffe Pinilla

Universidad Simón Bolívar

Katherine Escorcía Lindo

Universidad Simón Bolívar

Ibeth Luna-Rodríguez

Universidad Simón Bolívar

Elkin Navarro Quiroz

`elkin.navarro@unisimon.edu.co`

Universidad Simón Bolívar

Article

Keywords: metabolomics, repeated measurements, covariance, dimensionality reduction, methionine, reproducibility

Posted Date: October 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11144311/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

We evaluated what variation a two-coordinate metabolomic representation retains in held-out people. We compared personal means and between-visit deviations of eight signals in 224 participants and 581 samples from BEAT-COVID, Karu and Cornell. In Karu, learning from within-person deviations reduced their reconstruction loss by 0.0670 squared standardised log units versus learning from personal means; exploratory 2.5–97.5 percentiles were 0.0347–0.1646. Loss decreased in 25 of 33 people, while three signals were reconstructed better and five worse. Relative to total variation, the reduction was 30.2% with standardised logarithms and 3.3% without standardisation. These post hoc ratios describe differently weighted objectives, without their own intervals. The cost of reconstructing personal means remained uncertain. BEAT and Cornell did not establish an advantage as clearly, without demonstrating between-cohort differences. As complementary characterisation, methionine and its sulfoxide covaried in resources with repeated measurements and were associated cross-sectionally in 417 LEOCC participants, sampled a median of 1244 days before diagnosis. The contribution is to evaluate retrospective fidelity for measurements already acquired and its sensitivity to the objective; metabolic function and clinical utility are not established.

Introduction

When several blood samples are collected from a person, it is useful to distinguish the differences that separate that person from others from those observed between their own visits. Both levels may contain relationships between metabolites, but these relationships need not be the same. A person with higher mean abundances of two signals may contribute to a positive association across people even when deviations in those signals during follow-up are weakly related. This distinction affects how a set of measurements is interpreted and what is retained when it is summarised using a few variables.

Metabolomic studies of COVID-19 allow this question to be examined in blood. Amine and lipid profiles, immunometabolic associations, multi-omics signatures during hospitalisation, and individual recovery patterns have been described [1–4]. More recently, Petrova and colleagues studied one-carbon metabolism and COVID-19 severity by combining longitudinal metabolomics and genetics [5]. The resources differ in biological matrix, molecular coverage, treatment, and sampling schedule. Biobank samples collected before a subsequent diagnosis provide an additional context [6, 7]. These differences allow relationships to be examined under different conditions, while limiting direct comparisons between studies.

This study focuses on a common panel of eight signals: methionine, cystathionine, cysteine, glycine, serine, glutamic acid, methionine sulfoxide, and taurine. Their selection allows comparison of resources with partial coverage of compounds related to amino acids and sulphur metabolism. It does not amount to measuring a complete pathway. For example, experiments in the human hepatoma cell line HepG2 link the redox environment, transsulphuration, and glutathione synthesis [8], but covariance in blood does not measure those fluxes. Methionine and its sulfoxide offer a chemically interpretable relationship and

also a challenge: oxidation during sample handling can affect both measurements [9]. Interpretation must therefore begin with the signals actually observed.

Multilevel metabolomics and individual response models already distinguish variation between and within people [10, 11]. Repeated measures correlation also separates units of comparison [12], and metabolic network models show that a correlation can arise under different biochemical organisations [13, 14]. Multigroup components and pathway summaries extend these tools [15–18]. Cai and colleagues proposed dynamic factors for metabolites measured at irregular times, with an application to the COVID-19 resource studied by Ruffieux [19]. Our analysis uses a simpler representation to assess a defined objective: retaining observed measurements from held-out people.

Suppose that eight signals are summarised by two linear combinations. These combinations can be learned from differences between personal means or from deviations observed at their visits. To assess whether they serve the second objective, one can train on other people and measure how much of the held-out person's deviations falls outside that two-dimensional plane. This loss quantifies reconstruction fidelity. Its interpretation depends on the panel and the measurement scale; by itself, it does not measure physiological importance, Shannon information, or clinical prognosis.

The main question is whether the two coordinates learned from differences between people retain within-person deviations as well as those learned from these deviations, and what happens when reconstructing personal means. As a complementary analysis of the same panel, we examine whether previously prioritised molecular relationships recur in an external resource with repeated measurements and in a cross-sectional prediagnostic context. The first question evaluates the fidelity of a summary; the second characterises associations between signals. Bringing them together allows both properties to be described without assuming that recurrence of a pair explains the reconstruction gain for the full panel.

Materials and methods

Design, molecular coverage, and meaning of the measurements

We performed an exploratory reanalysis of publicly available processed measurements of circulating analytes. Earlier results from BEAT-COVID and Karu motivated the question and the prioritisation of pairs. The rules for the new external comparison were fixed locally on 23 September 2026, before calculating the new Cornell and LEOCC associations. This record preserves the sequence of decisions, but it is not a public preregistration and does not make the study confirmatory. The present revision retains the frozen analyses and incorporates post hoc descriptions of the distribution of losses across people and of ratios within each scale, calculated from archived results.

We refer to the common panel of eight analytes as C08. It includes five sulphur-containing compounds and three amino acids without sulphur; it is an analytical panel defined by chemical correspondence and

availability, rather than a validated functional module. It is a subset of the original M02 context of eleven signals. Homocysteine, glutathione, and methionine sulfone were not jointly available with adequate coverage and were not replaced by related compounds. Cysteine and cystine were also not treated as equivalent. The Cornell annotation did not resolve sulfoxide stereochemistry. We analysed relative assay signals, not molar concentrations, enzyme activity, or reaction rates.

For BEAT and Karu, we used the published supplementary materials [1, 2]. We retained the populations with repeated measurements previously selected using the complete-panel rules: 21 people and 78 samples in the first BEAT wave; 23 and 79 in the second; and 33 and 92 in Karu. Reducing the panel to eight signals did not reinstate previously excluded visits or participants. The recorded time references were days since symptom onset in BEAT and days since hospitalisation in Karu. The original analyses of thirty BEAT signals and eleven Karu signals remain identified in the Supplementary Information, without being equated with the new C08 comparison.

Cornell audit and selection

The serum liquid chromatography–mass spectrometry (LC-MS) matrix and annotations came from the version 1 Figshare deposit, DOI 10.6084/m9.figshare.19115972.v1 [3]. The deposited matrix preceded the authors' subsequent quotient normalisation, logarithmic transformation, and imputation. We inspected their scripts to document the provenance and applied our stated rules. The labels d1–d3 refer to the first three days of admission; they do not provide exact blood collection times.

After removing the GIM3 anchor replicates, 580 COVID-19 columns from 330 people remained, compared with the 582 samples described in the source article. This discrepancy remains unresolved. We excluded person 86 because of observations across batches and an ambiguous day; for person 362, we removed both day 3 observations, whose intubation indicators disagreed, while retaining their other eligible visits. Eligibility required an age of at least 18 years, documented absence of pregnancy, positive and complete signals for the corresponding analysis, and two or more distinct eligible days. We did not impute values.

The primary population retained samples with the source field `already_intubated` equal to `False`, described here as 'without an intubation flag at sampling'. This criterion does not require subsequent intubation or identify an untreated population. C08 included 147 people and 332 samples: GIM1, 47/131; GIM2, 100/201. For methionine–methionine sulfoxide, 147/334 were retained; for serine–cystathionine, 147/333; and for cystathionine–cysteine, 147/332. The sensitivity analysis allowing any intubation status represents a different observed population. We retained sample-level decisions, chemical mappings, and source cell coordinates.

The documented preanalytical context also differs. Cornell used remnant serum from clinical samples, usually stored at 4°C for 1–5 days before storage at – 80°C [3]. Karu describes EDTA plasma stored initially at – 20°C and subsequently at – 80°C, with sampling approximately every 3–4 days [2].

Specimen-specific storage durations were unavailable to adjust for these differences. These conditions motivate alternative explanations without demonstrating that the results are artefacts.

Covariance levels and reconstruction objectives

We fitted five strata separately: the two BEAT waves, Karu, and the Cornell GIM1 and GIM2 batches. For person i , x_{it} contains the natural logarithms of the eight signals at visit t ; n_i is the number of eligible visits and m_i the observed mean. Each person receives equal total weight, distributed across their visits. If μ is the average of the personal means and N is the number of people, we use the following empirical decomposition:

$$C_B = \frac{1}{N} \sum_i (m_i - \mu)(m_i - \mu)^T,$$

$$C_W = \frac{1}{N} \sum_i \frac{1}{n_i} \sum_t (x_{it} - m_i)(x_{it} - m_i)^T, C_T = C_B + C_W.$$

C_B describes differences between personal means; C_W , deviations around those means; and C_T is their sum under the same weighting. The letters B, W and T identify models learned from these three matrices, respectively. In this study, “within-person” means deviation from the mean of the available visits. It identifies neither a slope nor a temporal process

the calculation does not use visit order or remove a common trend in clinical day. With two visits per person, it is also not automatically equivalent to Pearson correlation between changes, because these change vectors are not centred across people.

The within-person divisor is n_i , not $n_i - 1$. C_W describes the empirical dispersion of the retained samples; it is not an unbiased estimator of a latent individual variance. With two visits, a person's contribution is one quarter of the outer product of their between-visit difference vector. Thus, visit count and the observed window affect the objective. Giving each person equal total weight prevents those with more samples from dominating the loss, but does not equalise the precision of their means or correct informative follow-up.

In each evaluation, we excluded all visits from one person before fitting. Training means and standard deviations derived from C_T provided a common scale for B, W and T. Two principal directions were learned from each standardised matrix, and the held-out person's observations were projected onto them. Loss L_W averages the squared residuals of their within-person deviations, first across eight signals and visits, then across people. L_B similarly evaluates that person's mean relative to the training centre. Both are expressed in squared standardised log units.

In the notation $L_W(B)$, the subscript identifies what is reconstructed and the parentheses identify how the representation was learned. The main contrasts were $L_W(B) - L_W(W)$ and $L_W(T) - L_W(W)$

a positive value favours W for reconstructing deviations. We also retained L_B for all three models. We did not retrospectively seek an optimal rank; all these comparisons fixed two coordinates. With scale and rank fixed, W minimises within-person loss in its training data among two-dimensional orthogonal projections. This property does not guarantee lower loss in a held-out person. The evaluation outside the fitting data addresses precisely that difference.

Projection uses all eight measured signals from the held-out person. Centring their visits also requires their observed mean. This is a retrospective evaluation of compression in participants excluded from training; it does not predict unmeasured metabolites, future values or outcomes at admission. Models were relearned in each stratum: a fixed signature learned in BEAT was not transported. In Cornell, losses from separately fitted batches were combined with weights 47/147 and 100/147. Geometry was not fitted by mixing uncorrected batches; the aggregate averages losses standardised within each batch, not errors on a common physical scale.

Molecular relationships and the LEOCC comparison

The primary pair for the external comparison was methionine–methionine sulfoxide, P05; serine–cystathionine, P02, and cystathionine–cysteine, P03, were secondary. This priority reflects earlier results and the chemical question fixed before the new external analysis. It does not mean that P05 has the highest correlation, is the only recurring pair or is superior in a confirmatory comparison. We retained all 28 possible C08 pairs in each stratum.

Each correlation was obtained by dividing the corresponding covariance by the product of its standard deviations. W correlation is not the average of individual correlations. In Cornell, we first combined the C_W matrices from both batches according to their numbers of people and then calculated the correlation; this operation incorporates variance magnitudes as well as batch sizes. B correlations are reported by batch. Each person belongs to a single batch

a constant shift on the log scale cancels when centring their visits, but a technical effect that varies between samples need not cancel.

We linked LEOCC ST002829 to the authors' exact public sample-to-person map [6, 7]. The contextual subset comprises 417 different people, each with one prediagnostic sample. The source protocol reports run-day normalisation, imputation, log transformation and Pareto scaling. We used the deposited processed values without a second logarithm or additional imputation; we did not reconstruct those previous operations or an imputation mask. We estimated Pearson correlation as the main analysis; sensitivities were Spearman correlation with average ranks and Pearson correlation between residuals from separate linear models that included an intercept, age at sampling, a female indicator, body mass index and days relative to diagnosis. The LEOCC cross-sectional association is distinct from B and W . Adjustment does not make these participants healthy controls or establish causality.

Normalisation, resampling and sensitivity analyses

The Cornell sensitivity analysis adapted probabilistic quotient normalisation, PQN [20]. After excluding the eight panel signals, 211 candidate signals remained for the reference. Within each population and batch, a signal required at least 75% weighted positive availability. Its reference was the lower weighted median, using the person's multiplicity divided by their number of visits. At least 20 reference signals were required. If m is the number of eligible signals in the learned reference, each sample required at least the larger of 20 and half of m rounded up in valid positive ratios. Its factor was the ordinary median of these signal/reference ratios. Signals were divided by the factor before taking logarithms, without imputation.

For P05 uncertainty, reference selection and factors were recalculated in each resample of people within each batch. For the geometric sensitivity analysis, the reference was learned only from the training participants in each exclusion. The latter provides point estimates without additional resampling. Held-out samples supplied their broad-panel measurements to calculate their factors; this sensitivity analysis therefore does not evaluate whether measuring only eight signals is sufficient.

The frozen analysis specified 1000 bootstrap resamples per stratum, with seeds and multiplicities retained. Each geometric replicate repeated the complete procedure, excluding all copies of the evaluated identity. Ranks and adjustments were recalculated in LEOCC. We report the 2.5th and 97.5th percentiles as exploratory resampling intervals: nominal 95% coverage and simultaneous coverage of the catalogue have not been established. These intervals do not turn selection informed by previous results into confirmatory discoveries. No threshold was applied to the catalogue to declare a functional network.

Omitting one person examined influence. We also evaluated the first two visits, analysis without standardising the logarithms, and inclusion of samples with any intubation status. These operations address different uncertainties. No unfavourable pairs, cohorts or sensitivity analyses were removed. Execution was specified to stop upon numerical degeneracy rather than silently omit fits; no such failures were recorded.

The preceding Spanish revision added four post-analysis descriptions under a separate specification: the interval for the paired contrast $L_B(W) - L_B(B)$ calculated within each already archived replicate; the number of distinct training identities in those replicates; the analyte-specific distribution of Karu losses; and visit frequencies and observed intervals between first and last visits. There were no new fits or changes to the frozen protocol. Paired differences were used, not differences between endpoints of marginal intervals. These descriptions are identified as post hoc in the supplement and revision record.

In this revision, without new fits, we calculated person-specific loss differences, their sign counts and medians, and relative reductions within each scale: one hundred times the difference between the comparator's loss and W loss, divided by the comparator's loss. All five strata and both comparators were retained. Counts were not subjected to binomial tests because evaluations share training

participants. Intervals for the ratios were not estimated. Supplementary Information S9 identifies these descriptions as post hoc and distinguishes the losses for each held-out person from the archived results obtained by omitting one person from the dataset and refitting the procedure.

Implementation and traceability

The analytical environment was Python 3.10.20, NumPy 2.2.6, pandas 2.3.3 and openpyxl 3.1.5. The package retains public processed data, chemical mappings, selection criteria, seeds, multiplicities and code. Full reproduction and separate implementations are documented separately in Supplementary Information S6, with their results, discrepancies and limitations. Reproduction begins with the available processed measurements; it does not reconstruct instrument acquisition.

The English submission includes a curated reproduction archive as Supplementary Data 1 and a figure-reproduction archive as Supplementary Data 2. The source inventory maps retained, separately acquired and omitted files; it does not imply that all historical archives are redistributed. OpenAI Codex provided material assistance with code, execution, literature retrieval, documentation, draft interpretation, translation, figure preparation and editorial review. Agent audits and separate numerical implementations are computational checks, not independent human peer review. Translation and packaging did not change the frozen scientific inputs or analysis rules. Scientific responsibility and final approval rest with the human authors.

Results

Populations and observation windows

The C08 analysis comprises 224 participants with repeated measurements and 581 samples from three data sources: 44 people from the two BEAT waves, 33 from Karu and 147 from Cornell. GIM1 and GIM2 are analytical strata within Cornell, rather than independent cohorts. LEOCC provides a separate comparison involving 417 people sampled once (Table 1).

Table 1
People, samples and the meaning of time in each analysis.

Data source	People / samples	Biological matrix and recorded time reference	Role in the analysis
BEAT, wave 1	21 / 78	Plasma; days since symptom onset	Original repeated-measures dataset
BEAT, wave 2	23 / 79	Plasma; days since symptom onset	Second wave of the same study
Karu	33 / 92	Plasma; days since hospitalisation	Additional repeated-measures dataset
Cornell GIM1	47 / 131	Serum; days 1–3 of admission	External C08 stratum
Cornell GIM2	100 / 201	Serum; days 1–3 of admission	External C08 stratum
LEOCC	417 / 417	Plasma; 42–3485 days before diagnosis	Cross-sectional context

The median intervals between the first and last visits were 10 days in BEAT wave 1, 7 in wave 2, 6 in Karu, 2 in GIM1 and 1 in GIM2; the respective ranges were 3–14, 3–39, 2–19, 1–2 and 1–2 days. In GIM2, 99 of 100 people had exactly two visits. These descriptions of the inputs, added during revision, are detailed in the supplement and do not make the sampling schedules equivalent. In LEOCC, the median of 1244 days before diagnosis precludes interpreting the samples as a baseline immediately preceding infection. In Cornell, P05/P02/P03 retain 334/333/332 samples, respectively, from the same 147 people; Figure S1 uses the 332 samples complete for C08. All subsequent denominators refer to the population included in this reanalysis.

Fidelity of representations for different objectives

In Karu, W reduced within-person loss from 0.1877 with B to 0.1207; T yielded 0.1728. The B – W contrast was 0.0670 [0.0347; 0.1646], and T – W was 0.0522 [0.0220; 0.0832]. Brackets indicate the exploratory 2.5–97.5 resampling percentiles unless otherwise stated. These results describe an advantage of W in reconstructing deviations under the fixed panel, scale and rank (Table 2; Fig. 1).

Table 2
Paired loss differences for two reconstruction objectives.

Stratum	$L_W(B) - L_W(W)$	$L_W(T) - L_W(W)$	$L_B(W) - L_B(B)$, post hoc
BEAT, wave 1	0.0408 [- 0.0355; 0.1652]	-0.0013 [- 0.0518; 0.0422]	-0.0333 [- 0.1971; 0.1151]
BEAT, wave 2	-0.0002 [- 0.0290; 0.0949]	-0.0114 [- 0.0310; 0.0313]	0.1080 [- 0.1681; 0.3213]
Karu	0.0670 [0.0347; 0.1646]	0.0522 [0.0220; 0.0832]	0.1353 [- 0.0106; 0.2315]
Cornell GIM1	0.0143 [- 0.0364; 0.0742]	0.0068 [- 0.0369; 0.0495]	0.0548 [- 0.1276; 0.1463]
Cornell GIM2	0.0164 [- 0.0127; 0.0403]	0.0102 [- 0.0159; 0.0256]	0.0823 [0.0006; 0.2043]
Cornell combined	0.0157 [- 0.0091; 0.0409]	0.0091 [- 0.0124; 0.0261]	0.0735 [- 0.0150; 0.1501]

The first two numerical columns assess within-person deviations: positive values favour W. The last assesses person-specific means: a positive value indicates greater loss with W. All quantities are in squared standardised logarithmic units. The final column and its intervals are a post hoc description calculated from the same archived replicates; the first two correspond to predefined contrasts. Combined Cornell estimates use fixed participant weights after fitting the batches separately. Absolute losses are retained in the supplement and in model_summary.tsv.

For person-specific means in Karu, L_B was 0.4688 with W, 0.3335 with B and 0.3237 with T. The W – B difference was 0.1353, but its paired interval, added during revision, was [- 0.0106; 0.2315]. The point estimate therefore indicates greater loss; the evidence does not establish a positive cost as clearly as it establishes the reduction in within-person loss.

BEAT did not reproduce a clear advantage: B – W was 0.0408 [- 0.0355; 0.1652] in the first wave and - 0.0002 [- 0.0290; 0.0949] in the second. T had a slightly lower within-person point loss than W in both. In the first wave, W also had a lower point loss than B when reconstructing means, so a within-person benefit did not necessarily imply an opposing cost in the excluded participants.

In Cornell, B – W was 0.0143 [- 0.0364; 0.0742] in GIM1 and 0.0164 [- 0.0127; 0.0403] in GIM2. The weighted aggregate was 0.0157 [- 0.0091; 0.0409]; T – W was 0.0091 [- 0.0124; 0.0261]. The external comparison did not establish a clear improvement with W. The maximum angles between the B and W planes ranged from 31.9° in the first BEAT wave to 83.3° in GIM1: a different geometric orientation did not ensure a reconstruction gain.

The subsequent descriptive decomposition for Karu (Fig. 2A) showed that the average advantage combined lower loss for taurine, methionine and glutamic acid with higher loss for the other five signals.

Their contributions to the panel contrast were + 0.0404, + 0.0282 and + 0.0187, respectively; methionine sulfoxide contributed – 0.0075. Each contribution is the analyte's loss difference divided by eight. The sum of all eight contributions at full precision reproduces the overall contrast, 0.0670 after rounding. Figure 2A and the supplement also show cystathionine, cysteine, glycine and serine, with their negative contributions. These estimates have no analyte-specific intervals; they locate the loss, rather than identifying causal effects or biological importance.

Distribution across people and magnitude of sensitivity to scale

In Karu, W had lower loss than B in 25 of 33 people and lower loss than T in 22 of 33 (Fig. 2B). The median B – W difference per person was 0.0489. When one person was omitted from the dataset and the procedure was rerun, the archived mean B – W contrasts remained between 0.0589 and 0.0802. This range describes influence within the retained dataset, rather than population uncertainty. The distribution across people complements the decomposition across signals: neither replaces the other.

Counting signs also does not replace the mean magnitude. In BEAT wave 2, W had lower loss than B in 16 of 23 people, although the mean B – W difference was – 0.0002. Unfavourable losses in some people can offset more frequent, smaller improvements. Supplement S9 retains the counts, means, medians and omission sensitivities for all five strata, without interpreting them as a population frequency of benefit.

In Karu, the relative loss reduction against B was 35.7% with standardised logarithms and 41.1% without standardisation. Against T, it was 30.2% and 3.3%, respectively. Each ratio uses the comparator's loss within its own scale. Retaining the sign does not imply retaining the relative magnitude of the advantage. These are post hoc descriptions without their own intervals; changing standardisation changes the weights and the objective. Absolute losses were not subtracted across scales, and the percentages were not interpreted as recovered biological information. Table S16 retains both scales and both comparators in every stratum.

Complementary panel covariation and previously prioritised relationships

The complete catalogue retains all 28 unordered pairs in the five strata (Figure S1), with B and W estimates, pointwise intervals and individual omissions. It contains positive associations other than P05. For example, glycine–serine has positive W correlations in all five strata. This descriptive result prevents interpreting the narrative priority of methionine–sulfoxide as having outperformed every other pair. The catalogue was not filtered to construct a discovery network.

The methionine–sulfoxide W correlation was positive in BEAT wave 1 (0.354), wave 2 (0.418) and Karu (0.595); their B correlations were 0.745, 0.209 and 0.126. Recurrence of the within-person sign did not imply a consistent ordering of B and W strengths. In the Cornell pair analysis, W was 0.522 [0.299; 0.684]

in GIM1 and 0.461 [0.295; 0.611] in GIM2. Combining the covariances first and then standardising yielded 0.482 [0.341; 0.596]. B was 0.511 [0.270; 0.692] and 0.330 [0.070; 0.584], respectively (Fig. 3).

Table 3
Prioritised relationships in repeated measurements and the cross-sectional context.

Pair; Cornell samples	Cornell W; 147 people	LEOCC Pearson; 417 people	LEOCC adjusted Pearson
Met / MetSO; 334	0.482 [0.341; 0.596]	0.493 [0.394; 0.614]	0.493 [0.393; 0.617]
Ser / Cth; 333	0.187 [0.056; 0.319]	-0.080 [- 0.177; 0.033]	-0.013 [- 0.120; 0.101]
Cth / Cys; 332	0.136 [- 0.036; 0.295]	0.146 [0.022; 0.266]	0.059 [- 0.062; 0.181]

Met: methionine; MetSO: methionine sulfoxide; Ser: serine; Cth: cystathionine; Cys: cysteine. Cornell combines within-person covariances from pair-specific populations. LEOCC contains one sample per person; adjustment includes age at sampling, sex, body mass index and days relative to diagnosis. Their coefficients do not estimate the same common parameter, and no formal test of between-study heterogeneity was performed.

The secondary relationships did not show the same pattern as P05 (Table 3). Serine–cystathionine was weakly positive in Cornell and not clearly different from zero in LEOCC. Cystathionine–cysteine was uncertain in Cornell; in LEOCC, the point coefficient decreased from 0.146 to 0.059 after adjustment. In contrast, methionine–sulfoxide remained positive in LEOCC using Pearson, 0.493 [0.394; 0.614]; Spearman, 0.536 [0.450; 0.602]; and adjusted Pearson, 0.493 [0.393; 0.617]. The cross-sectional association broadens the observed context, but does not replicate a within-person deviation.

Sensitivity to normalisation, visits and coverage

After PQN, the combined Cornell methionine–sulfoxide W coefficient was 0.309 [0.180; 0.423], compared with the primary estimate of 0.482. The batch-specific estimates were 0.317 and 0.307. Bootstrap references included 127–136 signals in GIM1 and 124–129 in GIM2; every replicate met the support requirements. Restricting the analysis to the first two visits yielded 0.481. Allowing any intubation status yielded 0.503 in 192 people and 437 samples. When one person was omitted, the coefficient ranged from 0.452 to 0.501; this range describes influence, rather than population uncertainty.

Sensitivity was greater for the secondary pairs. Serine–cystathionine changed from 0.182 to - 0.119 in GIM1 when only two visits were retained; its combined estimate changed from 0.187 to 0.105. In the geometry analysis, the point advantage of W in Karu retained its sign with two visits. Removing standardisation favoured B in BEAT wave 1; the signs and magnitudes of the other comparisons are presented in Table S16. With PQN, the combined Cornell geometric B - W contrast was 0.0266 as a point estimate without additional bootstrap analysis; it does not resolve the uncertainty in the primary contrast.

The original panels provide relationships that C08 does not cover. The 9,10-EpOME–9,10-DiHOME and 12,13-EpOME–12,13-DiHOME pairs had high W correlations in both BEAT waves; their intervals for between-level differences included zero. The negative cysteine–glutathione W relationship in Karu did not receive clear support in both BEAT waves. These results are retained in the supplement, with their own populations and panels. The absence of the required compounds prevented applying the same external C08 comparison to them.

Resampling diagnostics

The median numbers of distinct training identities per bootstrap evaluation were 12, 14, 20, 29 and 63 in the five strata, compared with 20, 22, 32, 46 and 99 in the original exclusions. Excluding all copies prevents identity sharing between training and evaluation, but the smaller number of distinct people can change fitting behaviour. The percentiles are retained as an exploratory description; calibrated coverage for training on $N - 1$ original people is not claimed. The diagnostic is detailed in Supplement S7 (Table S13); reproduction checks are described in S6.

Discussion

Reconstruction fidelity and heterogeneity among signals

The clearest evidence of reduced loss was observed in Karu. Two directions learned from within-person deviations reconstructed those deviations with lower loss than directions learned from personal means or total variation. The average combined lower loss for taurine, methionine and glutamic acid with higher loss for the other five signals. BEAT and Cornell did not establish an advantage with the same clarity. This pattern of intervals is not a test of differences between cohorts, whose sampling schedules, sample matrices and populations are also not equivalent. The practical use is to compare ways of retrospectively summarising an already measured panel when the aim is to preserve a particular level of variation.

The observed advantage is asymmetric. In Karu, the within-person contrast favoured W, whereas the paired interval for the cost of reconstructing personal means included zero. This prevents describing a general trade-off between the two targets or concluding that the models are equivalent. The analyte-level breakdown also shows why a lower average loss does not benefit every measurement. It is a post hoc descriptive diagnostic, without signal-specific intervals, and does not support selecting a confirmatory signature.

Scale determines how differences are weighted in reconstruction. Standardising by total variation in the training data reduces the influence of numerical amplitude without assigning equal physiological importance to the analytes. In Karu, the contrast against T retains its sign without standardisation, but its relative loss reduction is 3.3%, compared with 30.2% with standardisation. These ratios characterise different targets; they do not establish an inferential difference between scales. Some signs also change

in BEAT. Retaining both scales and the B and T comparators prevents summarising this sensitivity as uniform robustness or retrospectively choosing the most favourable weighting.

Molecular relationships and compatible explanations

The methionine–methionine sulfoxide association characterises a complementary property of the panel. It recurs within people in the repeated-measures resources and between people in the prediagnostic LEOCC samples. In Cornell, its sign is retained by batch, when participants are omitted and when the analysis is restricted to two visits. These sensitivity analyses show stability to specific decisions; they did not test whether the association explains the multivariate gain. Indeed, in Karu, methionine and methionine sulfoxide contribute with opposite signs to the reconstruction contrast. Statistical independence between these two properties was not established either.

There is direct chemical and clinical precedent. Potgieter and colleagues observed a methionine–methionine sulfoxide correlation of 0.352 in 127 healthy women that disappeared after removing two probable outliers [9]. That result justifies examining influence and sample handling without establishing a universal baseline correlation. Petrova and colleagues integrated longitudinal metabolomics and genetics in IMPACC and linked one-carbon metabolism to COVID-19 severity and outcomes [5]. Their clinical question differs from ours. That precedent does not confer prognostic or functional value on our covariation or reconstruction loss.

The decrease in the Cornell point estimate from 0.482 to 0.309 after PQN shows sensitivity to a transformation of the shared scale. PQN estimates a factor from a reference profile [20]. Excluding the eight panel signals and relearning the reference within resampling limits some dependencies in the adjustment. However, the factor may capture global biological variation, and a shared divisor may introduce dependence. Attenuation does not quantify the fraction of the association that was artefactual; the remaining coefficient does not measure *in vivo* oxidation either.

The sample matrix and handling remain relevant explanations. Cornell used residual clinical serum under conditions different from those of the Karu plasma [2, 3]. Experiments on serum, plasma, collection tubes and handling times show that these factors modify metabolomic profiles [21, 22], but do not establish the direction of their effect on this coefficient. Without specimen-specific timings or independent measurements of the processes involved, several explanations remain compatible with the observations (Table 4).

Table 4
Compatible explanations and the observations needed to distinguish them.

Observation	Compatible explanations	What remains to be observed
Methionine and methionine sulfoxide covary	Shared availability, transport or clearance; in vivo or ex vivo oxidation	Chemical and temporal measurements that separate processes, together with specimen-specific handling information
P05 changes after PQN	Technical scaling, global biological variation or dependence introduced by the divisor	Scaling indicators and analytical controls independent of the eight panel signals
W reconstructs better in Karu	Clinical window, population composition, noise and molecular coverage	Evaluation of the same target in comparable data and its relationship to an external observable
P05 appears in prediagnostic LEOCC	Between-person differences, pre-existing conditions and biobank selection	Appropriate comparators and repeated measurements addressing a defined clinical question

The table presents compatible hypotheses, not established mechanisms or a diagnosis of artefact. Each current sensitivity analysis examines a particular dependence of the result; none directly measures flux, cellular origin or antioxidant function.

Molecular coverage and observed time

The common panel maintains chemical correspondence across resources, but its coverage is partial. The absence of homocysteine and glutathione restricts the interpretation of transsulphuration and antioxidant defence. Evaluations of pathway scores and metabolomic simulations show that coverage and assumptions condition functional summaries [16, 17, 23]. In C08, the biochemical grouping guides the question without establishing a functional module. The EpOME–DiHOME results from the original panels are retained in the supplement with their own populations. Although experimental evidence supports the conversion of epoxides to diols [24], those compounds did not undergo the same external contrast, and their covariation does not estimate enzyme activity.

The observation window also defines the result. Cornell is concentrated within the first three days of admission; Karu and BEAT have different schedules and temporal reference points. C_W summarises the dispersion of available visits without using their order or adjusting for a common clinical trend. A person's mean combines their state during the window with measurement variation; it does not identify a stable trait. Giving each participant equal total weight does not correct sampling conditioned on their clinical course either.

LEOCC provides a cross-sectional association in samples often collected years before the subsequent diagnosis. It is neither a healthy reference nor the baseline state of the Cornell patients. Its related publications describe the same resource [6, 7]. Numerical similarity between coefficients does not permit treating them as estimates of a common parameter. Its contribution is to place the association

outside the acute window of the other cohorts, subject to the biobank's selection and processing limitations.

Contribution relative to earlier methods

Separation between levels and component-based reduction have precedents [10–12, 15–18]. Ruffieux and colleagues modelled recovery and relationships between molecular layers in COVID-19 [4]. Cai and colleagues reanalysed metabolites from that resource using dynamic factors that incorporate irregular sampling times [19]; these two applications are not independent biological replications. Our analysis addresses a narrower question: retrospectively evaluating how much two coordinates retain of observed measurements. We did not compare it with those models, so its simplicity does not establish superiority.

The empirical contribution is to quantify that fidelity under the same evaluation framework, showing magnitudes, their distribution among people and analytes, and the uncertain comparisons. Rank two was fixed for the comparison; it was not shown to be optimal. A methodological extension should specify which additional property it would preserve and how its improvement over this reference would be evaluated. Studying function would also require an independent observable. Neutrophil metabolomics accompanied by cellular assays [25] and joint studies of cytokines and metabolites [26] exemplify designs with other measurements; they do not identify the cellular origin of the circulating signals in this reanalysis.

Uncertainty and generalisation

Excluding a person entirely avoids using their visits to fit their own representation; individual evaluations share training participants and retain the uncertainty of small samples [27]. In the bootstrap, learning uses fewer distinct identities than in the original exclusions. This reduction is inherent to resampling with replacement and does not by itself establish invalidity. Reproducing its percentiles does not establish nominal coverage, and simultaneous coverage has not been established for the catalogue of pairs. Magnitudes, exploratory intervals and sensitivity analyses must therefore be interpreted together. A calibration assessment would require scenarios and criteria fixed before its execution; the procedure was not changed retrospectively to strengthen a conclusion.

Panel selection and priorities, informed by earlier analyses, limit confirmatory interpretation [28]. Complete-case selection and the requirement for repeated visits restrict the population represented. Retaining the historical BEAT and Karu populations facilitates traceability but may exclude samples eligible for the reduced panel. The Cornell count discrepancy and the lack of an original LEOCC imputation mask remain unresolved. Generalisation refers to excluded participants within each stratum, whose representations were relearned; no fixed signature was transported across cohorts. Nor did we evaluate whether better reconstruction improves a clinical decision or better characterises an immune function.

Conclusion

Separating personal means from deviations between visits allowed us to evaluate the fidelity of a two-coordinate metabolomic representation. Karu showed the clearest reduction in within-person loss, distributed across several people but unevenly among signals. Its relative magnitude against the total model depended on weighting: 30.2% with standardised logarithms and 3.3% without standardisation, as post hoc descriptions of different targets. The other evaluations remained uncertain, without establishing differences between cohorts. The cost of reconstructing personal means in Karu was not clearly established either. Methionine–methionine sulfoxide covariation provides a complementary characterisation and does not validate that reconstruction. The result guides the evaluation of retrospective summaries of a measured panel; its functional or clinical usefulness still requires independent observables.

Declarations

Author contributions

Elkin Navarro Quiroz: Conceptualization; Formal analysis.

Roberto Navarro Quiroz, Eloina Zarate Peñata, Cecilia Fernández-Ponce, Lisandro Pacheco Lugo, Leonardo Pacheco Londoño, Antonio Acosta-Hoyos, Nataly Galan Freyle, Andrea Jaruffe Pinilla, Katherine Escorcía Lindo and Ibeth Luna-Rodríguez: Writing – review & editing.

All authors approved the manuscript and its submission.

Funding

This work was supported by the General System of Royalties of Colombia (Sistema General de Regalías, SGR), project BPIN 2024000100078 (to Lisandro Pacheco Lugo).

The funder had no role in study design, data analysis or interpretation, manuscript preparation, or the decision to submit the manuscript.

Acknowledgements

The authors thank the Asociación Colombiana de Inmunología (ACOI) and the investigators who made the source data publicly available.

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study reanalysed publicly available, de-identified metabolomic data and associated study metadata; no participants were newly recruited and no biological specimens were collected. The ethical oversight

and informed-consent or consent-waiver procedures for the original studies are described in their respective publications [1–3,7].

Code availability

The analysis and verification code, fixed settings, dependency versions and execution instructions are provided in Supplementary Data 1. From a fresh extraction, `code/reproduce_submission.py` retrieves the specified missing sources, verifies their hashes, reconstructs the inputs, reruns the original scientific analyses and checks the descriptive outputs in a new output directory. Initial source acquisition requires network access; the subsequent calculations run from local files. The original 93-file scientific lock is preserved. Supplementary Data 2 provides the English figure-generation code, reference figures, plotting inputs and verification instructions. These materials establish the documented scope of computational reproduction; they do not reproduce instrument acquisition or undocumented processing in the source studies.

Data availability

We used public processed data from Goris and colleagues [1], Karu and colleagues [2], Cornell Figshare version 1 10.6084/m9.figshare.19115972.v1, and LEOCC Metabolomics Workbench ST002829, project PR001771 10.21228/M8SM6H. The public LEOCC map is identified by commit `e3a4c3e02ff46117f2203da8089ad40c5fa32b76`. The source studies, public accessions and selection procedures are identified in the Methods and Supplementary Information. Supplementary Data 1 provides the processed analysis inputs, source mappings, reference results, provenance records and integrity manifests used in this reanalysis. Nine required upstream files are acquired separately: five from fixed GitHub commits and four from official Metabolomics Workbench accession URLs. Their expected byte counts and SHA-256 values are recorded in `THIRD_PARTY_ACQUISITION.json` and verified before calculation. The original BEAT workbook is not redistributed; its earlier direct-workbook comparison is preserved as historical evidence and is not repeated by the current wrapper. Source-specific terms are documented in the archive. Supplementary Data 2 provides the derived numerical tables used to generate Figures 1–3 and Supplementary Figure S1.

References

1. Goris, M. et al. Tracking COVID-19 Severity and Progression Through Amines and Lipid Mediators. *J. Med. Virol.* **98** (7), e71030. 10.1002/jmv.71030 (2026).
2. Karu, N. et al. Severe COVID-19 Is Characterised by Perturbations in Plasma Amines Correlated with Immune Response Markers, and Linked to Inflammation and Oxidative Stress. *Metabolites* **12** (7), 618. 10.3390/metabo12070618 (2022).
3. Buyukozkan, M. et al. Integrative metabolomic and proteomic signatures define clinical outcomes in severe COVID-19. *iScience* **25** (7), 104612. 10.1016/j.isci.2022.104612 (2022).

4. Ruffieux, H. et al. A patient-centric modeling framework captures recovery from SARS-CoV-2 infection. *Nat. Immunol.* **24** (2), 349–358. 10.1038/s41590-022-01380-2 (2023).
5. Petrova, B. et al. MTHFR allele and one-carbon metabolic profile predict severity of COVID-19. *Proceedings of the National Academy of Sciences*, 122(51), e2509118122. (2025). 10.1073/pnas.2509118122
6. Chatelaine, H. A. S. et al. Nucleotide, Phospholipid, and Kynurenine Metabolites Are Robustly Associated with COVID-19 Severity and Time of Plasma Sample Collection in a Prospective Cohort Study. *Int. J. Mol. Sci.* **25** (1), 346. 10.3390/ijms25010346 (2023).
7. Chen, Y. et al. The value of prospective metabolomic susceptibility endotypes: broad applicability for infectious diseases. *EBioMedicine* **96**, 104791. 10.1016/j.ebiom.2023.104791 (2023).
8. Vitvitsky, V., Mosharov, E., Tritt, M., Ataulakhanov, F. & Banerjee, R. Redox regulation of homocysteine-dependent glutathione synthesis. *Redox Rep.* **8** (1), 57–63. 10.1179/135100003125001260 (2003).
9. Potgieter, H. C. et al. Spontaneous oxidation of methionine: effect on the quantification of plasma methionine levels. *Anal. Biochem.* **248** (1), 86–93. 10.1006/abio.1997.2075 (1997).
10. van Velzen, E. J. J. et al. Multilevel data analysis of a crossover designed human nutritional intervention study. *J. Proteome Res.* **7** (10), 4483–4491. 10.1021/pr800145j (2008).
11. Jansen, J. J., Szymańska, E., Hoefsloot, H. C. J. & Smilde, A. K. Individual differences in metabolomics: individualised responses and between-metabolite relationships. *Metabolomics* **8** (Suppl 1), 94–104. 10.1007/s11306-012-0414-8 (2012).
12. Bakdash, J. Z. & Marusich, L. R. Repeated Measures Correlation. *Front. Psychol.* **8**, 456. 10.3389/fpsyg.2017.00456 (2017).
13. Steuer, R., Kurths, J., Fiehn, O. & Weckwerth, W. Observing and interpreting correlations in metabolomic networks. *Bioinformatics* **19** (8), 1019–1026. 10.1093/bioinformatics/btg120 (2003).
14. Camacho, D., de la Fuente, A. & Mendes, P. The origin of correlations in metabolomics data. *Metabolomics* **1**, 53–63. 10.1007/s11306-005-1107-3 (2005).
15. Codesido, S. et al. Network principal component analysis: a versatile tool for the investigation of multigroup and multiblock datasets. *Bioinformatics* **37** (9), 1297–1303. 10.1093/bioinformatics/btaa954 (2021).
16. Wieder, C., Lai, R. P. J. & Ebbels, T. M. D. Single sample pathway analysis in metabolomics: performance evaluation and application. *BMC Bioinform.* **23**, 481. 10.1186/s12859-022-05005-1 (2022).
17. Wieder, C. et al. PathIntegrate: Multivariate modelling approaches for pathway-based multi-omics data integration. *PLoS Comput. Biol.* **20** (3), e1011814. 10.1371/journal.pcbi.1011814 (2024).
18. Guilmineau, C., Tremblay-Franco, M., Vialaneix, N. & Servien, R. Phoenix: a novel statistical approach for longitudinal metabolomic pathway analysis. *BMC Bioinform.* **26**, 105. 10.1186/s12859-025-06118-z (2025).

19. Cai, J., Goudie, R. J. B. & Tom, B. D. M. Dynamic Factor Analysis for Sparse and Irregular Longitudinal Data: An Application to Metabolite Measurements in a COVID-19 Study. *Stat. Med.* **45** (6–7). 10.1002/sim.70499 (2026). e70499.
20. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ¹H NMR metabolomics. *Anal. Chem.* **78** (13), 4281–4290. 10.1021/ac051632c (2006).
21. Yu, Z. et al. Differences between human plasma and serum metabolite profiles. *PLOS ONE*. **6** (7), e21230. 10.1371/journal.pone.0021230 (2011).
22. Cruickshank-Quinn, C. et al. Impact of Blood Collection Tubes and Sample Handling Time on Serum and Plasma Metabolome and Lipidome. *Metabolites* **8** (4), 88. 10.3390/metabo8040088 (2018).
23. Cooke, J. et al. Simulated metabolic profiles reveal biases in pathway analysis methods. *Metabolomics* **21** (5), 136. 10.1007/s11306-025-02335-y (2025).
24. Moghaddam, M. F. et al. Bioactivation of leukotoxins to their toxic diols by epoxide hydrolase. *Nat. Med.* **3**, 562–566. 10.1038/nm0597-562 (1997).
25. Li, Y. et al. Neutrophil metabolomics in severe COVID-19 reveal GAPDH as a suppressor of neutrophil extracellular trap formation. *Nat. Commun.* **14**, 2610. 10.1038/s41467-023-37567-w (2023).
26. Xiao, N. et al. Integrated cytokine and metabolite analysis reveals immunometabolic reprogramming in COVID-19 patients with therapeutic implications. *Nat. Commun.* **12**, 1618. 10.1038/s41467-021-21907-9 (2021).
27. Varoquaux, G. Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage* **180**, 68–77. 10.1016/j.neuroimage.2017.06.061 (2018).
28. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **7**, 91. 10.1186/1471-2105-7-91 (2006).

Figures

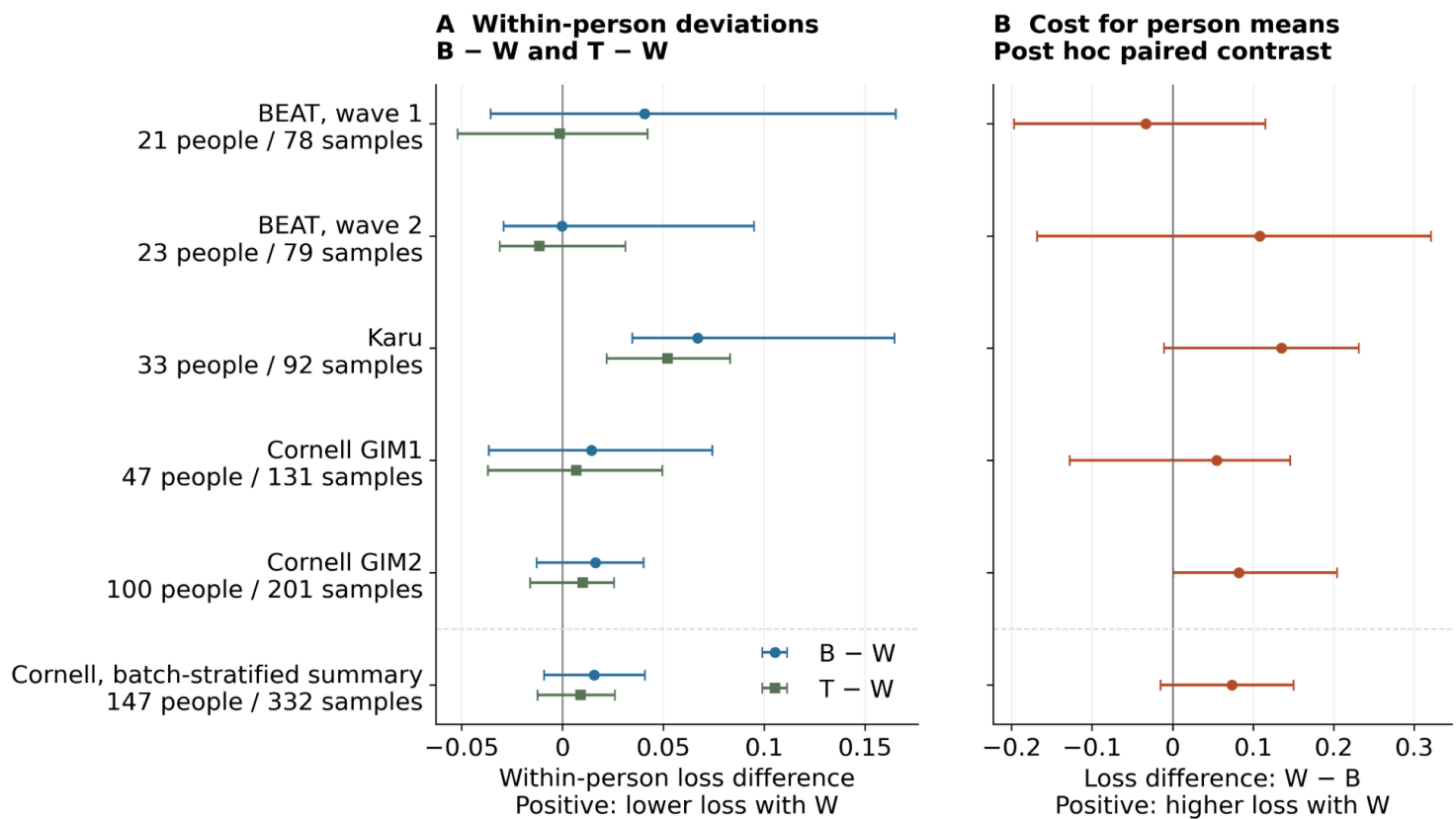


Figure 1

Representation fidelity according to the evaluated objective

All comparisons use two coordinates of the same eight signals. B is learned from personal means; W, from within-person deviations; and T, from total covariance. Panel A shows paired contrasts in within-person loss; positive values favour W. Bars show the exploratory 2.5–97.5 percentiles of 1000 person-level resamples, without established nominal or simultaneous coverage. Cornell combined weights separately fitted batches and is not another cohort. Panel B shows the paired contrast $L_B(W) - L_B(B)$ for personal means and its exploratory 2.5–97.5 percentiles, derived post hoc from the archived replicates. Positive values indicate greater loss with W. Both panels use squared standardised log units. Sources: geometry/paired_contrasts.tsv and diagnostics/quantitative/paired_between_costs.tsv.

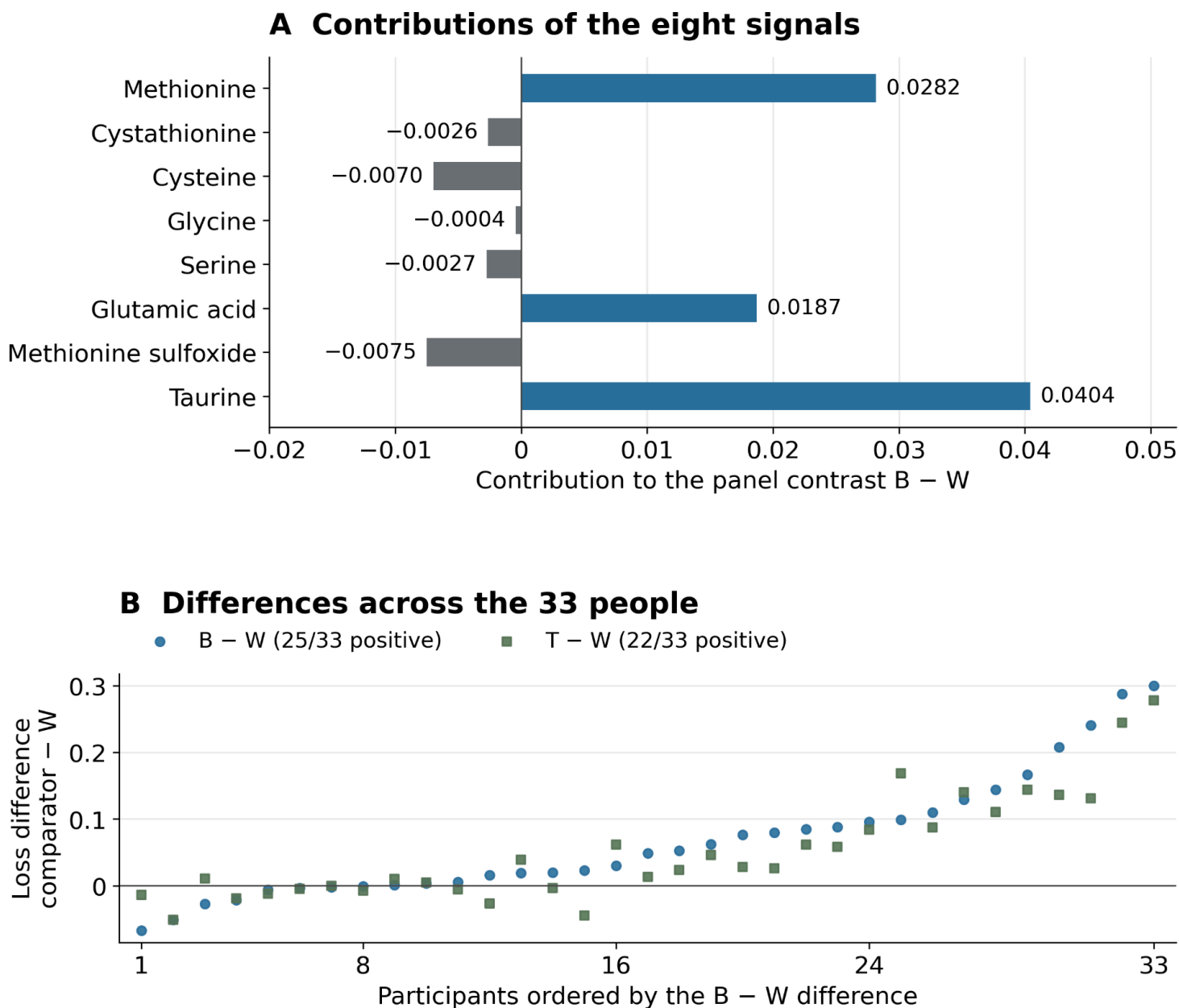


Figure 2

Distribution of reconstruction differences in Karu

A: contribution of each of the eight signals to the mean B-W contrast across the 33 people and 92 Karu samples. Each contribution is the difference in loss for that analyte divided by eight; their sum reproduces the panel contrast, 0.0670 after rounding. B: individual differences in B-W and T-W loss for the same 33 people, ordered by B-W. Positive values indicate lower loss with W; this ordering does not represent clinical time. Both panels are post hoc descriptions in squared standardised log units, with different aggregations. No signal-specific intervals or binomial tests were calculated: the fits for held-out people share training participants. All signals and people are retained. All five strata and both comparators are reported in the tables and TSV files of Supplementary Information S9. Sources:

derived/archived_descriptives/feature_decomposition_descriptive.tsv and person_differences_descriptive.tsv.

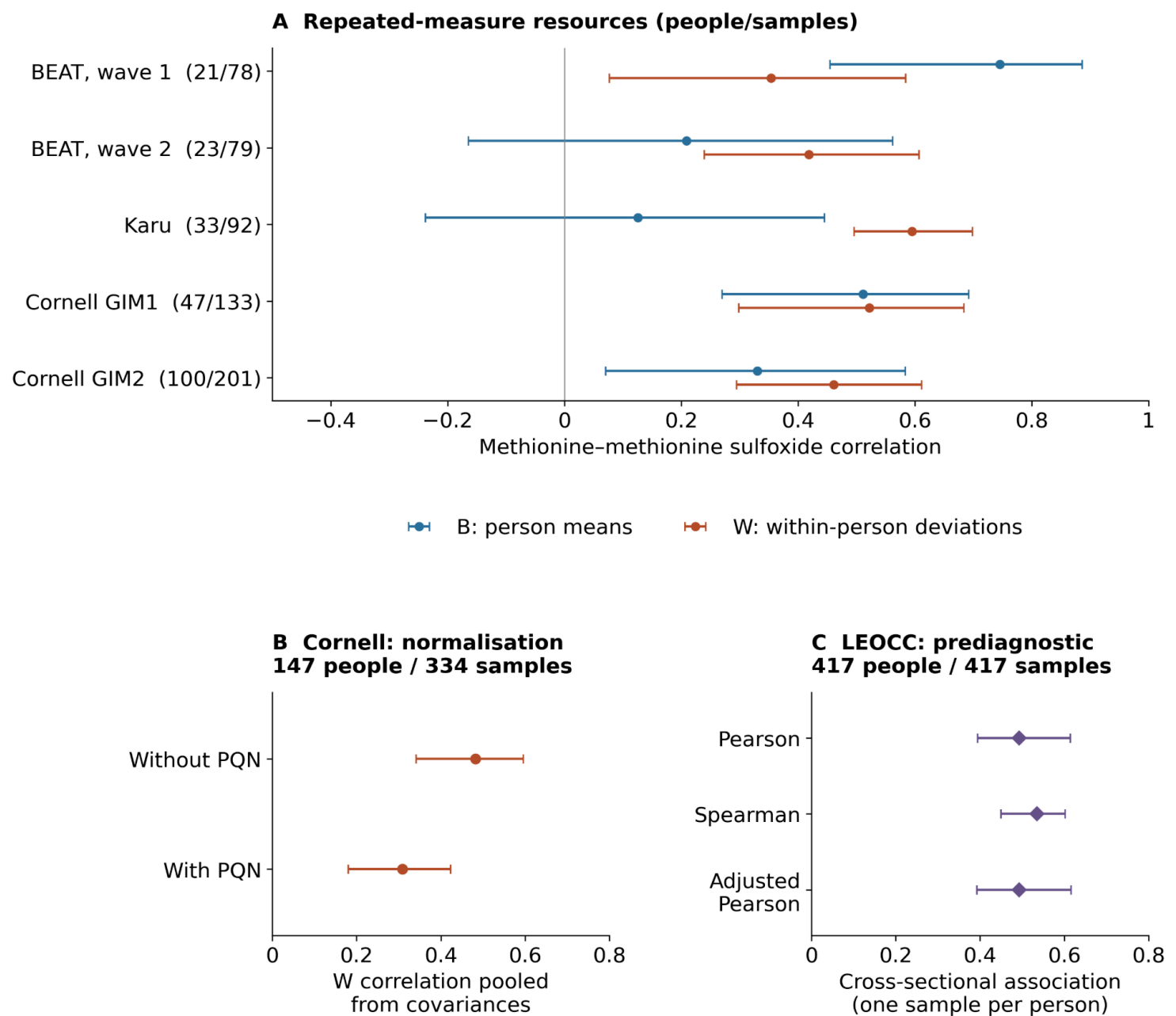


Figure 3

Methionine–methionine sulfoxide across contexts

A: B and W correlations in resources with repeated measurements; BEAT and Karu retain their historical populations, and Cornell uses the 147 people/334 samples specific to this pair. B: Cornell W estimates before and after PQN, combining batch-specific covariances and recalculating references within each resample. C: separate cross-sectional associations in 417 LEOCC participants, sampled a median of 1244 days before diagnosis; adjusted Pearson includes age, sex, body mass index and days relative to

diagnosis. Bars show the exploratory 2.5–97.5 percentiles. Similar values in A and C do not imply that the same parameter is being estimated. PQN is a sensitivity analysis in the same samples.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData1.zip](#)
- [SupplementaryInformation.pdf](#)
- [SupplementaryData2Figures.zip](#)