

Supplementary Information

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

Unless otherwise stated, bracketed limits are the exploratory 2.5th–97.5th percentiles of resampling by person. Nominal 95% coverage and simultaneous coverage have not been established. Tables S1–S10 retain the values from the integrated version; S11–S14 contain post hoc descriptions from the preceding revision; S15–S17 contain the post hoc descriptions added in the latest Spanish revision. Numbered references refer to the bibliography of the main article. Decimal points are used in both the English tables and the computational files.

S1. Sources, populations and molecular coverage

BEAT-COVID and Karu use plasma data from published supplementary materials [1,2]. The original full-panel selections comprised 27 people/84 visits and 45/101 in the two BEAT waves, and 44/103 in Karu. Requiring repeated measurements retained 21/78, 23/79 and 33/92, respectively. The original BEAT selection excluded the two ambiguous P41_d20 rows and visits with incomplete measurements for the thirty-signal panel. Reducing the panel did not reinstate these records. Time was retained as days since symptom onset in BEAT and days since hospitalisation in Karu.

Cornell uses serum signals measured by liquid chromatography–mass spectrometry (LC-MS) and annotations from Figshare deposit version 1, DOI 10.6084/m9.figshare.19115972.v1 [3]. The workbook contains 219 molecular rows and 714 measurement columns; 587 columns have documented COVID status. Removing seven GIM3 technical anchors leaves 580 columns/330 people, compared with the 582 samples reported in the publication. This discrepancy remains unresolved. The deposited values precede the authors' subsequent probabilistic quotient normalisation (PQN), logarithmic transformation and imputation; their R scripts were inspected but not executed.

Metadata exclusions removed four observations from person 86, who appeared in different batches and had an ambiguous day, and both samples from day 3 for person 362 because of discordant intubation indicators. The other eligible visits from person 362 were retained. This left 574 columns/329 people. Age of at least 18 years and documented absence of pregnancy were required; all documented COVID participants met these conditions. Unknown metadata were not inferred.

At most 50% missingness per sample across the 219 signals was allowed; no COVID sample exceeded this threshold. Each signal in the common eight-analyte panel (C08) had to have at most 25% missingness within its batch; all eight met this criterion. Positive, complete measurements were then required for each analysis, without imputation. Three panel measurements were missing: cystathionine and cysteine for person 21/day 1, and cysteine for person 154/day 2.

The primary Cornell population required the Boolean value False in the sample's already_intubated field: 480 samples/293 people remained before requiring completeness and repeated measurements. The label “No intubation flag at sampling” describes this filter; it neither presupposes later intubation nor uses its date. Each repeated-measures analysis required two distinct eligible days. GIM1 and GIM2 were fitted separately, and each retained person belonged to one batch. Exact sampling times were unavailable, and this is not considered an untreated population. The source procedures describe clinical serum usually stored at 4 °C for 1–5 days before storage at –80 °C; Karu describes EDTA plasma initially stored at –20 °C and subsequently at –80 °C [2,3]. Individual storage duration was not adjusted for.

Table S1. Primary Cornell populations and sensitivity analyses. Entries are people/samples. P05: methionine–methionine sulfoxide; P02: serine–cystathionine; P03: cystathionine–cysteine.

Population	Analysis	GIM1	GIM2	Total
No intubation flag at sampling	C08	47/131	100/201	147/332
No intubation flag at sampling	P05	47/133	100/201	147/334
No intubation flag at sampling	P02	47/132	100/201	147/333
No intubation flag at sampling	P03	47/131	100/201	147/332
Any intubation status	C08	61/172	131/263	192/435
Any intubation status	P05	61/174	131/263	192/437
Any intubation status	P02	61/173	131/263	192/436
Any intubation status	P03	61/172	131/263	192/435

LEOCC comprises Workbench analyses AN004619–AN004622 from ST002829, project PR001771, DOI 10.21228/M8SM6H [6,7]: 1108 deposited signals and 609 samples. Exact linkage to the authors' public mapping identified 461 samples from 461 people; 417 were prediagnostic, 25 were classified as collected during the episode and 19 afterwards. There were 148 unlinked samples. The fixed primary subset contains 417 people with one sample each, collected 42–3485 days before diagnosis, with a median of 1244 days. It is neither a healthy reference nor a longitudinal subset. The two LEOCC publications share source participants.

The LEOCC deposit reports normalisation, imputation, logarithmic transformation and Pareto scaling. That processing chain and the imputation mask were not independently reconstructed. No second logarithmic transformation or additional imputation was applied. Negative deposited values are compatible with that processing and were not treated as missing.

C08 contains methionine, cystathionine, cysteine, glycine, serine, glutamic acid, methionine sulfoxide and taurine. Exact names and acid/base equivalences were documented. The stereochemistry of Cornell methionine sulfoxide is unspecified. Homocysteine, glutathione and methionine sulfone were not replaced by related compounds: LEOCC lacks homocysteine/glutathione; Cornell annotates them but has no COVID measurements for them, and lacks methionine sulfone. EpOME–DiHOME pairs do not have complete external coverage. A combined 9-/13-HODE measurement does not replace separately measured isomers. C08 is a panel defined by coverage, not a complete functional pathway or an undisclosed modification of M02.

S2. Quantitative specification

Covariance with equal total weight per person

x_{ij} denotes the vector of log-transformed signals at visit j for person i , who has n_i visits. m_i is that person's observed mean; S_i is the covariance of deviations around that mean, with divisor n_i . The multiplicity c_i equals one in the ordinary analysis and the number of times the identity is selected in a bootstrap sample. C is the sum of the multiplicities.

$$\begin{aligned}
m_i &= \frac{1}{n_i} \sum_j x_{ij}, \quad S_i = \frac{1}{n_i} \sum_j (x_{ij} - m_i)(x_{ij} - m_i)^T, \\
C &= \sum_i c_i, \quad \mu = \frac{1}{C} \sum_i c_i m_i, \\
C_B &= \frac{1}{C} \sum_i c_i (m_i - \mu)(m_i - \mu)^T, \\
C_W &= \frac{1}{C} \sum_i c_i S_i, \quad C_T = C_B + C_W.
\end{aligned}$$

These definitions distribute each person's weight across their visits. C_B describes differences between observed personal means, and C_W describes deviations around those means; C_T is their empirical sum. They do not causally separate stable traits from disease states. C_W ignores visit order and does not remove common trends in clinical time. With two visits, S_i is one quarter of the outer product of the difference vector. It does not automatically equal the Pearson correlation between changes, because those changes are not centred across people. The divisor n_i describes the dispersion of retained visits; it does not provide an unbiased estimate of a latent individual variance. Equal weight does not imply equal precision or correction for informative follow-up.

Fixed rank and evaluation with entire people excluded

Each exclusion removed all visits from the evaluated person before learning the centre, scale or projection. Each signal was centred using its training mean and divided by its total standard deviation derived from C_T . B, W and T share this scale. Each model retains exactly two leading eigenvectors of its respective matrix. If Q contains the two orthonormal directions, $P=QQ^T$ is their projector. The number of components, grouping and decoder were not selected from the evaluated losses.

For the standardised vectors z_j of an excluded person, $\bar{z}$ is their observed mean. The W target is $z_j - \bar{z}$; the B target is $\bar{z}$ relative to the training centre, which is zero in these coordinates. With $p=8$:

$$\begin{aligned}
L_{W,i}(P) &= \frac{1}{n_i p} \sum_j \|(z_j - \bar{z})(I - P)\|^2, \\
L_{B,i}(P) &= \frac{1}{p} \|\bar{z}(I - P)\|^2, \\
L_W(P) &= \frac{1}{C} \sum_i c_i L_{W,i}(P), \quad L_B(P) = \frac{1}{C} \sum_i c_i L_{B,i}(P).
\end{aligned}$$

Projection uses the excluded person's same eight measured signals and their retrospective mean. It does not predict future measurements, missing metabolites or outcomes at admission. Losses average squared residuals over signals, visits and people, in squared standardised log units. They are neither entropy nor fractions of biological information. $L_W(B) - L_W(W)$ and $L_W(T) - L_W(W)$ evaluate the W target; a positive sign favours W. Absolute losses for both targets are also retained. Directions are relearned within each cohort; a fixed matrix is not transported between cohorts.

Principal angles describe differences between planes: their cosines are the singular values of the product of the transposed basis A and basis B; the maximum angle uses the smallest singular value. They are invariant to sign changes or rotations within the plane. A large angle does not establish better performance. Total variance had to exceed 10^{-12} , and the second eigenvalue and the second-third eigenvalue gap had to exceed 10^{-10} times the first eigenvalue. The rule stopped execution upon failure without replacing replicates; no failures were recorded.

Resampling and sensitivity analyses

Each stratum used 1000 multinomial resamples of people at its original sample size, with multiplicities archived. Models were relearned in every replicate and exclusion. If an identity appeared several times, all its copies were removed from training. Multiplicities weighted moments and losses. Contrasts were calculated within each replicate. Cornell combined losses using fixed weights of 47/147 and 100/147 after separate fits, on each batch's own standardised scale, without a common plane.

The 2.5th and 97.5th percentiles are exploratory pointwise limits; nominal or simultaneous coverage has not been established for pairs, strata, models or sensitivity analyses. No p values or confirmatory discoveries were generated. Omitting people assessed influence; the resulting ranges are not confidence intervals. Restricting to the first two visits, using logarithms without standardisation and allowing any intubation status addressed specific sensitivities. They change targets or populations without isolating a single cause of heterogeneity. Distinct training identities in the bootstrap are described in Table S13; their reduced number does not by itself establish that resampling is invalid.

Correlations and PQN

Correlations divide each covariance by the square root of the product of its marginal variances. In Cornell, C_W matrices were combined according to participant counts before standardisation. This is not the mean of individual or batch correlations; variances contribute to the association in the combined population. C_B was not pooled across unharmonised batches. A constant multiplicative factor for an analyte/batch cancels when logarithms are centred within a person in that batch; factors that vary between samples may remain.

PQN was a predefined sensitivity analysis using 211 candidate signals after excluding the eight panel signals. Each reference required at least 75% positive availability under person-based weighting. The lower weighted median was used with weight c_i/n_i . Integer weights derived from the least common multiple of visit counts resolved median and availability boundaries exactly. At least 20 reference signals were required. If m is the number of eligible reference signals, each sample required at least $\max(20, \text{ceiling of } m/2)$ valid quotients; its factor was the ordinary median of the positive quotients. Missing values were not imputed.

The logarithm of the factor was subtracted from the target logarithms. The molecular bootstrap recalculated eligibility, references and factors. PQN geometry learned references from training only, but the excluded sample supplied its broad panel to estimate its factor. This therefore does not establish the sufficiency of eight measurements. PQN geometry provides point estimates only, without an additional bootstrap. PQN may remove biological or technical variation; a shared divisor may introduce covariance. Attenuation does not separate these contributions.

LEOCC, decisions and seeds

Pearson correlation uses the deposited values. Spearman correlation uses average ranks, and adjusted Pearson correlation uses residuals from separate linear models containing an intercept, age at sampling, a female indicator, body mass index and days relative to diagnosis. All 417 people have the required covariates. Ranks and adjustments are recalculated in each resample; adjustment neither establishes causality nor removes unmeasured confounding.

The local specification preceded the new external fits and followed the earlier BEAT/Karu results; it was not a public preregistration. Geometry uses seed 2026092301 plus the stratum index, ordered as BEAT wave 1, wave 2, Karu, GIM1 and GIM2. Pairs use 2026092309 plus ten times the pair index plus the batch index, with pair order P05/P02/P03. LEOCC uses 2026092319. Multiplicities and 1000 replicates per analysis are retained. No models of clinical outcomes, new proteomics or topology were fitted.

S3. Complete representation results

Table S2. Absolute C08 losses. Model indicates the source of the directions; the two losses correspond to different targets. Units are squared standardised logarithms. Do not subtract marginal interval endpoints to obtain an interval for a difference.

Stratum	Model	Within-person loss	Personal-mean loss
BEAT wave 1	B	0.2388 [0.2115; 0.3935]	0.3139 [0.2441; 0.5745]
BEAT wave 1	W	0.1980 [0.1620; 0.3157]	0.2807 [0.2142; 0.5139]
BEAT wave 1	T	0.1967 [0.1663; 0.2887]	0.2763 [0.2182; 0.4867]
BEAT wave 2	B	0.1594 [0.1323; 0.2800]	0.3559 [0.3031; 0.7068]
BEAT wave 2	W	0.1596 [0.1122; 0.2512]	0.4639 [0.3713; 0.7320]
BEAT wave 2	T	0.1482 [0.1204; 0.2388]	0.3461 [0.2892; 0.6633]
Karu	B	0.1877 [0.1607; 0.3025]	0.3335 [0.2660; 0.5866]
Karu	W	0.1207 [0.1071; 0.1744]	0.4688 [0.4326; 0.6555]
Karu	T	0.1728 [0.1495; 0.2323]	0.3237 [0.2562; 0.5580]
Cornell GIM1	B	0.1920 [0.1662; 0.2609]	0.4263 [0.3754; 0.6573]
Cornell GIM1	W	0.1777 [0.1413; 0.2463]	0.4811 [0.4160; 0.6231]
Cornell GIM1	T	0.1845 [0.1612; 0.2395]	0.4175 [0.3726; 0.6316]
Cornell GIM2	B	0.1339 [0.1169; 0.1737]	0.3285 [0.2995; 0.4285]
Cornell GIM2	W	0.1175 [0.1010; 0.1586]	0.4108 [0.3680; 0.5397]
Cornell GIM2	T	0.1277 [0.1121; 0.1590]	0.3242 [0.2880; 0.4268]

Table S3. Paired contrasts in within-person loss. A positive value favours W; combined Cornell is not another cohort.

Stratum	B minus W	T minus W
BEAT wave 1	0.0408 [−0.0355; 0.1652]	−0.0013 [−0.0518; 0.0422]
BEAT wave 2	−0.0002 [−0.0290; 0.0949]	−0.0114 [−0.0310; 0.0313]
Karu	0.0670 [0.0347; 0.1646]	0.0522 [0.0220; 0.0832]
Cornell GIM1	0.0143 [−0.0364; 0.0742]	0.0068 [−0.0369; 0.0495]
Cornell GIM2	0.0164 [−0.0127; 0.0403]	0.0102 [−0.0159; 0.0256]
Cornell combined	0.0157 [−0.0091; 0.0409]	0.0091 [−0.0124; 0.0261]

Both Karu contrasts have entirely positive percentile limits; the BEAT and Cornell contrasts establish neither a clear advantage nor equivalence. The higher point-estimate loss for Karu personal means requires its own paired contrast, added in Table S11, whose interval includes zero. In BEAT wave 1, W has lower point-estimate loss than B for both targets. Maximum B–W angles were 31.9°, 52.4°, 80.6°, 83.3° and 47.1° in stratum order. Losses for the centre-only reference remain in model_summary.tsv.

Table S4. Predefined representation sensitivity analyses. Point-estimate differences; logarithms without standardisation use different units. PQN requires a broad panel and has no additional bootstrap.

Stratum	Sensitivity analysis	People / samples	B minus W	T minus W
BEAT wave 1	First two visits	21/42	0.0134	0.0055
BEAT wave 1	Log without standardisation	21/78	−0.0073	−0.0084
BEAT wave 2	First two visits	23/46	0.0023	0.0005
BEAT wave 2	Log without standardisation	23/79	0.0063	0.0064
Karu	First two visits	33/66	0.0566	0.0437
Karu	Log without standardisation	33/92	0.0146	0.0007
Cornell GIM1	First two visits	47/94	0.0284	0.0208
Cornell GIM1	Log without standardisation	47/131	0.0050	0.0054
Cornell GIM1	Any intubation status	61/172	0.0140	0.0088
Cornell GIM2	First two visits	100/200	0.0161	0.0102
Cornell GIM2	Log without standardisation	100/201	0.0012	0.0006
Cornell GIM2	Any intubation status	131/263	0.0124	0.0079
Cornell GIM1	PQN with training reference	47/131	0.0449	0.0293
Cornell GIM2	PQN with training reference	100/201	0.0179	0.0097

S4. Prioritised molecular relationships and sensitivity analyses

Table S5. Cornell pair-specific estimates before PQN. B is reported by batch. A dash denotes the deliberately undefined combined B estimate; combined W is derived from weighted covariances.

Pair	Stratum	People / samples	Between-person r	Within-person r
P05	GIM1	47/133	0.511 [0.270; 0.692]	0.522 [0.299; 0.684]
P05	GIM2	100/201	0.330 [0.070; 0.584]	0.461 [0.295; 0.611]
P05	Cornell combined	147/334	—	0.482 [0.341; 0.596]
P02	GIM1	47/132	−0.176 [−0.426; 0.087]	0.182 [0.018; 0.336]
P02	GIM2	100/201	−0.080 [−0.318; 0.200]	0.192 [0.002; 0.362]
P02	Cornell combined	147/333	—	0.187 [0.056; 0.319]
P03	GIM1	47/131	0.325 [0.096; 0.545]	0.147 [−0.110; 0.425]
P03	GIM2	100/201	0.493 [0.187; 0.675]	0.129 [−0.081; 0.320]
P03	Cornell combined	147/332	—	0.136 [−0.036; 0.295]

Table S6. P05 after PQN. References and factors are relearned in each replicate. Eligible reference signals: GIM1, 127–136; GIM2, 124–129. Minimum available quotients per sample: 122 and 119, respectively. All replicates met the requirements.

Stratum	People / samples	Normalised within-person r
GIM1	47/133	0.317 [0.120; 0.488]
GIM2	100/201	0.307 [0.154; 0.470]
Cornell combined	147/334	0.309 [0.180; 0.423]

Table S7. Combined Cornell sensitivity analyses. Two visits: 147 people/294 samples. Any intubation status: 192 people with 437/436/435 samples for P05/P02/P03. The omission range describes influence.

Pair	First two visits	Any intubation status	Individual-omission range
P05	0.481	0.503	0.452 to 0.501
P02	0.105	0.148	0.163 to 0.206
P03	0.150	0.153	0.115 to 0.188

GIM1 P02 changed from +0.182 to −0.119 with two visits, despite a positive combined primary point estimate. P03 remained uncertain. The stability observed for P05 is not attributed to these pairs.

Table S8. Cross-sectional LEOCC context, 417 people/417 samples. These coefficients estimate neither B nor W. All sensitivity analyses retain the same people.

Pair	Pearson	Spearman	Adjusted Pearson
P05	0.493 [0.394; 0.614]	0.536 [0.450; 0.602]	0.493 [0.393; 0.617]
P02	−0.080 [−0.177; 0.033]	−0.063 [−0.155; 0.040]	−0.013 [−0.120; 0.101]
P03	0.146 [0.022; 0.266]	0.140 [0.038; 0.241]	0.059 [−0.062; 0.181]

The new catalogue includes 28 C08 pairs in each geometry stratum and 28 Pearson pairs in LEOCC. Complete estimates, percentiles and omission results remain in computational tables. Presenting the catalogue does not create 168 independent confirmations or a functional network.

S5. Historical context: panels and targets that differ from C08

This section documents earlier analyses retained for traceability. Their results are neither new C08 fits nor additional validations in independent participants. The panel, population or number of directions differs by analysis; their magnitudes therefore cannot be compared directly with those in S3.

The earlier multilevel analysis used thirty BEAT signals divided into M02 (11), M07 (7) and M15 (12), with two directions per group; Karu retained M02. It retained the same repeated-measures populations but used targets and a dimensional budget different from C08. M02 adds homocysteine, glutathione and methionine sulfone. M07 includes linoleate, 9,10-/12,13-EpOME, their DiHOMEs and separate 9-/13-HODE signals. M15 includes LPE 16:0, 16:1, 18:0, 18:1, 18:2, 18:3, 20:3, 20:4, 20:5, 22:4, 22:5 and 22:6. Amines and lipids come from different assays; association between sets does not establish biological coupling.

Table S9. Historical full-panel contrasts. BEAT weights the sets by 11/30, 7/30 and 12/30; Karu uses eleven signals. These results do not replace C08 in S3.

Historical resource	Signals	B minus W	T minus W
BEAT wave 1	30	0.0363 [−0.0144; 0.0634]	0.0105 [−0.0209; 0.0362]
BEAT wave 2	30	0.0090 [−0.0078; 0.0336]	0.0010 [−0.0133; 0.0146]
Karu	11	0.0673 [0.0397; 0.1624]	0.0477 [0.0202; 0.0871]

Table S10. The twelve historically prioritised pairs, W correlations. Met: methionine; Cth: cystathionine; Cys: cysteine; MetSO: methionine sulfoxide; GSH: glutathione; LA: linoleate. Chemical positions are retained. NA denotes absent coverage, not zero correlation. Historical percentiles use their own replicates.

Pair	Measured relationship	BEAT wave 1	BEAT wave 2	Karu
P01	Homocysteine–Cth	0.463 [0.213; 0.670]	0.396 [0.038; 0.657]	0.140 [−0.132; 0.390]
P02	Ser–Cth	−0.059 [−0.389; 0.339]	0.254 [−0.116; 0.572]	−0.066 [−0.294; 0.177]

Pair	Measured relationship	BEAT wave 1	BEAT wave 2	Karu
P03	Cth–Cys	0.212 [–0.023; 0.429]	0.291 [0.071; 0.550]	–0.168 [–0.384; 0.047]
P04	Cys–GSH	0.171 [–0.118; 0.443]	–0.016 [–0.184; 0.215]	–0.316 [–0.520; –0.075]
P05	Met–MetSO	0.354 [0.062; 0.585]	0.418 [0.248; 0.611]	0.595 [0.496; 0.701]
P06	LA–9,10-EpOME	0.708 [0.546; 0.821]	0.461 [0.205; 0.646]	NA
P07	LA–12,13-EpOME	0.392 [0.192; 0.567]	0.225 [–0.024; 0.445]	NA
P08	9,10-EpOME–9,10-DiHOME	0.866 [0.799; 0.911]	0.825 [0.670; 0.914]	NA
P09	12,13-EpOME–12,13-DiHOME	0.895 [0.803; 0.943]	0.915 [0.802; 0.961]	NA
P10	9-HODE–13-HODE	0.889 [0.800; 0.934]	0.951 [0.891; 0.979]	NA
P11	MetSO–9-HODE	0.314 [0.158; 0.486]	0.156 [–0.091; 0.495]	NA
P12	LA–LPE 18:2	0.741 [0.493; 0.880]	0.214 [–0.203; 0.477]	NA

Epoxide–diol pairs were positive in BEAT, but their intervals for differences between levels included zero. The negative W association between cysteine and glutathione in Karu was not clearly supported in both BEAT waves. These contrasts lack complete external coverage. The between-set covariance reconstruction in the historical archive is a descriptive in-sample norm, not validated regulation, synergy or physiological information.

The earlier component-allocation study used all eligible people, including those with one visit: BEAT 27/84 and 45/101, and Karu 44/103. It evaluated contemporaneous reconstruction, not B/W decomposition. Uniform allocation of 2/2/2 produced relative error reductions of 72.39%/68.85% in BEAT; adaptive allocation produced 78.70%/69.64%; global PCA with six components produced 79.33%/70.05%. In Karu, the mean of eleven signals, the mean plus a contrast, and two unrestricted components produced 24.18%, 46.60% and 48.30%, respectively. Each percentage is one hundred times one minus the ratio of model loss to loss from the centre-only reference. It is not explained biological variance, Shannon information or clinical prediction. The local sign of the gain from adding components depends on nested projections; the chosen allocations and aggregate magnitudes are empirical. No additional inferential claim is drawn from these results.

The preserved scientific package contains the 89 files of the historical multilevel analysis and its manifest. It also preserves a partial component-allocation editorial folder and evidence/previous_allocation_complete.zip, containing the 1221 original scientific files, manifest and final verification. Hashes of the preserved files were checked; preservation does not mean these analyses were executed again.

S6. Numerical verification and its scope

Preserved checks from the scientific version

The original scientific package includes a rerun that reproduced 26 files byte for byte and 114 stored arrays exactly, with a maximum difference of zero, while leaving 93 pre-fit files intact. Equality of receipt timestamps or NPZ container files was not required. Its independent implementation using explicit observation rows and singular value decomposition (SVD) performed 2092 checks and 80076 scalar comparisons: all primary geometry fits, nine complete bootstrap replicates, molecular point estimates and 18 molecular

replicates, six P05 PQN replicates and six PQN geometry exclusions. Maximum differences were 4.57×10^{-14} for losses and 4.95×10^{-14} for projectors; PQN references and factors matched exactly. The entire geometry bootstrap was not independently refitted.

The separate LEOCC verification used weighted covariance, ranks with ties, QR residualisation and percentile interpolation. It checked all 1000 replicates and 417 omissions, with 50383 numerical checks and a maximum difference of 4.97×10^{-14} . The Cornell audit compared 264698 source values and confirmed 4640 target-cell rows. The portable reconstruction checked 5328 cells and reproduced four inputs byte for byte. These counts describe preserved receipts; they are not presented as executions carried out during this textual revision.

Verification of the first Spanish version

The earlier Spanish revision recalculated 224 primary exclusions using SVD without importing the original modules: three models and two losses per exclusion, giving 1344 scalar comparisons. Its maximum loss difference was 2.50×10^{-14} . It compared 23 TSV files and 114 stored arrays from the two stored runs; this scope differs from the 26-file receipt for the complete reproduction. The eight derived TSV files were reproduced byte for byte from a portable extraction of 64 files. The initial and portable receipts belong to the first Spanish archive identified in S8. That historical archive is not embedded in Supplementary Data 1; the later audit described below supplies the corresponding checks in the current delivery.

Subsequent audit on 23 September 2026

The subsequent editorial audit reran the complete analysis from a fresh extraction of the scientific package. It refitted the 1000 replicates specified per analysis, the exclusions and the sensitivity analyses using the original implementation: 26 scientific files were identical byte for byte and 114 stored arrays were exactly equal, with a maximum difference of zero. It also confirmed the integrity of the 560 inventoried scientific files and the 93 files frozen before fitting.

A separate SVD check refitted the primary geometry point estimates without importing the original modules: 224 person exclusions, 672 models and 1344 losses, with a maximum difference of 2.498×10^{-14} . The independent inspection of distributions used the saved replicates; it does not constitute a second independent refit of the entire bootstrap. The eight Spanish diagnostic TSV files were also reproduced byte for byte.

The audit directly checked 1992 BEAT/Karu values against their published, archived Excel workbooks, together with sample, person and day identities. For BEAT, it retained the historical eligibility rules for the thirty-signal panel without readjudicating the two excluded ambiguous rows. For Karu, it reconstructed the repeated-measures selection from Table S4. The portable procedure rechecked 5328 contextual cells, and Cornell preparation produced 56 identical files and the 264698 numerical comparisons and 4640 target-cell records described above. These counts overlap; they must not be added as independent biological observations.

Scope and commands are documented in `provenance/auditoria_20260923/reproducibility/AUDIT_COMPLETE_RECEIPT.json` and `FRESH_EXECUTION_RECEIPT.json`; the Excel comparison is in `DIRECT_WORKBOOK_RECEIPT.json` and `DIRECT_WORKBOOK_CELL_CHECKS.tsv` in the same folder. These checks were performed by AI agents and locally executed code; they are not human peer review. They start from processed public measurements and do not reproduce instrumental acquisition or the original LEOCC imputation. Local reproduction also does not establish inferential coverage, operation on every operating system or functional validation. Historical receipts, the subsequent audit and checks of the current delivery are identified separately in `README_EN.md` in Supplementary Data 1. Scientific paths in S1–S7 refer to its science directory; the later descriptions and their specifications are under editorial.

S7. Descriptions added after analysis

Tables S11–S14 derive from saved inputs and results; the bootstrap was not refitted and the frozen rules were not changed. The subsequent specification is preserved in `diagnostics/SPECIFICATION_QUANTITATIVE_POSTHOC.md`. All strata and all eight analytes are reported without selection by result.

Table S11. Paired cost for personal means. $L_B(W) - L_B(B)$. Differences were calculated within the same replicates; limits are percentiles of the distribution of those differences. This is a post hoc summary with uncalibrated coverage. A positive point estimate indicates higher loss with W.

Stratum	People	Difference [2.5th; 97.5th percentiles]
BEAT wave 1	21	−0.0333 [−0.1971; 0.1151]
BEAT wave 2	23	0.1080 [−0.1681; 0.3213]
Karu	33	0.1353 [−0.0106; 0.2315]
Cornell GIM1	47	0.0548 [−0.1276; 0.1463]
Cornell GIM2	100	0.0823 [0.0006; 0.2043]
Cornell combined	147	0.0735 [−0.0150; 0.1501]

The Karu cost is positive as a point estimate, but its interval includes zero. A universal trade-off between levels is not established. The positive lower limit for GIM2 is not selected as a new confirmatory test; the multiplicity and coverage limitations remain.

Table S12. Number of visits and first-to-last separation. A post hoc description of the inputs, not a temporal model. The interval uses each resource's recorded time scale without equating their time origins.

Stratum	Visits: number of people	Separation in days: minimum; median; maximum
BEAT wave 1	2: 5; 3: 2; 4: 8; 5: 6	3; 10; 14
BEAT wave 2	2: 11; 3: 3; 4: 5; 5: 1; 6: 1; 8: 1; 9: 1	3; 7; 39
Karu	2: 17; 3: 9; 4: 5; 5: 1; 6: 1	2; 6; 19
Cornell GIM1	2: 10; 3: 37	1; 2; 2
Cornell GIM2	2: 99; 3: 1	1; 1; 2

Table S13. Distinct training identities in the bootstrap. The original fit uses $N-1$; each replicate uses the number of distinct identities minus the evaluated identity. This describes the observed training support, not bias or coverage.

Stratum	Original training	Bootstrap: minimum; median; maximum
BEAT wave 1	20	8; 12; 19
BEAT wave 2	22	9; 14; 19

Stratum	Original training	Bootstrap: minimum; median; maximum
Karu	32	14; 20; 26
Cornell GIM1	46	22; 29; 36
Cornell GIM2	99	53; 63; 72

Bootstrap medians of losses and contrasts are also preserved in `bootstrap_medians.tsv`. For example, in Karu, $L_{\mathcal{W}}(\mathbf{B})$ has point estimate 0.1877 and bootstrap median 0.2180; $\mathbf{W}$ has 0.1207 and 0.1334; $\mathbf{B}-\mathbf{W}$ has 0.0670 and 0.0802. These shifts do not by themselves establish bias or invalidity of the bootstrap. A specific inferential calibration is needed if population coverage is intended.

Table S14. Complete breakdown of Karu within-person losses by analyte. Means over the 33 people. The contribution to the panel is the analyte's loss difference divided by eight. It has no analyte-specific intervals or causal meaning; it is not the effect of removing that molecule.

Analyte	Loss with $\mathbf{B}$	Loss with $\mathbf{W}$	Contribution to the C08 contrast
Methionine	0.2698	0.0445	0.0281685
Cystathionine	0.0711	0.0921	-0.0026256
Cysteine	0.1970	0.2529	-0.0069777
Glycine	0.0964	0.0998	-0.0004332
Serine	0.0721	0.0941	-0.0027468
Glutamic acid	0.2518	0.1022	0.0187026
Methionine sulfoxide	0.1100	0.1700	-0.0074939
Taurine	0.4333	0.1099	0.0404202

The unrounded sum is 0.06701396259681974. Reconstruction improves for taurine, methionine and glutamic acid; it worsens for the other five signals. Methionine and methionine sulfoxide have opposite contributions despite their positive correlation. The breakdown locates loss and does not identify mechanisms.

`P05_pooling_contributions.tsv` additionally documents GIM1 contributions to the combined variance/covariance, which differ from its fraction of participants; it does not change the estimand.

S8. Reproduction and file guide

Supplementary Data 1 contains the journal-sized reproducibility package, rooted at `MERITM_REPRODUCIBILITY`. Supplementary Data 2 contains the English figure scripts, their traced numerical inputs, reference images and separate reproduction instructions. `README_EN.md` is its entry point. The package retains the frozen analysis settings, required processed inputs, reference results, source mappings, audit records and executable code. Duplicate runs, earlier manuscripts and large historical archives are omitted. This curated delivery is not the complete historical archive: `CURATED_MANIFEST.json` identifies its own files, while the original scientific manifest is retained as a historical inventory.

Nine required third-party files are retrieved from public source URLs with SHA-256 verification, as specified in `THIRD_PARTY_ACQUISITION.json`. The original BEAT Excel workbook is omitted; its source-cell comparison is retained as a historical audit, not rerun by this wrapper. Initial acquisition requires network access. The analysis then runs from local files in a new output directory. Neither the original analysis lock nor its 93 expected file hashes is changed. A missing source or a hash mismatch is an error, not permission to substitute a newer version. The redistribution terms of third-party sources remain distinct from the licence of this manuscript.

Extract Supplementary Data 1 into a new, short directory with write permission. From `MERITM_REPRODUCIBILITY`, run the command below with a new output destination. It acquires any explicitly listed missing sources, reconstructs inputs, reruns the scientific analyses and reproduces the subsequent descriptions. It does not overwrite source files or reference results.

```
python -B code/reproduce_submission.py --acquire-missing --out ../replay_new
```

The recorded scientific environment is Python 3.10.20, NumPy 2.2.6, pandas 2.3.3 and openpyxl 3.1.5; figure production additionally uses Matplotlib 3.10.8. These versions document the tested environment and do not certify numerical or binary identity across all systems, BLAS libraries or document engines. PDF pagination is not a scientific invariant.

Complete reproduction rebuilds portable contextual inputs and reruns geometry, molecular pairs and LEOCC. BEAT/Karu start from the preserved repeated-measures inputs; their comparison against published Excel workbooks is documented separately in S6. LEOCC is rebuilt from Workbench text and source mapping/metadata workbooks. Cornell preparation runs in a fresh copy because its script writes beside the inputs. The wrapper compares the regenerated results with the preserved reference outputs and records the exact scope of each check.

The post hoc summaries in S7 and S9 are produced from the archived losses and fitted objects. They are kept separate from the frozen scientific analysis, retain their retrospective specifications and do not alter the original fits, selection rules or confirmatory status. Tables S15–S17 describe existing results and do not refit models. A separate audit of the primary fits uses explicit observations and singular value decomposition. Neither code execution nor numerical agreement establishes population coverage or biological function.

TSV files have a header and tab separators; identifiers should be read as strings. NPZ files do not use pickled objects. Metadata define the order of signals, people, models and targets. B/W/T models and W/B targets are not interchangeable. Full-precision values take precedence over rounded tables.

Path within <code>MERITM_REPRODUCIBILITY</code>	Content
<code>README_EN.md</code> and <code>code/reproduce_submission.py</code>	Reproduction instructions and execution wrapper.
<code>CURATED_MANIFEST.json</code> and <code>THIRD_PARTY_ACQUISITION.json</code>	Included-file integrity and hash-verified acquisition of specified public sources.
<code>science/inputs/*.tsv</code>	Person, sample, day and eight signals; LEOCC adds covariates and deposited values.
<code>science/cornell_inputs/analysis_membership.tsv</code>	Selections by population, analysis and visit.
<code>science/cornell_inputs/chemical_mapping.tsv</code> and <code>target_cell_trace.tsv</code>	Annotations, mappings and source-cell coordinates.
<code>science/results/run_01/geometry</code>	Absolute, person-level and analyte-level losses; contrasts; sensitivities; covariance; complete pair catalogue; bootstrap and fitted objects.

Path within MERITM_REPRODUCIBILITY	Content
science/results/run_01/molecular	Pair-specific W before/after PQN, B by batch, multiplicities, references and factors.
science/results/run_01/LEOCC	Cross-sectional estimates, complete pair catalogue, participant order, bootstrap associations and omissions.
science/traceability/ANALYSIS_LOCK.json	Original sealing of 93 files before fitting.
science/quant_review and science/verification_leocc	Preserved separate numerical checks with their stated scope.
editorial/diagnostics/quantitative	Eight diagnostic TSV files supporting S7.
editorial/derived/archived_descriptives	Seven descriptive TSV files supporting S9 and Figure 2.
editorial/code and editorial/traceability	Programs and retrospective specifications for the added descriptions.
provenance/auditoria_20260923/reproducibility	Historical rerun and direct source-workbook audit records.

The complete local scientific archive and the first Spanish archive are identified below for provenance; neither large archive is embedded in Supplementary Data 1. Their hashes identify those historical objects and are not the checksum of the current supplement.

Historical archive not embedded in this supplement	SHA-256
PAQUETE_CIENTIFICO_ORIGINAL.zip	1905a11aad4d11e40f86942abcb425a7b62a03b46e5cd80cb99d7ffc6783cea2
PAQUETE_ES_ORIGINAL.zip	daa9d83dcbb33cf25b7f1a68f38954913cc9a84db3c5ced4f2c44f316b084774

The current wrapper's output receipt distinguishes source integrity, numerical reproduction and the subsequent diagnostic checks. The retained historical receipts document earlier executions; they are not attributed to the current run. Reproduction does not resolve the Cornell discrepancy, reconstruct undocumented upstream processing or establish the functional meaning of covariation.

S9. Person-level distribution and relative sensitivity described after analysis

This section addresses the technical note received after the Spanish revision. Its results were known when traceability/SPEC_DESCRIPTIVOS_POSTHOC_ES.md was written; that specification is therefore not a preregistration. Only the frozen fold_losses.tsv, feature_losses.tsv, model_summary.tsv, sensitivities.tsv and jackknife.tsv tables were summarised. There were no new fits, resampling, participant selection or inferential tests. All five strata and both comparators, B and T against W, were retained.

Table S15. Person-level distribution of within-person differences. Positive values favour W. Evaluated identities share training participants; the counts are descriptive. A majority count does not guarantee a positive mean.

Stratum	Comparator	Lower error with W / people	Mean difference	Median difference
BEAT wave 1	B	15 / 21	0.0408	0.0428
BEAT wave 1	T	14 / 21	−0.0013	0.0074
BEAT wave 2	B	16 / 23	−0.0002	0.0074
BEAT wave 2	T	14 / 23	−0.0114	0.0037
Karu	B	25 / 33	0.0670	0.0489
Karu	T	22 / 33	0.0522	0.0263
Cornell GIM1	B	29 / 47	0.0143	0.0122
Cornell GIM1	T	24 / 47	0.0068	0.0086
Cornell GIM2	B	69 / 100	0.0164	0.0147
Cornell GIM2	T	63 / 100	0.0102	0.0116

Differences subtract W loss from comparator loss, in squared standardised log units. There were no exact ties in these tables. The favoured majority and the mean can disagree: BEAT wave 2 favours W over B in 16/23 people, with a mean of −0.0002. Every individual loss is retained in person_differences_descriptive.tsv.

Table S16. Losses and relative reduction within each scale. The percentage is $100 \times (\text{comparator loss} - \text{W loss}) / \text{comparator loss}$. E: standardised logarithms; NE: logarithms without standardisation. Absolute losses are not compared across scales, and no intervals were estimated for the percentages.

Stratum	Scale	Comparator	Comparator loss	W loss	Relative reduction %
BEAT wave 1	E	B	0.23880	0.19799	17.1
BEAT wave 1	NE	B	0.03352	0.04078	−21.6
BEAT wave 1	E	T	0.19672	0.19799	−0.6
BEAT wave 1	NE	T	0.03243	0.04078	−25.8
BEAT wave 2	E	B	0.15935	0.15959	−0.2
BEAT wave 2	NE	B	0.03262	0.02631	19.4
BEAT wave 2	E	T	0.14820	0.15959	−7.7
BEAT wave 2	NE	T	0.03270	0.02631	19.5
Karu	E	B	0.18769	0.12068	35.7

Stratum	Scale	Comparator	Comparator loss	W loss	Relative reduction %
Karu	NE	B	0.03553	0.02094	41.1
Karu	E	T	0.17283	0.12068	30.2
Karu	NE	T	0.02165	0.02094	3.3
Cornell GIM1	E	B	0.19202	0.17768	7.5
Cornell GIM1	NE	B	0.03420	0.02920	14.6
Cornell GIM1	E	T	0.18449	0.17768	3.7
Cornell GIM1	NE	T	0.03462	0.02920	15.7
Cornell GIM2	E	B	0.13387	0.11747	12.2
Cornell GIM2	NE	B	0.02327	0.02210	5.0
Cornell GIM2	E	T	0.12765	0.11747	8.0
Cornell GIM2	NE	T	0.02274	0.02210	2.8

E losses are in squared standardised log units; NE losses are in squared log units. A negative percentage indicates higher W loss for that target. Full values and denominators are in `relative_reduction_by_scale.tsv`. The percentages are not fractions of biological information and do not justify choosing the scale according to the result.

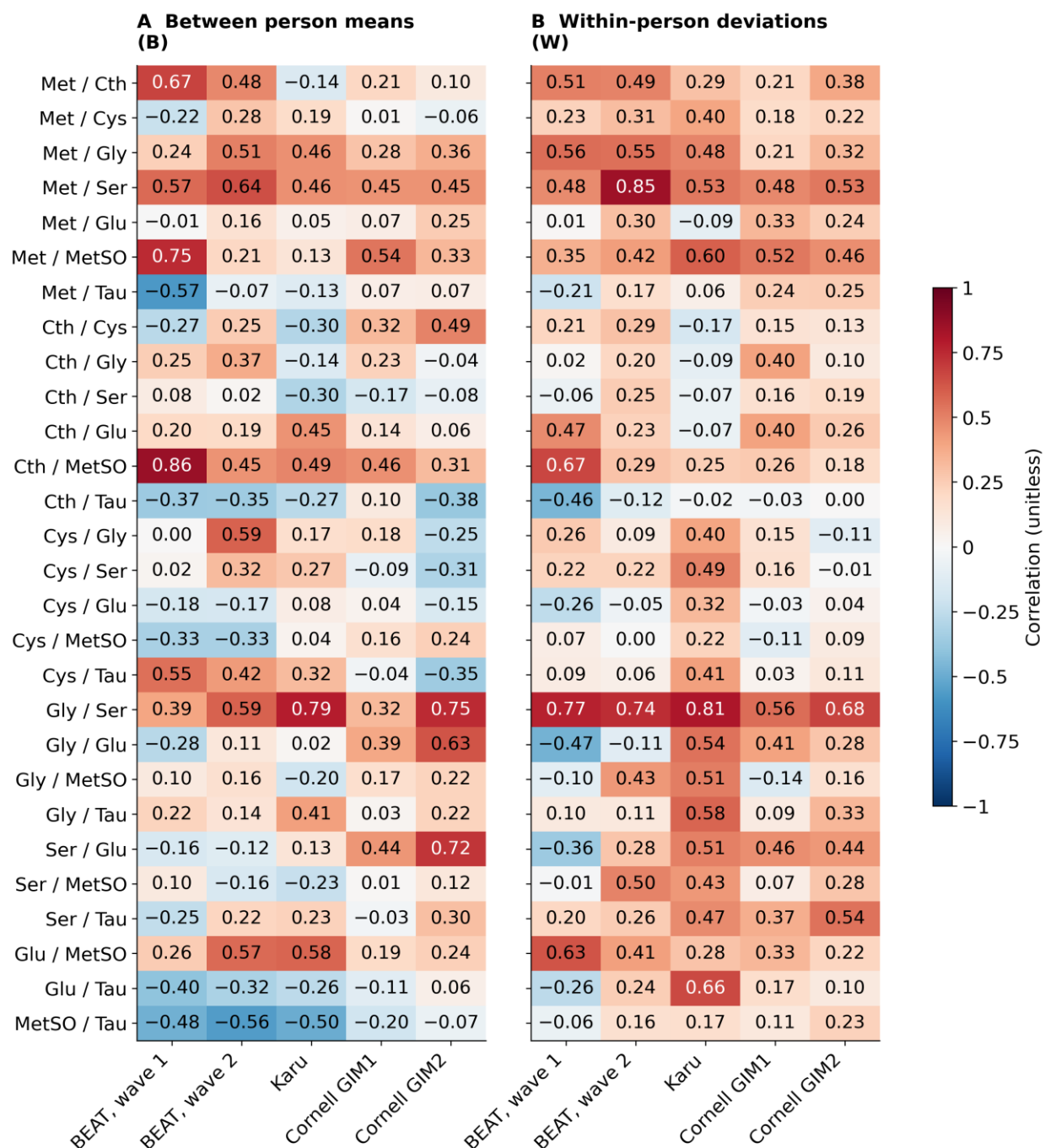
Table S17. Range of mean contrasts after omitting one person from the dataset and refitting the procedure. Summary of archived jackknife results; these are neither the extremes of individual-person losses nor confidence intervals.

Stratum	Comparator	Minimum	Maximum
BEAT wave 1	B	0.0259	0.0690
BEAT wave 1	T	-0.0059	0.0105
BEAT wave 2	B	-0.0105	0.0188
BEAT wave 2	T	-0.0187	0.0010
Karu	B	0.0589	0.0802
Karu	T	0.0460	0.0589
Cornell GIM1	B	0.0037	0.0333
Cornell GIM1	T	-0.0031	0.0251
Cornell GIM2	B	0.0136	0.0228
Cornell GIM2	T	0.0078	0.0148

Ranges are in squared standardised log units. In Karu, B–W remains between 0.0589 and 0.0802 after omitting one person from the dataset. This constrains an explanation based exclusively on a single retained person, without ruling out systematic biases or estimating population coverage.

The figures in the note were checked locally against the original tables. The external review script and TSV files mentioned in that note were not among the files received for the earlier audit. The independent execution of 672 models reported by the note is not attributed to this check. The local scope was to reproduce archived descriptions and check their arithmetic; the historical receipts and the current reproduction receipt identify these separate executions.

Figure S1. Complete catalogue of molecular relationships in C08



All 28 unordered pairs are shown by stratum (people/samples: BEAT wave 1, 21/78; wave 2, 23/79; Karu, 33/92; GIM1, 47/131; GIM2, 100/201): A, correlation between personal means; B, correlation of within-person deviations with equal total weight per person. Colours and numbers are descriptive estimates without a significance threshold. Intervals and omission diagnostics for every cell are retained in geometry/pair_summary.tsv in the scientific package. Cornell uses 332 C08 samples here; P05 in Figure 3 uses 334. Met: methionine; Cth: cystathionine; Cys: cysteine; Gly: glycine; Ser: serine; Glu: glutamic acid; MetSO: methionine sulfoxide; Tau: taurine. Associations establish neither reaction direction nor function.