

Supporting Information

Deterministic Verification of Audit Outputs for Downstream Metabolomics Result Tables on a Frozen Synthetic Benchmark

This supplement accompanies the frozen Validex 0.2.0 verification study. It provides reproduction instructions, additional results and the corpus-screening record. Machine-readable files in `supporting_data` retain full records where a Word table uses selected fields. All revision analyses are post hoc; the primary held-out result remains 147/160 exact and 13/160 mismatched.

S1 Reproduction and evidence inventory

Extract `Validex_Review_Package.zip` into a new directory. Create a Python environment, install `environment/requirements-minimal.txt`, and run `python reproduce.py` from the archive root. The script checks 720 freeze-inventory entries, executes all 240 cases and writes `reproduction_output/summary.json`. It recomputes held-out, development and challenge results from the retained reference. `supporting_data` contains the detailed revision tables and the exact screening-rule script.

The evaluated snapshot is commit `86be06dfc64e33851b61c7b5d261657fec32e778`. The freeze chronology is supported by mutable file timestamps and a next-day commit, not an external preregistration. `SHA256SUMS.txt` checks the delivered archive contents. Historical inputs and outputs remain separate from revision analyses.

Table S1. Contents of `Validex_Review_Package.zip` by directory.

Directory	Contents
<code>benchmark_v2</code>	Inputs, specifications, reference, freeze inventory, generator
<code>historical_outputs_validex_0_2_0_2026-07-23</code>	Frozen outputs and original result tables
<code>current_outputs_2026-09-04_rerun</code>	Repeated execution outputs
<code>revision_analyses_2026-09-04</code>	Post hoc analysis results
<code>supporting_data</code>	Screening records, source rules and supplementary tables
<code>environment</code>	Historical environment record and reproduction requirements

S2 Case results and field performance

The primary endpoint compares ingestion, five active mappings and five ambiguity states. In each of the 13 non-exact held-out cases, the expected active `p_value` and `fdr` names were replaced by `null`. The structurally selected columns and all five ambiguity states agreed. Each probability column contained three invalid cells. The remaining three active mappings agreed. The interpretation of the conflict as a reference-design mismatch remains an author-led post hoc hypothesis; all 13 cases remain failures.

Table S2. Reproduced from manuscript Table 4. The 13 held-out cases that did not match. Each case carries three invalid cells in each probability column, so the valid fraction is $(n - 3)/n$ and falls below the 0.80 gate in every case. Structural selection and all five ambiguity states agreed with the reference in all 13. Each discrepancy is a prospective endpoint failure; the post hoc hypothesis is a structural-versus-active reference mismatch. Full mappings and ambiguity states are in `thirteen_case_table.csv`.

Case	Rows	Valid fraction	Expected → observed <code>p / FDR</code>
<code>SYNV2_0045</code>	11	0.727	<code>p_value / fdr</code> → <code>null / null</code>
<code>SYNV2_0057</code>	9	0.667	<code>p_value / fdr</code> → <code>null / null</code>
<code>SYNV2_0069</code>	14	0.786	<code>p_value / fdr</code> → <code>null / null</code>

Case	Rows	Valid fraction	Expected → observed p / FDR
SYNV2_0081	12	0.750	p_value / fdr → null / null
SYNV2_0093	10	0.700	p_value / fdr → null / null
SYNV2_0105	8	0.625	p_value / fdr → null / null
SYNV2_0117	13	0.769	p_value / fdr → null / null
SYNV2_0129	11	0.727	p_value / fdr → null / null
SYNV2_0141	9	0.667	p_value / fdr → null / null
SYNV2_0153	14	0.786	p_value / fdr → null / null
SYNV2_0165	12	0.750	p_value / fdr → null / null
SYNV2_0177	10	0.700	p_value / fdr → null / null
SYNV2_0189	8	0.625	p_value / fdr → null / null

Table S3. Held-out field precision, recall and F1 with Wilson 95% recall intervals. Counts and support are in manuscript Table 2; the intervals are descriptive because the fixtures are dependent.

Field	Precision	Recall	F1	Recall 95% interval
compound id	1.000	1.000	1.000	0.977–1.000
effect size	1.000	1.000	1.000	0.977–1.000
p value	1.000	0.919	0.958	0.866–0.952
fdr	1.000	0.912	0.954	0.855–0.948
annotation	1.000	1.000	1.000	0.972–1.000

The counts and support are in manuscript Table 2. Intervals are descriptive because fixtures are dependent. Full expected/observed dictionaries, cell classifications, warnings and failure interpretations are retained in `thirteen_case_table.csv`; family-specific field metrics are in `field_metrics_by_family.csv`.

S3 Sensitivity and output coverage

The usability threshold controls active probability mappings. The separate ordering threshold controls warning emission. Neither 0.80 value has a recorded statistical calibration. The usability analysis is in manuscript Table 5 and `threshold_sensitivity.csv`. The table below gives ordering warnings; historical uses coerced pairs and restricted uses pairs finite and within [0, 1]. Values are counts of warnings, not estimates of true statistical errors.

Table S4. Ordering warnings emitted at six warning thresholds under the historical coerced-pair rule (H) and the post hoc restricted rule (R), which requires both values to be numeric, finite and within [0, 1]. Partition sizes are 40, 160 and 40.

Threshold	Development H / R	Held-out H / R	Challenge H / R
0.5	0 / 0	0 / 0	0 / 0
0.6	3 / 3	13 / 13	4 / 4
0.7	3 / 3	13 / 13	4 / 4
0.8	3 / 3	13 / 13	4 / 4
0.9	5 / 3	21 / 13	6 / 4
1.0	6 / 3	26 / 13	8 / 4

H and R denote historical and restricted rules. Partition sizes are 40, 160 and 40. At 0.80, restricting pairs changes 20 denominators and zero warning decisions; at 0.90 and 1.00 it changes decisions. The benchmark has no infinite-valued case. Infinity handling is checked only in the characterization tests. The frozen endpoint covers eleven components. Audit score, confidence, completeness, user-facing findings and provenance have no frozen expected values. Post hoc reconstructed invalid-cell counts and

ordering/duplicate flags agreed in all 160 held-out cases. Byte equality on repetition is distinct from correctness of an unscored component. `output_coverage.csv` records these boundaries. `ablation_results.csv`, `format_performance.csv`, `challenge_performance.csv` and `duplicate_aware_metrics.csv` retain the corresponding supplementary results.

S4 Corpus search and screening

The source archive contains 140 files from 58 studies, retrieved on 23 July 2026 through several search waves. MetaboLights provides study data and metadata; Metabolomics Workbench was also consulted. The retained registry contains 33 queries covering Europe PMC and public repositories. It is incomplete for the earliest pilot wave: queries QRY_0001 through QRY_0005 are absent. A second bounded search on 24 July screened 330 unique article records. The available queries, limits and eligibility criteria follow below, and all 195 table dispositions are listed in S5.

A row unit counted as resolved under the screening script when a normalized header contained a compound, metabolite, feature, gene, name or mass-to-charge term. Sample or assay headers and unresolved headers received separate dispositions. This permissive header rule does not establish metabolomics modality. An uncontaminated table had no recorded prior Validex output; eight tables with Validex 0.1.0 demonstration outputs were excluded before screening. Eligibility required a traceable study, lawful access, an archived and hashed source, a bounded table with interpretable rows, documented metabolomics or lipidomics modality, at least one in-scope field, adequate documentation, resolved licensing and reproducible extraction. Abundance-only matrices, patient-level raw data, prior-exposure cases and prohibited duplicates were excluded. Reporting recommendations describe additional scientific metadata beyond this structural screen.

Sampling followed bounded repository queries and returned-record order, with no random draw or representative sampling frame. The expanded first-wave search capped records at 300 and downloads at 150; per-query limits were 12 for Europe PMC, eight for Zenodo and Figshare, and five for Dryad and OSF. It deduplicated candidates against existing and current records using lowercased DOI, accession and title tuples. The later search used Europe PMC requests under twelve source-labelled waves, capped at 500 records and 200 downloads, retaining at most 35 returned records per request. It deduplicated article identifiers and download URLs; extracted-file hashes defined table duplicates. Eligible selection was to sort by study and table identifier and allow two tables per study. No file was retrieved in that search, so table deduplication and the selection cap had no effect. These procedures support a bounded feasibility assessment, not an exhaustive search across twelve independent services.

A deterministic script assigned the 187 screening statuses from stored headers and metadata. Its second rule variant added probability-field and documentation requirements; no model was called at decision time. During revision, two authors separately verified the 13 mismatch cases and 195 screening records. The corresponding author confirmed checking original source tables and summaries. The retained records contain no paired human ratings or reconciliation log, so human inter-rater agreement and reviewer-specific decision changes cannot be calculated. The frozen screening manifest is unchanged from its first committed version; this establishes record stability, not agreement between the two authors.

The 33 retained first-wave queries and twelve later source-labelled waves are reproduced below. Source labels in the later registry describe intended retrieval categories; the implementation queried Europe PMC for each wave. The registry does not establish twelve independent platform searches. All original query fields and limits are preserved in `corpus_search_strings.csv`.

1. QRY_0006 | Zenodo | 2026-07-23

"metabolomics" "fold change" "p-value" "adjusted p-value"

Limits and filters: SW_0002_AMD_0002; file-first tabular result-name filter; max 12 records; max 8 files; <=10 MB/file; per-query return limit: 12.

2. QRY_0007 | Europe PMC | 2026-07-23

"metabolomics" "differential metabolites" supplementary

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

3. QRY_0008 | Europe PMC | 2026-07-23

"metabolomics" "fold change" "adjusted p"

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

4. QRY_0009 | Europe PMC | 2026-07-23

"metabolomics" "log2 fold change" FDR

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

5. QRY_0010 | Europe PMC | 2026-07-23

"metabolomics" "volcano plot" supplementary table

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

6. QRY_0011 | Europe PMC | 2026-07-23

"untargeted metabolomics" "significant metabolites" table

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

7. QRY_0012 | Europe PMC | 2026-07-23

"LC-MS metabolomics" "differential features"

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

8. QRY_0013 | Europe PMC | 2026-07-23

"GC-MS metabolomics" "fold change" p-value

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

9. QRY_0014 | Europe PMC | 2026-07-23

"NMR metabolomics" comparison p-value metabolites

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

10. QRY_0015 | Europe PMC | 2026-07-23

"lipidomics" differential analysis supplementary

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

11. QRY_0016 | Europe PMC | 2026-07-23

"metabolomics" q-value feature table

Limits and filters: SW_0003_AMD_0003; source_priority=1; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 12.

12. QRY_0017 | Zenodo | 2026-07-23

metabolomics filetype:xlsx "fold change"

Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

13. QRY_0018 | Zenodo | 2026-07-23
metabolomics filetype:csv "adjusted p"
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

14. QRY_0019 | Zenodo | 2026-07-23
metabolomics filetype:tsv logFC
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

15. QRY_0020 | Zenodo | 2026-07-23
metabolomics supplementary "p_value"
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

16. QRY_0021 | Zenodo | 2026-07-23
metabolomics supplementary FDR
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

17. QRY_0022 | Zenodo | 2026-07-23
metabolomics supplementary "significant metabolites"
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

18. QRY_0023 | Zenodo | 2026-07-23
metabolomics results table "q-value"
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

19. QRY_0024 | Zenodo | 2026-07-23
metabolite statistical results spreadsheet
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

20. QRY_0025 | Zenodo | 2026-07-23
metabolomics differential results dataset
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

21. QRY_0026 | Zenodo | 2026-07-23
metabolomics comparison statistics file
Limits and filters: SW_0003_AMD_0003; source_priority=2; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

22. QRY_0027 | Figshare | 2026-07-23
metabolomics differential results
Limits and filters: SW_0003_AMD_0003; source_priority=3; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

23. QRY_0028 | Figshare | 2026-07-23
 metabolomics fold change adjusted p
 Limits and filters: SW_0003_AMD_0003; source_priority=3; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

24. QRY_0029 | Figshare | 2026-07-23
 untargeted metabolomics significant metabolites
 Limits and filters: SW_0003_AMD_0003; source_priority=3; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

25. QRY_0030 | Figshare | 2026-07-23
 lipidomics differential analysis
 Limits and filters: SW_0003_AMD_0003; source_priority=3; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 8.

26. QRY_0031 | Dryad | 2026-07-23
 metabolomics fold change adjusted p
 Limits and filters: SW_0003_AMD_0003; source_priority=4; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

27. QRY_0032 | Dryad | 2026-07-23
 untargeted metabolomics significant metabolites
 Limits and filters: SW_0003_AMD_0003; source_priority=4; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

28. QRY_0033 | Dryad | 2026-07-23
 lipidomics differential analysis
 Limits and filters: SW_0003_AMD_0003; source_priority=4; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

29. QRY_0034 | Dryad | 2026-07-23
 metabolomics volcano supplementary table
 Limits and filters: SW_0003_AMD_0003; source_priority=4; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

30. QRY_0035 | OSF | 2026-07-23
 untargeted metabolomics significant metabolites
 Limits and filters: SW_0003_AMD_0003; source_priority=5; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

31. QRY_0036 | OSF | 2026-07-23
 lipidomics differential analysis
 Limits and filters: SW_0003_AMD_0003; source_priority=5; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

32. QRY_0037 | OSF | 2026-07-23
 metabolomics volcano supplementary table
 Limits and filters: SW_0003_AMD_0003; source_priority=5; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

33. QRY_0038 | OSF | 2026-07-23
 metabolomics q value feature table

Limits and filters: SW_0003_AMD_0003; source_priority=5; max_records=300; max_files=150; max_file_size_bytes=52428800; per-query return limit: 5.

34. WAVE_01 | publisher_supplementary_pages | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND (metabolomics OR lipidomics) AND (supplementary table OR supplementary data) AND (fold change OR p-value OR FDR)

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

35. WAVE_02 | europe_pmc_direct_supplement_links | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND untargeted metabolomics AND differential metabolites AND supplementary

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

36. WAVE_03 | pubmed_linked_publisher_supplements | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND targeted metabolomics AND case-control AND supplementary table

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

37. WAVE_04 | paper_associated_zenodo | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND lipidomics AND differential abundance AND adjusted p-value

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

38. WAVE_05 | paper_associated_figshare | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND metabolite fold change adjusted p-value table

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

39. WAVE_06 | paper_associated_dryad | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND metabolomics association results regression coefficient

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

40. WAVE_07 | metabolights_attachments | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND metabolomics intervention comparison supplementary table

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

41. WAVE_08 | metabolomics_workbench_files | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND longitudinal metabolomics differential results

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

42. WAVE_09 | gnps_or_massive_exports | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND multi-group metabolomics differential results

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

43. WAVE_10 | paper_associated_github | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND LC-MS metabolomics supplementary results

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

44. WAVE_11 | osf_project_files | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND GC-MS metabolomics supplementary results

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

45. WAVE_12 | institutional_repositories | 2026-07-24

OPEN_ACCESS:y AND FIRST_PDATE:[2020-01-01 TO 2024-12-31] AND metabolomics FDR supplementary xlsx

Limits and filters: OPEN_ACCESS:y; FIRST_PDATE 2020-2024 (Europe PMC syntax); per-query return limit: not recorded.

S5 Case-level corpus dispositions

The following table lists all 195 candidates, including eight prior-exposure exclusions. The four screened statuses sum to 187: 115 ineligible, 66 unresolved row units, six indeterminate modalities and zero eligible. The full CSV adds source identifiers, sheet, dimensions, headers, rule rationales and both script-variant statuses. Historical verification fields describe the September 4 export and are retained as historical evidence. The separate revision verification record records the subsequent author confirmation.

Table S5. All 195 candidate tables with their frozen disposition, including the eight prior-exposure exclusions. The four screened statuses sum to 187.

Table	Study	Disposition and exclusion reason
TBL_0001	STU_0001	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0002	STU_0040	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0003	STU_0001	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0004	STU_0040	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0005	STU_0001	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0006	STU_0041	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0007	STU_0041	row unit unresolved: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0008	STU_0041	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0009	STU_0057	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0010	STU_0001	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0011	STU_0041	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0012	STU_0041	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0013	STU_0041	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0014	STU_0057	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0015	STU_0001	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0016	STU_0041	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0017	STU_0041	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0018	STU_0041	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0019	STU_0001	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0020	STU_0002	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0021	STU_0042	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0022	STU_0002	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0023	STU_0002	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0024	STU_0002	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0025	STU_0057	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0026	STU_0002	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0027	STU_0002	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0028	STU_0003	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0029	STU_0003	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0030	STU_0003	row unit unresolved: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0031	STU_0003	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0032	STU_0003	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0033	STU_0003	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0034	STU_0004	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0035	STU_0004	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0036	STU_0004	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0037	STU_0004	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0038	STU_0005	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0039	STU_0005	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0040	STU_0005	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0041	STU_0043	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0042	STU_0043	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0043	STU_0043	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0044	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0045	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0046	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0047	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0048	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0049	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0050	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0051	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0052	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0053	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0054	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0055	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0056	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0057	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0058	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0059	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0060	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0061	STU_0043	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0062	STU_0043	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0063	STU_0043	indeterminate ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0064	STU_0043	indeterminate ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0065	STU_0043	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0066	STU_0043	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0067	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0068	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0069	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0070	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0071	STU_0043	indeterminate ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0072	STU_0043	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0073	STU_0043	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0074	STU_0043	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0075	STU_0045	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0076	STU_0045	row unit unresolved: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0077	STU_0045	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0078	STU_0045	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0079	STU_0045	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0080	STU_0045	indeterminate ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0081	STU_0046	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0082	STU_0046	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0083	STU_0006	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0084	STU_0006	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0085	STU_0007	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0086	STU_0007	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0087	STU_0008	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0088	STU_0008	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0089	STU_0009	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0090	STU_0009	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0091	STU_0010	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0092	STU_0010	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0093	STU_0011	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0094	STU_0011	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0095	STU_0012	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0096	STU_0012	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0097	STU_0013	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0098	STU_0013	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0099	STU_0014	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0100	STU_0014	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0101	STU_0015	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0102	STU_0015	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0103	STU_0016	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0104	STU_0016	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0105	STU_0017	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0106	STU_0017	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0107	STU_0018	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0108	STU_0018	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0109	STU_0019	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0110	STU_0019	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0111	STU_0020	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0112	STU_0020	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0113	STU_0021	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0114	STU_0021	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0115	STU_0022	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0116	STU_0022	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0117	STU_0023	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0118	STU_0023	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0119	STU_0024	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0120	STU_0024	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0121	STU_0025	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0122	STU_0026	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0123	STU_0026	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0124	STU_0027	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0125	STU_0027	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0126	STU_0028	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0127	STU_0028	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0128	STU_0029	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0129	STU_0029	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0130	STU_0030	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0131	STU_0030	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0132	STU_0031	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0133	STU_0031	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0134	STU_0032	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0135	STU_0032	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0136	STU_0033	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0137	STU_0033	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0138	STU_0034	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0139	STU_0034	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0140	STU_0035	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0141	STU_0035	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0142	STU_0036	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0143	STU_0036	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0144	STU_0037	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0145	STU_0037	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0146	STU_0038	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0147	STU_0038	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0148	STU_0039	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0149	STU_0039	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0150	STU_0047	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0151	STU_0047	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0152	STU_0048	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0153	STU_0049	indeterminate ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0154	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0155	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0156	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0157	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0158	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0159	STU_0049	indeterminate ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0160	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0161	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0162	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0163	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0164	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0165	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0166	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0167	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0168	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0169	STU_0049	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0170	STU_0052	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0171	STU_0053	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0172	STU_0053	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0173	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0174	STU_0054	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0175	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0176	STU_0054	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0177	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0178	STU_0054	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0179	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0180	STU_0054	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0181	STU_0054	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0182	STU_0054	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0183	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0184	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0185	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0186	STU_0054	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0187	STU_0055	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0188	STU_0055	row unit unresolved: Pass A and Pass B assigned the same precise final status.
TBL_0189	STU_0059	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0190	STU_0060	contaminated prior product output: Prior Validex output exists before the fully AI-authored primary evaluation.
TBL_0191	STU_0061	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.

Table	Study	Disposition and exclusion reason
TBL_0192	STU_0062	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0193	STU_0063	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0194	STU_0064	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.
TBL_0195	STU_0065	confirmed ineligible ai assessed: Pass A and Pass B assigned the same precise final status.

S6 Author verification and historical changes

Two authors separately verified the 13 mismatch cases and 195 corpus-screening records during revision. The corresponding author confirmed checking original source tables and summaries. This is author verification after execution. No paired human decisions or reconciliation log are retained, so inter-rater agreement cannot be calculated and no external-review claim is made. The screening manifest is byte-identical to its first committed version, b336c77 of 24 July 2026. No saved screening-disposition changes were found. That comparison cannot establish the authors' initial agreement or exclude unrecorded changes during their review.

The 13-case interpretation remains a hypothesis about the conflict between structural expectations and gated emission. Verification does not remove any mismatch from the primary endpoint. No signature or separately dated review form is represented by this statement.

S7 Machine-readable supporting files

thirteen_case_table.csv

field_metrics_by_family.csv

family_performance.csv

format_performance.csv

challenge_performance.csv

threshold_sensitivity.csv

fdr_rule_comparison.csv

fdr_warning_threshold_sweep.csv

ablation_results.csv

output_coverage.csv

case_fingerprints.csv

duplicate_groups.csv

duplicate_aware_metrics.csv

benchmark_timeline.csv

corpus_exclusion_log.csv

corpus_search_strings.csv

corpus_screening_rules_verbatim.py.txt

ground_truth_internal_consistency.csv

screening_scripts supplies the retained search and screening implementations. corpus_source_records supplies the frozen source manifest, source identifiers, table boundaries and eligibility decisions. These

records distinguish the 188-table semantic-review workflow from the later 187-table deterministic eligibility screen. The main manuscript retains its separate AI-use statement.