Supplementary Material

Selection, adjustment, and unmeasured confounding in routinely collected laboratory data: a transparent bias-analysis framework applied to 25-hydroxyvitamin D and two exemplar analytes (calcium and urate) in 82,279 adult records

S1. Detailed cohort-flow counts and plausibility bounds

Filter step	Records	Cumulative excluded
Source records (all 25(OH)D-linked laboratory rows 2014–2023)	107,900	—
Numeric 25(OH)D result	107,856	44
Adults 18–100 with valid demographics (master analytic pool)	82,279	25,621
Calcium + albumin co-tested and within plausibility bounds	42,518	65,382
First eligible calcium patient-year (analytic — calcium cohort)	34,924	—
First eligible urate patient-year (reference — urate cohort)	22,059	—

Bounds are the same for both cohorts. Consolidation to one representative record per patient-year is deterministic (first date within year in the source workbook).

S2. IPW model diagnostics

Weight statistic	Urate cohort	Calcium cohort
Adults with modelling covariates	82,194	82,194
Included records (S=1)	34,557	50,524
P(S=1) — marginal selection probability	0.420	0.615
Selection-model c-statistic (AUC)	0.592	0.573
Raw stabilised weight — mean (SD)	1.000 (0.241)	1.001 (0.108)
Raw stabilised weight — median (P25, P75)	0.965 (0.865, 1.081)	0.997 (0.918, 1.081)
Raw stabilised weight — min, max	0.483, 8.617	0.679, 1.425
Trim thresholds (P1, P99 among S=1)	0.699, 1.814	0.792, 1.250
Trimmed weight — mean (SD)	0.993 (0.187)	1.000 (0.107)

Trimmed weight — min, max	0.699, 1.814	0.792, 1.250

Stabilised weights are trimmed at the 1st and 99th percentiles among included records. Selection model: logistic regression on age spline (df=4), sex, 25(OH)D per 10 nmol/L, calendar month and calendar year. Both weight distributions are well-behaved (trimmed mean ~1.0, SD < 0.2), and the selection model has modest discrimination (AUC 0.57–0.59), consistent with selection into a laboratory-tested subgroup that is only weakly predictable from measured covariates.

S3. Spline sensitivity (M4 primary, ns df variations)

Spline df	Urate M4 primary (µmol/L per 10 nmol/L)	Calcium M4 primary (µmol/L per 10 nmol/L)
3	−0.98 (−1.58, −0.38)	+4.10 (3.54, 4.66)
4	−1.00 (−1.60, −0.40)	+4.12 (3.55, 4.69)
5	−1.02 (−1.63, −0.41)	+4.15 (3.58, 4.72)

Estimates for the linear slope of 25(OH)D on the outcome are stable across natural-spline degrees of freedom, consistent with a broadly linear exposure–response after M4 adjustment.

S4. Leave-one-out negative-control benchmark

Cohort	Held-out control	Remaining-control range |Std M4|	Target |Std M4|	Verdict
Urate	Chloride	0.008–0.068	0.026	within
Urate	Potassium	0.008–0.098	0.026	within
Urate	Sodium	0.046–0.098	0.026	outside (target below)
Urate	Total CO_2	0.008–0.098	0.026	within
Urate	Phosphate	0.008–0.098	0.026	within
Calcium	Chloride	0.016–0.069	0.071	outside (target above)
Calcium	Potassium	0.016–0.083	0.071	within
Calcium	Sodium	0.050–0.083	0.071	within
Calcium	Total CO_2	0.016–0.083	0.071	within
Calcium	Phosphate	0.016–0.083	0.071	within

For each cohort, we compute the range of absolute standardised M4 effects among the remaining four controls when one control at a time is held out. The verdict is “within” when the absolute target effect lies within the remaining range; verdict changes flag potentially informative controls. In the urate cohort, holding out sodium leaves the target below the remaining range (i.e., the target effect is smaller than the M4 effect on the remaining four controls). In the calcium cohort, holding out chloride leaves the target above the remaining range.

S5. Software environment

Python 3.14, statsmodels 0.15.0, patsy, scipy, pandas, numpy, matplotlib. R environment (mirror): R ≥4.3 with sandwich, splines, survey, EValue, data.table, lmtest.

S6. Deposited files

- build_cohort.py — deterministic Python cohort builder (validated against the previously published urate paper).
- analysis_stage2.py — Python analysis pipeline (M0→M5, IPW, splines, PR, E-values, negative-control benchmark).
- analysis.R — R mirror of the analysis pipeline for independent replication.
- cohort_calcium.csv, cohort_urate.csv, master_records.csv — aggregated analytic tables (no individual identifiers).
- table2_cascade.csv, table3_evalues_splines.csv, table4_negcontrols.csv — machine-readable versions of the main tables.
- fig1.png, fig2_forest.png, fig3_evalues.png, fig4_negcontrols.png — figure files at 250–300 dpi.

S7. STROBE-style reporting checklist

- Study design: single-centre analytical study of routinely-collected laboratory data.
- Setting: UMC Ljubljana, Institute of Clinical Chemistry and Biochemistry, 2014–2023.
- Participants: adults with a same-day 25(OH)D and outcome measurement satisfying plausibility bounds; consolidated to one record per patient-year.
- Variables: exposure 25(OH)D (nmol/L); outcomes serum urate (µmol/L) and serum total calcium (mmol/L); phenotypes hyperuricemia, hypocalcaemia, hypercalcaemia; covariates age, sex, meteorological season, era, serum albumin, eGFR-MDRD.
- Data sources: hospital laboratory information system, single Siemens Advia platform throughout.
- Bias: quantified through the framework described (adjustment cascade, IPW, splines, E-values, negative-control-outcome benchmark).

- Statistical methods: as described in Section 2.3–2.6 of the main manuscript.
- Missing data: covariates were required to be present for each model; the M5 same-sample subset restricts to records with a paired creatinine.
- Sensitivity analyses: IPW, spline df variation, leave-one-out negative-control benchmark, E-value calculations.