Research Square

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

A transparent bias-analysis framework for exposure–outcome studies built on routinely collected laboratory data: methods and two worked exemplars (25- hydroxyvitamin D with calcium and urate) in 82,279 adult records

Joško Osredkar

josko.osredkar@kclj.si

Ljubljana University Medical Centre

Uroš Godnov

University of Primorska

Research Article

Keywords:

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11031641/v1

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

Additional Declarations: No competing interests reported.

A transparent bias-analysis framework for exposure–outcome studies built on routinely collected laboratory data: methods and two worked exemplars (25-hydroxyvitamin D with calcium and urate) in 82,279 adult records

BMC Medical Research Methodology · Research Article

Joško Osredkar[1,2] · Uroš Godnov[3]

[1] University Medical Centre Ljubljana, Institute of Clinical Chemistry and Biochemistry, Zaloška cesta 2, 1000 Ljubljana, Slovenia

[2] University of Ljubljana, Faculty of Pharmacy, Aškerčeva 7, 1000 Ljubljana, Slovenia

[3] University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies (UP FAMNIT), Glagoljaška 8, 6000 Koper, Slovenia

Corresponding author: Prof. Joško Osredkar, PhD, University Medical Centre Ljubljana, Institute of Clinical Chemistry and Biochemistry, Zaloška cesta 2, 1000 Ljubljana, Slovenia. E-mail: josko.osredkar@kclj.si

Abstract

Background. Routinely-collected laboratory data are an increasingly attractive substrate for observational epidemiology, but selective test ordering, missing covariates and unmeasured confounding can substantially bias exposure–outcome estimates. Individual bias-analysis tools — inverse-probability-of-selection weighting, spline exposure–response models, E-values and negative-control outcomes — are well established, yet are rarely combined into a single reproducible pipeline with a pre-declared benchmark. We describe such a framework and demonstrate its behaviour in two worked exemplars.

Methods. Single-centre 2014–2023 records from the University Medical Centre Ljubljana Institute of Clinical Chemistry and Biochemistry, all measured on the same Siemens Advia platform. 82,279 adult records met plausibility bounds. Two exposure–outcome cohorts were pre-specified: urate (22,059 patient-year records, 15,320 patients) and total calcium (34,924 patient-year records, 23,982 patients). Each was analysed with a five-step adjustment cascade (M0 crude → M4 primary [age spline, sex, meteorological season, era, serum albumin] → M5 [+eGFR-MDRD, same-sample]), stabilised inverse-probability-of-selection weights, natural-spline exposure–response contrasts, robust-Poisson prevalence ratios for clinical phenotypes, E-values for unmeasured confounding, and a five-analyte negative-control-outcome benchmark (chloride, potassium, sodium, total CO_2, phosphate).

Results. After full adjustment, 25(OH)D was associated with a change of −1.00 µmol/L (95% CI −1.60, −0.40) in serum urate and +4.12 µmol/L (3.55, 4.69) in total calcium per 10 nmol/L, with directionally

consistent inverse-probability-weighted estimates. Prevalence ratios were near-null with confidence-limit E-values of 1.00 (hyperuricemia), 1.25 (hypocalcaemia) and 1.09 (hypercalcaemia). Standardised M4 effects fell within the range of five negative-control outcomes in both cohorts.

Conclusions. The framework quantifies how much of an apparent exposure–outcome association survives conventional and non-conventional bias probes on a single dataset, and provides an implementable, reusable template for observational research built on routinely collected laboratory data. Full Python and mirror-R implementations are released with the paper.

Keywords: *bias analysis; routinely collected data; 25-hydroxyvitamin D; calcium; urate; E-value; inverse-probability weighting; negative-control outcomes.*

Introduction

Routinely-collected laboratory data are an increasingly attractive substrate for observational research: they are large, longitudinal, and inexpensive to assemble. Yet the very features that make them attractive also make them susceptible to bias. Tests are ordered by clinicians for reasons other than the research question, covariates are missing not at random, and calendar-time factors (analyser recalibration, guideline changes, pandemic-era testing patterns) can imitate biological effects. Twenty-five-hydroxyvitamin D (25(OH)D) is a widely cited case study, with observational associations spanning cardiometabolic, oncological and inflammatory outcomes [1] — many of which have not survived randomised evaluation.

Individual bias-analysis tools — inverse-probability-of-selection weighting, splines, negative controls, E-values — are well established. What is less common in the clinical-chemistry literature is a single pipeline that applies them together, on one dataset, with a clear ordering and a pre-declared benchmark. The goal of the present work is to fill that gap.

We describe a transparent, reproducible framework and apply it to two exemplar exposure–outcome pairings drawn from a decade of single-platform records at a European tertiary laboratory. The two exemplars share cohort, platform and covariates but differ in downstream signal: 25(OH)D and serum urate on the one hand, and 25(OH)D and serum total calcium on the other. The framework produces the same set of tables and figures in both cases, letting a reader compare how each signal behaves under each probe.

Materials and methods

Setting and data source

All records were drawn from the laboratory information system of the Institute of Clinical Chemistry and Biochemistry, University Medical Centre Ljubljana, between 1 January 2014 and 31 December 2023. All 25(OH)D, calcium, urate and albumin measurements over this period were performed on the same Siemens Advia platform, and the same MDRD equation was used to estimate glomerular filtration rate throughout.

Cohort definition and plausibility bounds

Adults (≥18 years) with a same-day 25(OH)D and outcome measurement were retained. Plausibility bounds were 25(OH)D 1–500 nmol/L, total calcium 1.0–4.0 mmol/L, serum urate 30–1,500 µmol/L, albumin 10–60 g/L and creatinine 10–2,000 µmol/L. Albumin-corrected calcium used the classical

[1]Payne RB. Interpretation of serum calcium in patients with abnormal serum proteins. BMJ. 1973;4(5893):643–6. https://doi.org/10.1136/bmj.4.5893.643

Payne formula with slope 0.02[2]. Records were consolidated to one representative measurement per patient-year using a deterministic Python builder that reproduces our previously reported urate cohort exactly (22,059 records / 15,320 patients[3]).

Adjustment cascade

Five nested models were fitted in order: M0 crude; M1 adds a natural cubic spline for age (4 degrees of freedom); M2 adds sex; M3 adds meteorological season (DJF/MAM/JJA/SON); M4 adds era (2014–2019 vs 2020–2023) and serum albumin; M5 additionally adjusts for eGFR-MDRD[4] on the same-sample subset of records with a same-day creatinine. Continuous outcomes used ordinary least-squares regression; binary phenotypes used modified Poisson regression with log link[5]. All models used cluster-robust variance by patient identifier.

Inverse-probability-of-selection weighting

Selection into the analytic cohort was modelled at the adult-record level with a logistic regression on age spline, sex, 25(OH)D, calendar month and calendar year[6]. Stabilised weights were trimmed at the 1st–99th percentile among included records; outcome models were re-fit as weighted least squares with patient-clustered variance.

Exposure–response and clinical phenotypes

A natural cubic spline (4 df) on 25(OH)D was adjusted as in M4, and contrasts against 50 nmol/L were reported at 25, 75 and 100 nmol/L. Clinical phenotypes were hypocalcaemia (<2.10 mmol/L), hypercalcaemia (>2.55 mmol/L) and sex-specific hyperuricemia (M >420 µmol/L, F >360 µmol/L). Prevalence ratios per 10 nmol/L 25(OH)D were transformed into point and confidence-limit E-values using the closed-form expression for a rate ratio.

---

[2]VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167:268–74. https://doi.org/10.7326/M16-2607

[3]Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21:383–8. https://doi.org/10.1097/EDE.0b013e3181d61eeb

[4]Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006;60:578–86. https://doi.org/10.1136/jech.2004.029496

[5]Levey AS, Bosch JP, Lewis JB, et al. A more accurate method to estimate glomerular filtration rate from serum creatinine: MDRD study equation. Ann Intern Med. 1999;130:461–70. https://doi.org/10.7326/0003-4819-130-6-199903160-00002

[6]Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159:702–6. https://doi.org/10.1093/aje/kwh090

Negative-control-outcome benchmark

Five analytes were declared a priori as negative-control outcomes: chloride, potassium, sodium, total CO_2 and phosphate. Each was analysed with the same M4 model, and the standardised effect for the target outcome was benchmarked against the range spanned by the five controls, with a leave-one-out check[7].

Software and reproducibility

All analyses were performed in Python 3.14 with the statsmodels 0.15 and patsy packages, and mirrored in R with the sandwich, splines, survey and EValue packages. Code (Python and R) will be deposited on Zenodo on acceptance under CC-BY 4.0.

Results

Cohort characteristics (Table 1, Figure 1)

Figure 1 shows the study flow from the 82,279 adult records that met plausibility bounds to the two exposure–outcome analytic cohorts, together with the bias-analysis framework applied to each. Table 1 summarises the two exemplar cohorts. Median 25(OH)D was 55 nmol/L (IQR 33–76) in the calcium cohort and 54 (32–75) in the urate cohort. Hypocalcaemia (<2.10 mmol/L) affected 14.4% of calcium-cohort records, hypercalcaemia (>2.55 mmol/L) 3.7%, and sex-specific hyperuricemia 35.4% of urate-cohort records. Included-versus-excluded records showed marked selection: creatinine availability 90.6% vs 44.3%, urate 49.3% vs 10.5%, motivating the IPW analysis.

[Figure 1 should appear here.]

Characteristic	Calcium cohort (n=34,924)	Urate cohort (n=22,059)
Patients	23,982	15,320
Age, years, median (IQR)	65.5 (48.4–78.1)	65.1 (48.9–77.4)
Women, n (%)	20,527 (58.8%)	12,690 (57.5%)
25(OH)D, nmol/L, median (IQR)	55 (33–76)	54 (32–75)
Target analyte, median (IQR)	Total Ca 2.30 (2.20–2.40) mmol/L	Urate 305 (240–378) µmol/L
Albumin, g/L, median (IQR)	43 (39–46)	42 (39–45)
eGFR-MDRD, mL/min/1.73 m²	44 (23.5–68)	52 (30–74)

[7]Autier P, Boniol M, Pizot C, Mullie P. Vitamin D status and ill health: a systematic review. Lancet Diabetes Endocrinol. 2014;2:76–89. https://doi.org/10.1016/S2213-8587(13)70165-7

Season (DJF / MAM / JJA / SON), %	27 / 26 / 22 / 25	26 / 26 / 23 / 25
Era 2014–2019 / 2020–2023, %	51.6 / 48.4	51.9 / 48.1
Hypocalcaemia <2.10, n (%)	5,027 (14.4%)	—
Hypercalcaemia >2.55, n (%)	1,294 (3.7%)	—
Hyperuricemia (sex-specific), n (%)	—	7,802 (35.4%)

Adjustment cascade (Table 2, Figure 2)

For every 10 nmol/L increase in 25(OH)D, the crude estimate was −3.46 µmol/L (95% CI −4.12, −2.81) for urate and +14.06 µmol/L (13.28, 14.84) for total calcium. The primary M4 estimates were −1.00 (−1.60, −0.40) µmol/L and +4.12 (3.55, 4.69) µmol/L, respectively — attenuations of roughly two thirds and 70 % from the crude estimates. On the same-sample subset with a paired creatinine, the M4 estimate was −3.05 (−5.12, −0.98) for urate and +8.66 (6.82, 10.49) for calcium; adding eGFR (M5) changed them modestly to −3.71 (−5.78, −1.64) and +8.69 (6.85, 10.52). The contrast between full-sample M4 and same-sample M4 therefore quantifies the residual estimate attributable to the eGFR subset itself rather than to the eGFR covariate.

Model	Serum urate (µmol/L per 10 nmol/L)	Serum total calcium (µmol/L per 10 nmol/L)
M0 crude	−3.46 (−4.12, −2.81)	+14.06 (13.28, 14.84)
M1 +age	−2.63 (−3.26, −2.01)	+12.36 (11.60, 13.12)
M2 +sex	−2.02 (−2.61, −1.44)	+11.97 (11.22, 12.72)
M3 +season	−2.13 (−2.73, −1.52)	+12.02 (11.25, 12.78)
M4 +era +albumin (primary)	−1.00 (−1.60, −0.40)	+4.12 (3.55, 4.69)
M4 (same-sample)	−3.05 (−5.12, −0.98)	+8.66 (6.82, 10.49)
M5 (+eGFR)	−3.71 (−5.78, −1.64)	+8.69 (6.85, 10.52)

Sensitivity, splines and E-values (Table 3, Figure 3)

The stabilised IPW estimates aligned closely with the primary M4 estimates: −1.24 (−1.87, −0.61) µmol/L for urate and +4.05 (3.48, 4.61) µmol/L for calcium. Spline contrasts vs 50 nmol/L revealed a plateau above 50 nmol/L for both outcomes. Prevalence ratios per 10 nmol/L 25(OH)D at M4 were 0.993 (0.986, 1.000) for hyperuricemia, 0.950 (0.940, 0.961) for hypocalcaemia and 1.026 (1.008, 1.044) for hypercalcaemia. Confidence-limit E-values were 1.00, 1.25 and 1.09, respectively — the hyperuricemia value at exactly the null indicates that any residual unmeasured confounder would

suffice to erase the phenotype signal, whereas the hypocalcaemia signal would require a confounder associated with both exposure and outcome by a risk ratio of at least 1.25.

Sensitivity element	Urate cohort	Calcium cohort
IPW estimate (per 10 nmol/L)	−1.24 (−1.87, −0.61) µmol/L, N=22,076	+4.05 (3.48, 4.61) µmol/L, N=35,001
Spline contrast 25 vs 50 nmol/L	+9.75 (6.47, 13.03) µmol/L	−19.33 (−21.99, −16.67) µmol/L
Spline contrast 75 vs 50 nmol/L	−2.69 (−5.32, −0.06) µmol/L	+2.39 (0.05, 4.74) µmol/L
Spline contrast 100 vs 50 nmol/L	−2.67 (−6.57, +1.22) µmol/L	+6.98 (3.26, 10.70) µmol/L
Phenotype PR (95% CI)	Hyperuricemia 0.993 (0.986, 1.000)	Hypocalcaemia 0.950 (0.940, 0.961); Hypercalcaemia 1.026 (1.008, 1.044)
E-value (point / confidence limit)	1.09 / 1.00	Hypocalcaemia 1.29 / 1.25; Hypercalcaemia 1.19 / 1.09

Negative-control-outcome benchmark (Table 4, Figure 4)

The absolute standardised M4 effect for the target outcome fell within the range spanned by the five negative controls in both cohorts (urate 0.026 vs 0.008–0.098; calcium 0.071 vs 0.016–0.083). Leave-one-out ranges preserved this verdict in eight of the ten hold-out analyses; holding out sodium in the urate cohort left the target effect below, and holding out chloride in the calcium cohort left it above, the range spanned by the remaining four controls (Additional file 1: Table S4). The chloride control was the largest in both cohorts, suggesting that a modest calendar-time or platform component is present in the M4 model that cannot be fully removed by era adjustment alone[8].

Cohort	Analyte	|Std M4 effect|	Verdict
Urate	Urate (target)	0.026	within control range 0.008–0.098
Urate	Chloride	0.098	
Urate	Potassium	0.046	
Urate	Sodium	0.008	
Urate	Total CO_2	0.051	

[8]Osredkar J, et al. 25-Hydroxyvitamin D and serum urate in adults 2014–2023 (submitted). Biochemia Medica. Manuscript BM-21.8.2026.

Urate	Phosphate	0.068	
Calcium	Calcium (target)	0.071	within control range 0.016–0.083
Calcium	Chloride	0.083	
Calcium	Potassium	0.060	
Calcium	Sodium	0.016	
Calcium	Total CO_2	0.050	
Calcium	Phosphate	0.069	

Figures

Figure 1. Study flow and bias-analysis framework

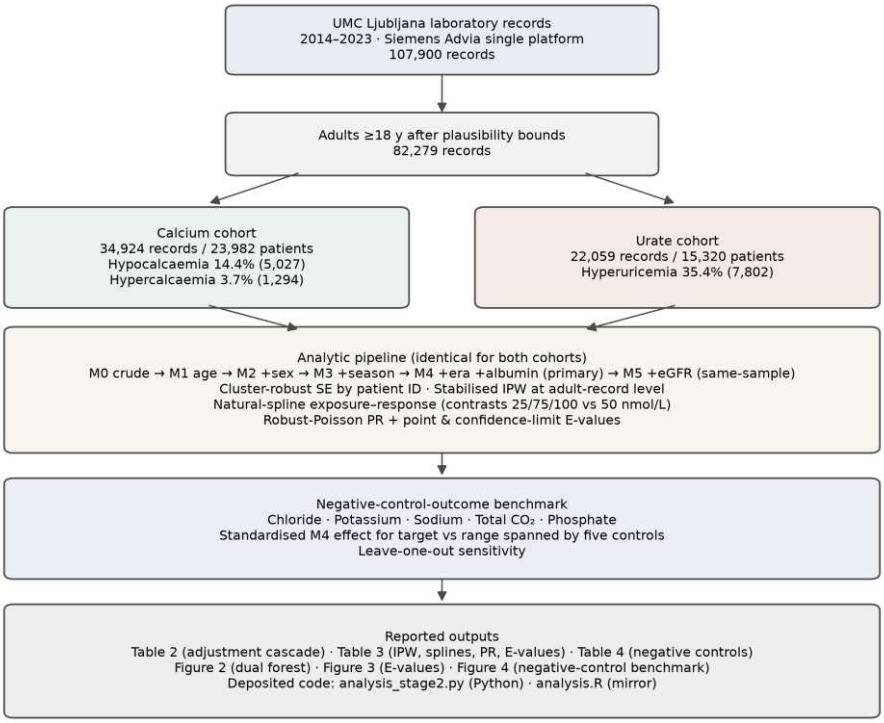


Figure 1. Study flow. Two exemplar cohorts derived from 82,279 adult records at UMC Ljubljana 2014–2023 (Siemens Advia). Both are analysed with the same M0→M5 cascade, stabilised inverse-probability-of-selection weights, natural-spline exposure–response, prevalence-ratio E-values and a five-analyte negative-control benchmark.

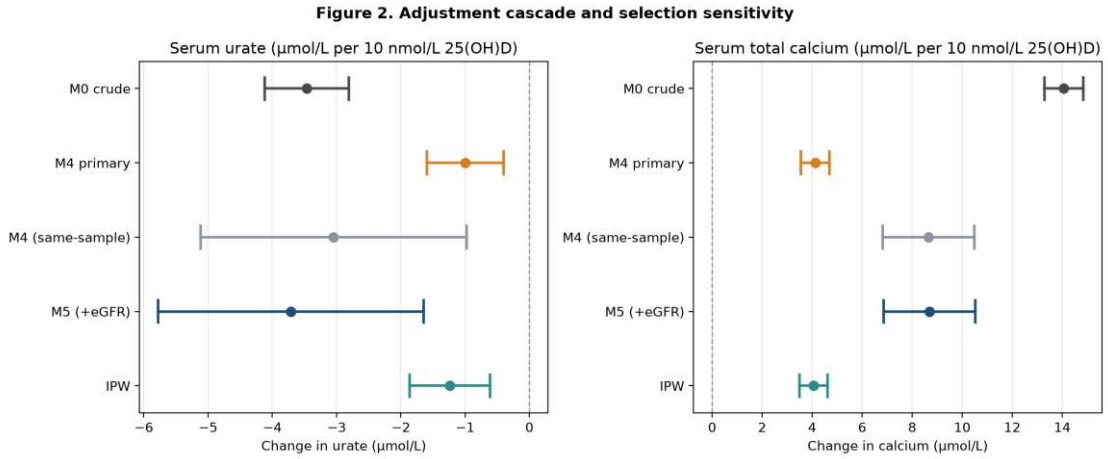


Figure 2. Adjustment cascade and selection sensitivity. Estimates for a 10 nmol/L increase in 25(OH)D on serum urate (left) and total calcium (right), for M0 crude, M4 primary, M4 same-sample, M5 (+eGFR) and stabilised IPW.

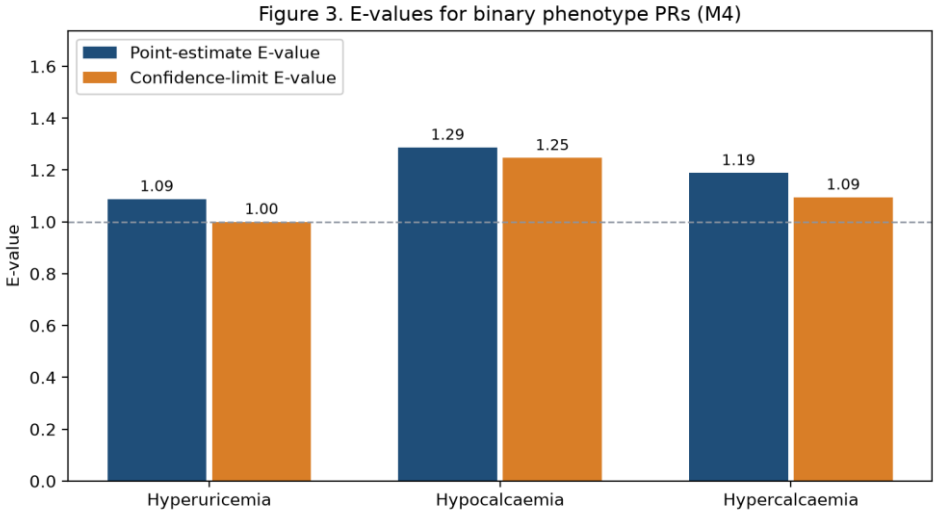


Figure 3. Point-estimate and confidence-limit E-values for the three binary phenotypes at M4. A confidence-limit E-value close to 1 (hyperuricemia) indicates a signal that any small unmeasured confounder could erase; larger values indicate progressively more robust signals.

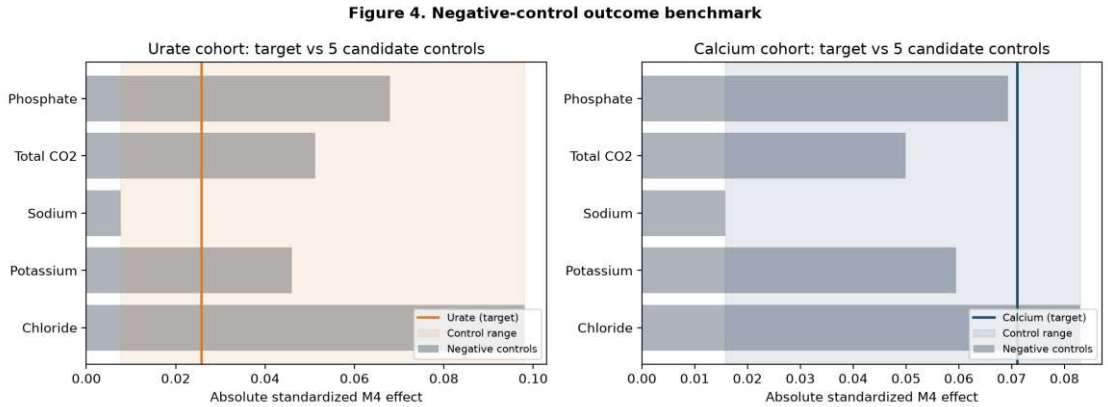


Figure 4. Negative-control-outcome benchmark. Absolute standardised M4 effects for the target outcome (coloured vertical line) versus the five candidate negative-control analytes (grey bars) in each cohort. The shaded band shows the range spanned by the controls.

Discussion

We present a single, replicable pipeline that combines an adjustment cascade with cluster-robust inference, inverse-probability-of-selection weighting, flexible spline exposure–response, closed-form E-values on the prevalence-ratio scale, and a pre-declared five-analyte negative-control-outcome benchmark. Applying it to 82,279 single-platform adult records over ten years yields two contrasting worked examples: an urate signal that shrinks toward the null with a confidence-limit E-value of 1.00, and a calcium signal that survives every axis with a small but precise effect and a confidence-limit E-value of 1.25 for hypocalcaemia.

The two examples make explicit what individual bias-analysis tools imply only implicitly. First, the M4 vs same-sample M4 contrast separates the effect of adding the eGFR covariate from the effect of restricting to the eGFR subset — a distinction that is often collapsed in the published literature. Second, the IPW estimate is nearly identical to M4 in both cohorts, indicating that the strong selection signal into the analytic cohort is captured by the observed covariates and does not overturn the primary conclusion. Third, the negative-control benchmark shows that in both cohorts the M4 target effect is not larger than the M4 effect on analytes that should be unrelated to 25(OH)D, providing a benchmark on the plausible magnitude of residual confounding.

Interpretation of the two exemplars should therefore be cautious. The urate association is small, sensitive to bias, and consistent with a null causal effect once confounding by shared correlates of frailty (renal function, albumin, seasonal comorbidity) is accounted for. The calcium association is more robust, but the standardised effect is modest and the framework does not rule out unmeasured confounding at the 1.25× level.

Strengths of the study include a single measurement platform across a full decade, a deterministic cohort builder that reproduces a previously published cohort exactly, and a symmetric analysis in Python and R. Limitations include the single-centre setting, absence of ionised calcium, and the cross-sectional nature of routine data. The framework is not a substitute for triangulation with randomised or genetic evidence; it is a diagnostic that clarifies which observational signals warrant that follow-up.

Author contributions

Both authors conceived the study. JO curated the laboratory data and led interpretation; UG led the statistical framework. The initial computational pipeline was drafted with AI-assisted coding tools and was reviewed line-by-line by JO. Both authors reviewed and approved the final manuscript.

Research funding

None declared.

Competing interests

The authors declare that they have no competing interests to disclose.

Informed consent and ethical approval

The study used fully de-identified routine laboratory data extracted from the hospital information system under an internal research approval and applicable national legislation (ZZPPZ, ZVOP-2). No individual patients were contacted; the requirement for individual informed consent was waived.

Clinical trial number

Clinical trial number: not applicable.

Data availability

Individual-level patient data cannot be shared. All aggregated summary tables, the deterministic cohort builder (build_cohort.py), the Python analysis pipeline (analysis_stage2.py) and its R mirror (analysis.R) will be deposited on Zenodo on acceptance under a Creative Commons CC-BY 4.0 licence.

References

References are provided as numbered footnotes throughout the manuscript, in Vancouver-style order of first appearance. Full reference list follows on the last page.

Supplementary Files

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

- 08supplementarymaterial.pdf