

Supplementary Information

Serum 25-hydroxyvitamin D, liver enzymes and FIB-4 fibrosis risk: a statistically robust but clinically negligible association from 31,607 adult laboratory records

Joško Osredkar, Uroš Godnov, Aleš Jerin, Darko Siuka, David Drobne

Contents

Item	Title
Supplementary Methods	Derivation of eGFR, FIB-4 and the ALT-elevation flag; the 6-record exposure-plausibility caveat; corrections applied after independent verification
Table S1	Missing-data and co-testing (collider) analysis
Table S2	Spearman correlations with 25(OH)D, stratified by sex, with formal Fisher z tests
Table S3	Dunn post-hoc comparisons between 25(OH)D categories, all outcomes
Table S4	Kruskal–Wallis tests across 25(OH)D categories, all outcomes
Table S5	Outcomes by 25(OH)D category, with confidence intervals, and Cliff’s delta (deficient vs optimal)
Table S6	Full five-step model output, all six continuous outcomes
Table S7	Likelihood-ratio tests for effect modification
Table S8	Subgroup estimates (interpretable only via Table S7)
Table S9	Seasonal summary and seasonally stratified correlations
Table S9b	Monthly summary
Table S10	Annual trend, 2014–2023
Table S11	Sensitivity analysis: one record per patient
Table S12	Sensitivity analyses: ALT threshold, clustering, reclassification arithmetic and E-value
Table S13	Completed STROBE checklist for cross-sectional studies
Table S14	Negative-control outcome analysis: the control set and its tiers, the clean band at model 4, the extended nine-control band, the contaminated controls, the full M0–M3 cascade, the continuous forms, the patient-level frame and the positive control

Supplementary Methods

S.1 Derived variables

Estimated glomerular filtration rate. eGFR was computed with the race-free CKD-EPI 2021 creatinine equation [30]. Serum creatinine, delivered in $\mu\text{mol/L}$, was converted to mg/dL by dividing by 88.4. With $\kappa = 0.7$ (women) or 0.9 (men) and $\alpha = -0.241$ (women) or -0.302 (men):

$$\text{eGFR} = 142 \times \min(\text{Scr}/\kappa, 1)^{\alpha} \times \max(\text{Scr}/\kappa, 1)^{-1.200} \times 0.9938^{\text{age}} \times 1.012 \text{ (if female)}$$

eGFR was computable in 31,253 of 31,607 records (98.9%). The binary flag $\text{eGFR} < 60 \text{ mL/min/1.73 m}^2$ is **missing, not negative**, where creatinine is missing.

FIB-4. $\text{FIB-4} = \text{age (years)} \times \text{AST (U/L)} / [\text{platelets (10}^9\text{/L)} \times \sqrt{\text{ALT (U/L)}}]$ [1]. Because the laboratory delivers catalytic activities in $\mu\text{kat/L}$, AST and ALT were first multiplied by 60. FIB-4 was computable in **30,232 of 31,607 records (95.6%)**; the 1,375 records without a platelet count have FIB-4 = missing. The conventional high-risk threshold of 2.67 was applied [1,2], and — critically — **the binary flag FIB-4 > 2.67 is missing wherever FIB-4 is missing** (see S.3). The low-risk threshold of 1.30 was recorded but is not used for any inference here.

Age-free FIB-4 surrogate. To test whether the 25(OH)D–FIB-4 association is arithmetically forced by the age term in the FIB-4 numerator, we defined $\text{AST (U/L)} / [\text{platelets (10}^9\text{/L)} \times \sqrt{\text{ALT (U/L)}}]$ and repeated the entire adjustment cascade on its natural logarithm.

Elevated ALT. The primary definition uses the sex-specific “true-normal” upper limits recommended by the ACG, $\text{ALT} > 25 \text{ U/L}$ in women and $> 33 \text{ U/L}$ in men [8], which derive from the screened blood-donor analysis of Prati and colleagues [9]. These are *not* the reporting laboratory’s own reference interval and they flag 43.5% of this cohort; that is a property of the threshold, not of the population, and is stated explicitly in the main text. Prespecified sensitivity definitions: (i) the routine IFCC 37°C clinical-chemistry upper reference limits $0.58 \mu\text{kat/L}$ (women) and $0.74 \mu\text{kat/L}$ (men), i.e. 34.8 and 44.4 U/L; (ii) a single threshold of 40 U/L in both sexes. Results are in Table S12.

Season and era. Season from month of sampling (winter = December–February; spring = March–May; summer = June–August; autumn = September–November). Era: pre-pandemic 2014–2019 versus 2020–2023.

S.2 The 6-record exposure-plausibility caveat

The shared cohort builder filters on 25(OH)D *is not missing before* applying physiological plausibility ranges. Consequently, 6 records entered the cohort with a 25(OH)D value that was subsequently set to missing because it lay outside the plausible range of 1–500 nmol/L. **Cohort-level counts therefore use $n = 31,607$ while every analysis involving the exposure uses $n = 31,601$.** This is not an error and not a discrepancy to be reconciled; it is documented here so that a reader recomputing our numbers is not surprised by it. FIB-4 analyses use smaller n again: FIB-4 is computable in 30,232 records (a platelet count is additionally required), and analyses that involve *both* FIB-4 and the exposure therefore run on **30,226** records.

S.3 Corrections applied after independent verification

The analysis was independently re-derived through a separate code path before submission. That verification reproduced the STROBE cascade exactly and every Spearman coefficient to $|\Delta\rho| \leq 0.00024$. It identified two defects in the shared data layer; a second audit round subsequently identified two more. All four were corrected and the entire analysis was re-run; **every number in this article comes from the corrected run.** The defects and their consequences are recorded here in the interest of full transparency, including the one whose correction removed a statistically significant finding.

1. **Binary flags did not preserve missingness.** Flags were constructed as $(x > \text{cutoff})$, and because $\text{NaN} > \text{cutoff}$ evaluates to `False` in the underlying numeric library, records in which the variable could not be computed were coded as *below* the threshold rather than as missing. For FIB-4 this mis-assigned the 1,375 records without a platelet count to the low-risk group. Correcting it changes the prevalence of high-risk FIB-4 from 14.7% of all 31,607 records to **15.4% of the 30,232 records in which FIB-4 is computable**. The same defect affected the $\text{eGFR} < 60$ flag.
2. **Deduplication returned the first non-null value per column, not the first record.** The first-record-per-patient-per-year rule was implemented with a grouped `.first()` operation which, in

the library used, returns the first non-missing value *column by column*. Where a patient had more than one qualifying record in a calendar year (4,179 patient-years), this could splice a platelet count or an albumin value from one visit onto aminotransferases from another — assembling a laboratory profile that no visit ever produced. Correcting it to take the true first row changes 164 albumin cells, 75 platelet cells, 58 glucose cells, 28 creatinine cells and 2 exposure cells, and moves the number of computable FIB-4 values from 30,307 to 30,232. Every downstream estimate changed only in the third decimal place (e.g. rho(25(OH)D, FIB-4) $-0.0808 \rightarrow -0.0807$), but the reported n values changed and are now correct.

3. **The same splicing defect in the one-record-per-patient sensitivity cohort.** The correction in item 2 was applied to the primary cohort but an independent second occurrence, in the construction of the patient-level sensitivity cohort of Supplementary Table S11, was missed in the first audit round. It supplied platelet counts from later visits to 208 patients whose earliest record had none, fabricating a FIB-4 value for them; the FIB-4-evaluable sample was thereby inflated from 21,058 to 21,266. **Correcting it removed a statistically significant result:** the fully adjusted patient-level odds ratio for FIB-4 > 2.67 moves from 0.963 (0.929–0.997, $p = 0.033$) to 0.967 (0.933–1.002, $p = 0.063$). We draw attention to this because the correction happens to favour the deflationary conclusion of this paper, and it should be scrutinised accordingly. It does not affect the primary analysis, which is unchanged to every reported digit.
4. **Category boundaries were closed on the wrong side.** The 25(OH)D and FIB-4 category variables were cut with right-closed intervals, so that the label “Deficient (<30)” in fact contained records at exactly 30 nmol/L and “Optimal (≥ 75)” in fact excluded records at exactly 75 nmol/L — the printed labels contradicted the code, and the strata counts contradicted the text, which computed < 30 correctly throughout. Because 25(OH)D is reported to the nearest whole nmol/L, the boundaries are densely populated. Correcting the interval closure moved **350 records at exactly 30 nmol/L from “deficient” to “insufficient”, 377 records at exactly 50 from “insufficient” to “sufficient”, and 364 records at exactly 75 from “sufficient” to “optimal”** — in each case into the stratum whose label they had always satisfied. The strata sizes change from 5,430 / 7,082 / 10,032 / 9,057 to **5,080 / 7,055 / 10,045 / 9,421**. This affects Table 1, Supplementary Tables S3–S5 and Figure 3. **Every stratified contrast became slightly stronger** — for example the Kruskal–Wallis ϵ^2 for FIB-4 rose from 0.0185 to 0.0198, the deficient-versus-optimal Cliff’s delta for FIB-4 from 0.180 to 0.193, and the prevalence of high-risk FIB-4 in the deficient stratum from 28.1% to 29.0%. It does not affect any correlation, regression or interaction estimate, none of which use the categories.

The direction of these corrections, stated plainly. Three of the four corrections moved results in the direction that favours this paper’s own argument: correction 3 removed a statistically significant finding that we had been reporting as support (Supplementary Table S11), and correction 4 slightly strengthened every stratified contrast between vitamin D categories, including the FIB-4 contrast on which the paper’s descriptive section rests. We record this because a reader is entitled to see that the corrections were not neutral and that we disclosed it rather than banking the benefit silently. None of the four touched the primary analysis: the crude and adjusted correlation, regression and interaction estimates that carry the paper’s conclusion are unchanged to every reported digit, and the primary cohort size (31,607 records from 22,103 patients) is identical under the defective and corrected code.

The retired FIB-4 availability anchor. Related to correction 2, the value 30,307 was used during development as the expected count of records with a computable FIB-4. It has been retired: it was itself a product of the column-wise de-duplication, which supplied platelet counts from later visits to 75 records that had none at the index visit. The correct count is **30,232**, the number of cohort records that actually carry a platelet result at the index visit. The provenance of this change is recorded in the analysis code.

Both the original and the corrected analysis scripts, together with the independent verification script, are included in the code deposit cited in the Data availability statement.

S.4 Software

Python 3 with pandas, NumPy, SciPy (spearmanr, kruskal, mannwhitneyu, chi2), statsmodels (OLS with HC3 covariance, Logit) and matplotlib. All figures are programmatic output from the study data; no image-generation tool of any kind was used. The analysis is deterministic apart from bootstrap confidence intervals for medians, which use a fixed random seed.

Supplementary Table S1. Missing-data and co-testing (collider) analysis

For each analyte that may be co-ordered with 25(OH)D but is not required by the cohort rule, this table compares the exposure distribution between records in which the analyte was and was not measured. A non-zero Cliff's delta means that the decision to order the additional test is not independent of vitamin D status — the empirical signature of the selection (collider) problem discussed in the main text. Two-tailed Mann–Whitney p values.

Analyte	n available	% available	Median 25(OH)D if available	Median 25(OH)D if missing	Cliff's δ	p	% deficient if available	% deficient if missing
Creatinine ($\mu\text{mol/L}$)	31,253	98.9	59	57	+0.013	0.684	16.09	14.12
Leucocytes ($10^9/\text{L}$)	30,256	95.7	58	61	−0.024	0.129	16.11	15.25
Haemoglobin (g/L)	30,256	95.7	58	61	−0.024	0.129	16.11	15.25
Platelets ($10^9/\text{L}$)	30,232	95.6	58	60	−0.019	0.237	16.10	15.56
Glucose (mmol/L)	29,466	93.2	58	67	−0.149	1.0×10^{-30}	16.48	10.42
Albumin (g/L)	26,728	84.6	58	60	−0.023	0.011	16.32	14.74
Total protein (g/L)	22,570	71.4	59	58	+0.018	0.012	15.36	17.86
Calcium (mmol/L)	21,769	68.9	59	57	+0.045	2.3×10^{-10}	15.17	18.07
Urate ($\mu\text{mol/L}$)	14,064	44.5	61	57	+0.083	7.1×10^{-37}	13.57	18.08

Interpretation. Co-ordering is clearly non-random in the exposure for glucose and urate and mildly so for calcium, total protein and albumin. Because entry into this cohort itself requires that both 25(OH)D and a liver panel were ordered, the sample is conditioned on a variable (the decision to test) that is downstream of both illness and vitamin D status — i.e. a collider [33]. Neither the magnitude nor the direction of the residual adjusted association is therefore formally identified.

Supplementary Table S2. Spearman correlations with 25(OH)D stratified by sex, with formal tests of the sex difference

Fisher z test of two independent correlations. Note that a difference in two rho values is claimed only where this test supports it; the corresponding interaction tests on the modelled scale are in Table S7.

Analyte	rho (women)	n (women)	rho (men)	n (men)	z	p for difference
AST (U/L)	−0.039	18,431	−0.123	13,170	7.43	1.1×10^{-13}
ALT (U/L)	−0.049	18,431	−0.053	13,170	0.44	0.663
GGT (U/L)	−0.121	18,431	−0.156	13,170	3.21	1.3×10^{-3}
ALP (U/L)	−0.123	18,431	−0.127	13,170	0.35	0.729
De Ritis ratio	+0.007	18,431	−0.085	13,170	8.03	1.1×10^{-15}
FIB-4 index	−0.031	17,576	−0.132	12,650	8.78	$< 1 \times 10^{-16}$
Platelets ($10^9/L$)	+0.014	17,576	+0.048	12,650	−2.86	4.2×10^{-3}
Albumin (g/L)	+0.161	15,677	+0.217	11,049	−4.67	3.0×10^{-6}
Age (years)	−0.019	18,431	−0.100	13,170	7.15	8.6×10^{-13}
eGFR	−0.026	18,225	−0.028	13,022	0.20	0.845

Supplementary Table S3. Dunn post-hoc comparisons between 25(OH)D categories

All six pairwise comparisons for each outcome, with Dunn z statistics, raw p values and Bonferroni-corrected p values (six comparisons per outcome). D = deficient (<30 nmol/L), I = insufficient (30–<50), S = sufficient (50–<75), O = optimal (≥ 75). The two-tailed p value is computed as $2[1 - \Phi(|z|)]$, which underflows to exactly zero in double precision once $|z|$ exceeds approximately 8.3; such entries are shown as “ $<1 \times 10^{-300}$ ” and should be read as “smaller than this computation can resolve”, not as an estimate of the true tail area. The z statistic itself is exact and is the quantity to compare across rows.

Outcome	Comparison	z	p (raw)	p (Bonferroni)
AST	D vs I	11.63	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
AST	D vs S	18.55	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
AST	D vs O	17.18	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
AST	I vs S	6.79	1.2×10^{-11}	6.9×10^{-11}
AST	I vs O	5.40	6.6×10^{-8}	4.0×10^{-7}
AST	S vs O	−1.42	0.156	0.933
ALT	D vs I	0.26	0.795	1.000
ALT	D vs S	6.52	7.2×10^{-11}	4.3×10^{-10}
ALT	D vs O	9.99	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
ALT	I vs S	6.92	4.7×10^{-12}	2.8×10^{-11}
ALT	I vs O	10.74	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
ALT	S vs O	4.30	1.7×10^{-5}	1.0×10^{-4}
GGT	D vs I	13.02	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
GGT	D vs S	24.50	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
GGT	D vs O	26.63	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
GGT	I vs S	11.73	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$

Outcome	Comparison	z	p (raw)	p (Bonferroni)
GGT	I vs O	14.23	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
GGT	S vs O	2.91	3.6×10^{-3}	0.022
ALP	D vs I	12.39	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
ALP	D vs S	21.75	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
ALP	D vs O	22.92	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
ALP	I vs S	9.42	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
ALP	I vs O	10.85	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
ALP	S vs O	1.71	0.088	0.529
De Ritis	D vs I	14.83	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
De Ritis	D vs S	15.50	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
De Ritis	D vs O	8.93	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
De Ritis	I vs S	-0.38	0.703	1.000
De Ritis	I vs O	-7.45	9.0×10^{-14}	5.4×10^{-13}
De Ritis	S vs O	-7.77	7.8×10^{-15}	4.7×10^{-14}
FIB-4	D vs I	18.92	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
FIB-4	D vs S	23.95	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
FIB-4	D vs O	18.06	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
FIB-4	I vs S	4.16	3.2×10^{-5}	1.9×10^{-4}
FIB-4	I vs O	-2.14	0.032	0.194
FIB-4	S vs O	-6.84	7.7×10^{-12}	4.6×10^{-11}
Platelets	D vs I	-7.56	4.0×10^{-14}	2.4×10^{-13}
Platelets	D vs S	-9.99	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
Platelets	D vs O	-8.26	2.2×10^{-16}	1.3×10^{-15}
Platelets	I vs S	-2.13	0.034	0.201
Platelets	I vs O	-0.30	0.764	1.000
Platelets	S vs O	1.97	0.049	0.293
Albumin	D vs I	-18.55	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
Albumin	D vs S	-29.12	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
Albumin	D vs O	-31.13	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
Albumin	I vs S	-10.30	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
Albumin	I vs O	-12.71	$<1 \times 10^{-300}$	$<1 \times 10^{-300}$
Albumin	S vs O	-2.76	5.7×10^{-3}	0.034

Note the FIB-4 rows. The insufficient-versus-optimal comparison is not supported after correction ($p = 0.194$) and the sufficient-versus-optimal comparison is *negative*, i.e. the optimal group has the higher FIB-4. The exposure gradient in FIB-4 is not monotone; the optimal group is older than the two middle groups and includes supplemented patients.

Supplementary Table S4. Kruskal–Wallis tests across the four 25(OH)D categories

Three degrees of freedom throughout. ε^2 is the epsilon-squared effect size (proportion of variance in the ranks shared with category).

Outcome	n	H	p	ε^2
AST (U/L)	31,601	393.3	6.22×10^{-85}	0.0124
ALT (U/L)	31,601	163.6	3.02×10^{-35}	0.0051
GGT (U/L)	31,601	863.0	9.51×10^{-187}	0.0272

Outcome	n	H	<i>p</i>	ε^2
ALP (U/L)	31,601	637.6	7.16×10^{-138}	0.0201
De Ritis ratio	31,601	302.3	3.20×10^{-65}	0.0095
FIB-4 index	30,226	602.4	3.10×10^{-130}	0.0198
Platelets (10 ⁹ /L)	30,226	105.9	8.46×10^{-23}	0.0034
Albumin (g/L)	26,726	1,131.4	5.64×10^{-245}	0.0422

Every ε^2 is below 0.05; the largest belongs to albumin, not to any hepatobiliary analyte.

Supplementary Table S5. Outcomes by 25(OH)D category, and Cliff's delta for the extreme contrast

Panel A — outcomes by category (median with bootstrap 95% CI for FIB-4; Wilson 95% CI for proportions).

25(OH)D category	n	Median FIB-4		FIB-4 > 2.67, %		Elevated ALT, %		Median		Median		Median		Median	
		[95% CI]	n with FIB-4	[95% CI]	%	[95% CI]	%	GGT (U/L)	ALP (U/L)	Median age (y)	albu- min (g/L)	Median platelets (10 ⁹ /L)			
Deficient (<30)	5,080	1.56 [1.51–1.62]	4,866	28.98 [27.72–30.27]	47.13 [45.76–48.50]	31.2	81.0	65.2	41.0	224					
Insufficient (30–<50)	7,055	1.17 [1.15–1.19]	6,770	15.18 [14.35–16.06]	46.21 [45.05–47.37]	24.0	73.8	58.9	44.0	235					
Sufficient (50–<75)	10,045	1.13 [1.11–1.15]	9,582	11.40 [10.78–12.05]	42.27 [41.31–43.24]	20.4	70.2	59.0	44.0	236					
Optimal (≥75)	9,421	1.22 [1.20–1.25]	9,008	12.33 [11.67–13.03]	40.72 [39.73–41.71]	19.8	69.6	61.3	44.0	233					

Panel B — Cliff's delta, deficient versus optimal. Positive delta = higher values in the deficient group.

Outcome	δ	<i>p</i>	n (deficient)	n (optimal)
GGT	+0.265	1.0×10^{-152}	5,080	9,421
ALP	+0.228	3.0×10^{-114}	5,080	9,421
FIB-4	+0.193	1.6×10^{-78}	4,866	9,008
AST	+0.173	8.5×10^{-67}	5,080	9,421
ALT	+0.099	1.0×10^{-22}	5,080	9,421
De Ritis ratio	+0.096	1.6×10^{-21}	5,080	9,421
Platelets	−0.087	3.3×10^{-17}	4,866	9,008
Albumin	−0.338	5.8×10^{-213}	4,361	7,953

By Cohen's conventional benchmarks ($|\delta| < 0.147$ negligible, < 0.33 small, < 0.474 medium), every hepatobiliary contrast is small; the albumin contrast is the largest in the panel and is the only one that crosses out of the small band into the medium band ($\delta = -0.338$).

Supplementary Table S6. Full five-step model output for all six continuous outcomes

Percentage change in the outcome per +25 nmol/L 25(OH)D with 95% CI, exact p, and attenuation relative to the crude model. Ordinary least squares on the natural log of the outcome, HC3 robust standard errors. Model 1 crude; 2 + age, sex; 3 + albumin; 4 + season, era; 5 + eGFR.

Outcome	Model	n	% change [95% CI]	<i>p</i>	Attenuation
log FIB-4	1	30,226	-5.21 [-5.90, -4.52]	9.7×10^{-47}	—
log FIB-4	2	30,226	-2.78 [-3.32, -2.24]	4.4×10^{-23}	47.3%
log FIB-4	3	25,588	-0.39 [-0.94, +0.17]	0.168	92.7%
log FIB-4	4	25,588	-1.21 [-1.77, -0.65]	2.2×10^{-5}	77.2%
log FIB-4	5	25,504	-1.06 [-1.62, -0.49]	2.4×10^{-4}	80.2%
log GGT	1	31,601	-10.11 [-10.98, -9.23]	9.4×10^{-102}	—
log GGT	2	31,601	-8.10 [-8.94, -7.25]	3.8×10^{-72}	20.7%
log GGT	3	26,726	-4.65 [-5.55, -3.75]	2.8×10^{-23}	55.3%
log GGT	4	26,726	-5.01 [-5.92, -4.09]	1.7×10^{-25}	51.8%
log GGT	5	26,570	-4.59 [-5.50, -3.66]	1.8×10^{-21}	56.0%
log ALP	1	31,601	-3.98 [-4.40, -3.57]	8.9×10^{-77}	—
log ALP	2	31,601	-3.62 [-4.03, -3.22]	8.0×10^{-66}	9.2%
log ALP	3	26,726	-2.21 [-2.64, -1.78]	2.7×10^{-23}	45.0%
log ALP	4	26,726	-2.33 [-2.77, -1.89]	1.9×10^{-24}	42.0%
log ALP	5	26,570	-2.42 [-2.87, -1.97]	1.3×10^{-25}	39.8%
log ALT	1	31,601	-2.79 [-3.34, -2.24]	1.2×10^{-22}	—
log ALT	2	31,601	-2.17 [-2.71, -1.63]	6.3×10^{-15}	22.4%
log ALT	3	26,726	-1.82 [-2.40, -1.23]	1.2×10^{-9}	35.2%
log ALT	4	26,726	-2.23 [-2.82, -1.64]	2.1×10^{-13}	20.3%
log ALT	5	26,570	-1.53 [-2.12, -0.94]	3.7×10^{-7}	45.4%

Outcome	Model	n	% change [95% CI]	<i>p</i>	Attenuation
log AST	1	31,601	−3.68 [−4.21, −3.15]	3.4×10^{-40}	—
log AST	2	31,601	−2.89 [−3.41, −2.37]	1.5×10^{-26}	21.8%
log AST	3	26,726	−1.03 [−1.58, −0.47]	3.0×10^{-4}	72.5%
log AST	4	26,726	−2.03 [−2.58, −1.48]	1.0×10^{-12}	45.4%
log AST	5	26,570	−1.28 [−1.82, −0.73]	4.8×10^{-6}	65.6%
log De Ritis	1	31,601	−0.92 [−1.30, −0.54]	2.3×10^{-6}	—
log De Ritis	2	31,601	−0.73 [−1.10, −0.37]	7.6×10^{-5}	20.2%
log De Ritis	3	26,726	+0.80 [+0.42, +1.19]	4.3×10^{-5}	186.6% (sign reversal)
log De Ritis	4	26,726	+0.20 [−0.18, +0.59]	0.304	122.1% (sign reversal)
log De Ritis	5	26,570	+0.26 [−0.14, +0.65]	0.200	127.7% (sign reversal)

Full covariate output, fully adjusted FIB-4 model (n = 25,504, log scale): intercept 0.114 (SE 0.051); 25(OH)D per 25 nmol/L −0.0106 (SE 0.0029, $p = 2.4 \times 10^{-4}$); age +0.0253 per year (SE 0.00025); female −0.1425 (SE 0.0075); albumin −0.0326 per g/L (SE 0.00098); spring +0.0213; summer −0.0091; autumn +0.0641; pandemic era +0.1450 (SE 0.0072); eGFR +0.00070 per mL/min/1.73 m² (SE 0.00017). Model R² = 0.485.

Same-rows attenuation check. Refitting the crude model on the 25,588 rows with albumin measured gives a crude coefficient of −5.25% (rho for 25(OH)D–FIB-4 on those rows −0.078), against a fully adjusted −1.06%: **79.8% attenuation on essentially identical data.** The attenuation is therefore not an artefact of the changing sample size.

Age-free FIB-4 surrogate, AST/(platelets × $\sqrt{\text{ALT}}$): crude rho −0.065 (95% CI −0.076 to −0.054; n = 30,226); cascade −3.94% → −2.87% → −0.43% → −1.27% → **−1.13%**, i.e. 71.7% attenuation. The residual association is therefore not an artefact of the age term inside FIB-4.

Supplementary Table S7. Likelihood-ratio tests for effect modification

Each test compares the fully adjusted model with and without the 25(OH)D × modifier interaction term; 1 degree of freedom throughout. **No subgroup claim in this paper rests on anything other than these tests.**

Outcome	Modifier	n	LR χ^2	<i>p</i>	Supported?
log FIB-4	Sex	25,588	19.42	1.0×10^{-5}	Yes
log FIB-4	Age (continuous)	25,588	11.97	5.4×10^{-4}	Yes
log FIB-4	Era (pandemic)	25,588	0.25	0.616	No
log GGT	Sex	26,726	5.79	0.016	Weak

Outcome	Modifier	n	LR χ^2	<i>p</i>	Supported?
log ALP	Sex	26,726	0.19	0.664	No
log ALP	Age (continuous)	26,726	3.07	0.080	No
FIB-4 > 2.67	Sex	25,588	1.72	0.190	No
FIB-4 > 2.67	Age (continuous)	25,588	13.25	2.7×10^{-4}	Yes
Elevated ALT	Sex	26,726	2.33	0.127	No

The sex modification is scale-dependent: it is supported on the continuous log FIB-4 scale but not on the dichotomised FIB-4 > 2.67 scale. The sex-specific statement in the main text is therefore reported as provisional and hypothesis-generating. Age modification is supported on both scales.

Supplementary Table S8. Subgroup estimates from fully adjusted models

These estimates are interpretable only in the light of Table S7. They are reported for completeness and must not be read as demonstrating a subgroup difference in the absence of a supported interaction test.

Subgroup	n (linear)	% change in FIB-4 per +25 nmol/L [95% CI]	n (logistic)	OR for FIB-4 > 2.67 [95% CI]
Women	14,908	−0.05 [−0.71, +0.61]	14,908	0.966 [0.925, 1.009]
Men	10,596	−2.55 [−3.55, −1.54]	10,596	0.904 [0.862, 0.947]
Age ≤ 60 y	11,993	+0.37 [−0.39, +1.14]	11,993	0.973 [0.903, 1.049]
Age > 60 y	13,511	−2.42 [−3.22, −1.60]	13,511	0.920 [0.889, 0.953]
Pre-pandemic 2014–2019	9,405	−1.45 [−2.37, −0.51]	9,405	0.858 [0.798, 0.924]
Pandemic era 2020–2023	16,099	−0.91 [−1.60, −0.21]	16,099	0.960 [0.927, 0.995]

The era contrast is a good example of why the interaction test governs: the two era-specific odds ratios look different, but the formal interaction test on the continuous scale is not supported ($p = 0.616$), so the apparent difference should not be claimed.

Supplementary Table S9. Seasonal summary and seasonally stratified correlations

Panel A — summary by season. Kruskal–Wallis across seasons: 25(OH)D $H = 1,258.8$, $p = 1.3 \times 10^{-272}$, $\epsilon^2 = 0.0397$ ($n = 31,601$); FIB-4 $H = 164.1$, $p = 2.4 \times 10^{-35}$, $\epsilon^2 = 0.0053$ ($n = 30,232$); GGT $H = 83.4$, $p = 5.8 \times 10^{-18}$, $\epsilon^2 = 0.0025$; ALP $H = 45.0$, $p = 9.1 \times 10^{-10}$, $\epsilon^2 = 0.0013$.

Season	n	Median 25(OH)D (nmol/L)	% deficient	Median FIB-4	% FIB-4 > 2.67	Median GGT (U/L)	Median ALP (U/L)
Winter	9,600	52	21.30	1.280	17.69	23.4	73.8
Spring	8,377	53	20.82	1.178	13.58	21.6	72.0
Summer	5,517	65	8.32	1.127	10.94	21.0	71.4
Autumn	8,113	65	10.26	1.264	17.41	22.8	71.4

Panel B — correlations with 25(OH)D within season.

Season	Outcome	n	rho [95% CI]	<i>p</i>
Winter	FIB-4	9,193	−0.050 [−0.071, −0.030]	1.4×10^{-6}
Winter	ALP	9,599	−0.106 [−0.126, −0.087]	1.4×10^{-25}
Winter	GGT	9,599	−0.152 [−0.171, −0.132]	1.1×10^{-50}
Spring	FIB-4	8,026	−0.030 [−0.052, −0.009]	6.3×10^{-3}
Spring	ALP	8,376	−0.135 [−0.156, −0.114]	2.5×10^{-35}
Spring	GGT	8,376	−0.163 [−0.183, −0.142]	1.1×10^{-50}
Summer	FIB-4	5,255	−0.145 [−0.171, −0.118]	4.9×10^{-26}
Summer	ALP	5,514	−0.146 [−0.172, −0.120]	1.0×10^{-27}
Summer	GGT	5,514	−0.167 [−0.193, −0.142]	6.4×10^{-36}
Autumn	FIB-4	7,752	−0.133 [−0.155, −0.111]	7.3×10^{-32}
Autumn	ALP	8,112	−0.139 [−0.161, −0.118]	1.8×10^{-36}
Autumn	GGT	8,112	−0.150 [−0.171, −0.128]	6.8×10^{-42}

Note that the FIB-4 correlation is three to five times larger in summer and autumn than in winter and spring — a further reminder that the crude association is entangled with the seasonal structure of both testing behaviour and exposure.

Supplementary Table S9b. Monthly summary

Month	n	Median 25(OH)D (nmol/L)	Median FIB-4	% FIB-4 > 2.67
January	4,034	54	1.254	16.08
February	2,828	49	1.226	15.90
March	3,205	52	1.187	13.88
April	2,276	51	1.222	15.86
May	2,896	55	1.146	11.44
June	2,404	61	1.153	9.97
July	1,533	67	1.130	11.55
August	1,580	72	1.079	11.78
September	2,466	71	1.228	15.24
October	2,646	66	1.214	16.11
November	3,001	60	1.347	20.33
December	2,738	53	1.381	21.90

Supplementary Table S10. Annual trend, 2014–2023

Spearman rho of calendar year with 25(OH)D = +0.107 (95% CI +0.096 to +0.118; $n = 31,601$; $p = 4.7 \times 10^{-81}$); with FIB-4 = +0.041 (+0.029 to +0.052; $n = 30,232$; $p = 1.4 \times 10^{-12}$). Era contrast (2020–2023 vs 2014–2019): Cliff's $\delta = -0.126$ for 25(OH)D ($p = 4.1 \times 10^{-76}$) and -0.113 for FIB-4 ($p = 2.5 \times 10^{-59}$).

Year	n	Median 25(OH)D	% deficient	Median FIB-4	% FIB-4 > 2.67	Median GGT (U/L)	Median ALP (U/L)	Median albumin (g/L)
2014	1,200	52.5	13.00	1.190	8.87	22.8	74.4	43
2015	1,315	53	13.46	1.344	13.78	21.6	72.0	43
2016	1,754	53	22.86	1.204	11.02	23.4	71.1	43
2017	2,061	54	19.89	1.109	10.30	22.2	70.2	44
2018	2,089	55	19.44	1.101	10.73	21.0	70.8	44
2019	2,713	55	19.17	1.037	7.06	19.8	69.6	45
2020	3,271	60	15.10	1.178	18.02	22.2	70.2	44
2021	6,353	59	15.28	1.440	24.17	24.6	69.0	43
2022	5,623	61	16.18	1.331	18.44	22.2	74.4	43
2023	5,228	64	12.15	1.129	10.92	22.2	78.0	45

The 2021 excursion in FIB-4 (median 1.44; 24.2% high-risk) is large and is the reason calendar era was retained as a covariate. We make no claim about its cause; plausible contributions include changes in referral patterns, in who was tested, and in the acuity of the tested population during the pandemic, none of which we can distinguish in these data.

Supplementary Table S11. Sensitivity analysis — one record per patient

The primary analysis uses the first record per patient per calendar year (31,607 records from 22,103 patients; 78.3% of patients contribute exactly one record, maximum 10). This sensitivity analysis retains only the earliest record for each patient ($n = 22,103$), eliminating within-patient clustering entirely.

Panel A — correlations with 25(OH)D

Analyte	n	rho [95% CI]	p	Primary-cohort rho
GGT (U/L)	22,101	−0.155 [−0.168, −0.143]	1.6×10^{-119}	−0.156
ALP (U/L)	22,101	−0.148 [−0.161, −0.135]	5.3×10^{-108}	−0.131
FIB-4 index	21,058	−0.126 [−0.139, −0.113]	2.3×10^{-75}	−0.081
AST (U/L)	22,101	−0.084 [−0.097, −0.071]	3.6×10^{-36}	−0.090
ALT (U/L)	22,101	−0.042 [−0.055, −0.029]	3.2×10^{-10}	−0.071
Albumin (g/L)	18,238	+0.229 [+0.215, +0.243]	1.3×10^{-215}	+0.189

Panel B — adjustment cascade, log FIB-4 (% change per +25 nmol/L)

Model	n	% change [95% CI]	<i>p</i>	Attenuation
1 Crude	21,058	−6.81 [−7.69, −5.93]	5.5×10^{-49}	—
2 + Age, sex	21,058	−2.69 [−3.35, −2.01]	9.9×10^{-15}	61.4%
3 + Albumin	17,397	+0.43 [−0.26, +1.13]	0.222	106.1%
4 + Season, era	17,397	−0.36 [−1.06, +0.34]	0.314	94.9%
5 Full (+ eGFR)	17,335	−0.40 [−1.10, +0.30]	0.261	94.3%

Panel C — adjustment cascade, FIB-4 > 2.67 (odds ratio per +25 nmol/L)

Model	n / events	OR [95% CI]	<i>p</i>
1 Crude	21,058 / 3,875	0.769 [0.745, 0.794]	2.1×10^{-59}
2 + Age, sex	21,058 / 3,875	0.846 [0.819, 0.874]	5.4×10^{-24}
3 + Albumin	17,397 / 3,406	0.999 [0.966, 1.034]	0.969
4 + Season, era	17,397 / 3,406	0.968 [0.934, 1.003]	0.070
5 Full (+ eGFR)	17,335 / 3,395	0.967 [0.933, 1.002]	0.063

Panel D — adjustment cascade, elevated ALT (odds ratio per +25 nmol/L)

Model	n / events	OR [95% CI]	<i>p</i>
1 Crude	22,101 / 9,906	0.961 [0.942, 0.981]	1.4×10^{-4}
5 Full (+ eGFR)	18,127 / 8,172	0.971 [0.948, 0.995]	0.016

Interpretation. At the patient level the crude association is *larger* than in the primary analysis (because the earliest record is more often the sickest presentation), but adjustment removes it: the fully adjusted binary estimate has a confidence interval that includes 1 (OR 0.967, 0.933–1.002, $p = 0.063$), and the continuous fully adjusted estimate is likewise not distinguishable from zero (−0.40% per +25 nmol/L, $p = 0.261$). **The sensitivity analysis therefore does not confirm the primary adjusted association at the patient level.** We report this plainly rather than resolving it in the paper’s favour. Two readings are compatible with the data and we cannot separate them here: the patient-level analysis discards roughly 30% of the records and is correspondingly less precise, so a genuine effect of this size would not reliably reach conventional significance; equally, the primary estimate may be partly sustained by repeated sampling of the same patients. Either way the point estimate is of the same negligible magnitude as the primary one, and the paper’s conclusion — that the association is too small to be clinically actionable — does not depend on which reading is correct.

History of this table. An earlier version reported the patient-level adjusted association as null; that was traced to the binary-flag defect described in Supplementary Methods S.3 and corrected, which made it significant (OR 0.963, 0.929–0.997, $p = 0.033$). A subsequent audit found a third, independent instance of the record-splicing defect of Supplementary Methods S.3 in the construction of this sensitivity cohort: the earliest record per patient was assembled with a column-wise “first non-missing value” rule rather than by taking the earliest row whole, which supplied platelet counts from later visits to 208 records that had none on the index date and so fabricated a FIB-4 for them. With the earliest row taken whole, the FIB-4-evaluable sample falls from 21,266 to 21,058 and the fully adjusted estimate returns to the null. The values in this table are from the corrected construction.

Supplementary Table S12. ALT threshold sensitivity, clustering, reclassification arithmetic and E-value

Panel A — elevated ALT under three threshold definitions. Fully adjusted = model 5 (age, sex, albumin, season, era, eGFR).

Definition	Threshold (women / men)	n evaluable	n elevated (%)	Crude OR [95% CI]	Fully adjusted OR [95% CI]
ACG “true-normal” (primary) [8]	25 / 33 U/L	31,607	13,736 (43.46)	0.938 [0.922, 0.954]	0.940 [0.921, 0.959]
Routine IFCC 37 °C limits	0.58 / 0.74 μkat/L (34.8 / 44.4 U/L)	31,607	7,120 (22.53)	0.896 [0.877, 0.916]	0.940 [0.917, 0.964]
Conventional single limit	40 / 40 U/L	31,607	6,977 (22.07)	0.874 [0.854, 0.893]	0.927 [0.903, 0.951]

The fully adjusted odds ratio is between 0.927 and 0.940 under all three definitions; the conclusion does not depend on the threshold. The 43.5% prevalence under the primary definition is a property of the deliberately low “true-normal” limits and is stated as such in the main text.

Panel B — clustering of the primary cohort. 31,607 records from 22,103 patients; mean 1.43 records per patient; maximum 10.

Records contributed	Patients	% of patients
1	17,297	78.26
2	2,639	11.94
3	1,034	4.68
4	486	2.20
5	281	1.27
6	158	0.71
7	100	0.45
8	58	0.26
9	31	0.14
10	19	0.09

Panel C — what the adjusted estimate means for a patient. In the fully adjusted model, +25 nmol/L of 25(OH)D corresponds to a **−1.056%** multiplicative change in FIB-4. At the cohort median FIB-4 of 1.219 (IQR 0.758–1.986) this is an absolute shift of **−0.0129 FIB-4 units**. Only **138 of 30,232 records (0.46%)** lie close enough to the 2.67 decision threshold for a shift of that magnitude to move them across it. No individual patient’s fibrosis-risk classification would change.

Panel D — E-value for unmeasured confounding [34]. Marginal prevalence of FIB-4 > 2.67 = 15.35%. Converting the fully adjusted odds ratio 0.935 (0.906–0.966) to the approximate risk-ratio scale gives RR ≈ 0.945 (upper limit 0.971). The corresponding **E-value is 1.31 at the point estimate and 1.21 at the confidence limit**: an unmeasured confounder associated with both 25(OH)D and high-risk FIB-4 by a risk ratio of about 1.3, above and beyond the measured covariates, would fully explain the residual association. Given that measured age carries an odds ratio of 1.061 *per year* and calendar era an odds ratio of 1.87, an unmeasured confounder of that strength is entirely plausible.

Supplementary Table S13. STROBE checklist

Manuscript: Serum 25-hydroxyvitamin D, liver enzymes and FIB-4 fibrosis risk: a statistically robust but clinically negligible association from 31,607 adult laboratory records **Authors:** Osredkar J, Godnov U, Jerin A, Siuka D, Drobne D **Reporting guideline:** von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. *Lancet* **370**, 1453–1457 (2007).

#	Item	Recommendation	Where reported (section pointer — layout-independent)
Title and abstract			
1a	Title/abstract	Indicate the study's design with a commonly used term in the title or the abstract	Title (no design term); Abstract (no design term). The design is named only in Methods §Study design (“retrospective cross-sectional analysis of routine clinical laboratory data”) and in Discussion, “Other limitations” (“The design is cross-sectional”)
1b	Title/abstract	Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract (single unstructured paragraph) — setting, n, crude and adjusted effect sizes with 95% CIs, the attenuation, and the explicit conclusion that the association is “statistically unambiguous, clinically negligible”
Introduction			
2	Background/rationale	Explain the scientific background and rationale for the investigation	Introduction, paragraphs 1–5 (vitamin D and liver disease; FIB-4 in routine practice; the failure of the interventional evidence; 25(OH)D as a consequence of ill health; the effect-size/collider argument)

#	Item	Recommendation	Where reported (section pointer — layout-independent)
3	Objectives	State specific objectives, including any prespecified hypotheses	Introduction, final paragraph — three explicitly numbered questions (magnitude; attenuation; clinical relevance), with the answer to the third stated in advance
Methods			
4	Study design	Present key elements of study design early in the paper	Methods §Study design, setting and data source, first sentence
5	Setting	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods §Study design — Institute of Clinical Chemistry and Biochemistry, University Medical Centre Ljubljana, Slovenia; 2 January 2014 to 29 December 2023
6a	Participants	Give the eligibility criteria, and the sources and methods of selection of participants	Methods §Cohort definition (five prespecified steps, plus the 6-record exposure-plausibility caveat); Results §Cohort; Figure 1 legend
6b	Participants	<i>(cohort/case-control only — not applicable to cross-sectional studies)</i>	Not applicable
7	Variables	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Methods §Variables and units (exposure, all derived variables, thresholds, units in both $\mu\text{kat/L}$ and U/L , FIB-4 > 2.67, ACG ALT limits, De Ritis ratio, eGFR, season, era); confounder set and effect modifiers defined in §Statistical analysis; Supplementary Methods S.1 gives every derivation formula

#	Item	Recommendation	Where reported (section pointer — layout-independent)
8	Data sources/measurement	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Data source: Methods §Study design, setting and data source (single laboratory information system, single tertiary centre, whole-database extract). Units and derived-variable definitions: Methods §Variables and units; Supplementary Methods S.1. The analyser platform and assay are not named in the manuscript.
9	Bias	Describe any efforts to address potential sources of bias	Methods §Statistical analysis (five-step adjustment cascade; complete-case analysis with a missing-data analysis; refit of the crude model on the complete-case rows to exclude sample-size change as the explanation; derived-variable artefact check; E-value); Results (the same-rows crude refit on the 25,588 albumin rows); Discussion, “The albumin adjustment is a bound, not a control” (over-adjustment, adjustment-for-the- exposure and collider bias, with the empirical co-testing analysis); Supplementary Table S1

#	Item	Recommendation	Where reported (section pointer — layout-independent)
10	Study size	Explain how the study size was arrived at	The study size is the entire eligible database; no a priori power calculation was performed. This study-size justification is not explicitly stated in the manuscript. Nearest content: Methods §Cohort definition (the rule that determines the size); Results §Cohort (full attrition); Figure 1 legend.
11	Quantitative variables	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Methods §Variables and units — 25(OH)D analysed continuously (per +25 nmol/L) <i>and</i> categorised at the IOM/Endocrine Society cut-points, with the reason given; §Statistical analysis justifies the log transformation of all skewed continuous outcomes and the non-parametric univariable methods
12a	Statistical methods	Describe all statistical methods, including those used to control for confounding	Methods §Statistical analysis — software, alpha, two-tailed tests, exact p values, normality justification, Spearman with Fisher-z intervals, Mann–Whitney with Cliff’s delta, and the prespecified five-step adjustment cascade with its covariate blocks and estimators

#	Item	Recommendation	Where reported (section pointer — layout-independent)
12b	Statistical methods	Describe any methods used to examine subgroups and interactions	Methods §Statistical analysis, “Effect modification” — likelihood-ratio tests of the exposure × modifier term in the fully adjusted model; subgroup estimates reported only alongside them. Results §Sex and age modify...; Supplementary Tables S7–S8
12c	Statistical methods	Explain how missing data were addressed	Methods §Statistical analysis (“Complete-case analysis was used; no imputation was performed”) plus the missing-data analysis; Supplementary Table S1; n reported for every model in Table 3; Discussion quantifies the missingness (albumin 15.4%, platelets 4.4%)
12d	Statistical methods	<i>(cross-sectional)</i> If applicable, describe analytical methods taking account of sampling strategy	Methods §Cohort definition — the primary unit is the first record per patient per calendar year; a one-record-per-patient analysis eliminating clustering entirely is reported (Results §Sensitivity...; Supplementary Table S11), and the clustering distribution is given in Results §Cohort and in Supplementary Table S12, Panel B

#	Item	Recommendation	Where reported (section pointer — layout-independent)
12e	Statistical methods	Describe any sensitivity analyses	Methods §Variables and units (prespecified ALT-threshold sensitivity analysis), §Statistical analysis (derived-variable artefact check; E-value) and §Negative-control outcomes (explicitly labelled as not prespecified); Results §Sensitivity and missing-data analyses and §How large a residual does this adjustment set produce where no effect can exist?; Supplementary Tables S11 (one record per patient), S12 (ALT thresholds, reclassification, E-value), S6 (same-rows attenuation, age-free FIB-4 surrogate), S14 (negative-control calibration, both analysis frames, binary and continuous, with a positive control)
Results 13a	Participants	Report numbers of individuals at each stage of the study	Results §Cohort, first paragraph; Figure 1 legend (107,900 → 107,856 → 82,235 → 37,456 → 31,607 records from 22,103 adults)

#	Item	Recommendation	Where reported (section pointer — layout-independent)
13b	Participants	Give reasons for non-participation at each stage	Figure 1 legend and Results §Cohort — exclusions are missing/implausible exposure, invalid demographics, age outside 18–100, and absent or implausible index analytes; no individual declined, since no participant was approached
13c	Participants	Consider use of a flow diagram	Figure 1 (STROBE participant flow).
14a	Descriptive data	Give characteristics of study participants and information on exposures and potential confounders	Table 1 (by 25(OH)D category); Results §Cohort narrative, including the two features of Table 1 that set up the confounding argument
14b	Descriptive data	Indicate number of participants with missing data for each variable of interest	Results §Cohort (FIB-4 computable in 30,232 of 31,607, 95.6%); Table 1 footnotes (FIB-4 denominator 30,232; creatinine/eGFR denominator 31,253); Results §Adjustment removes... (25,588 rows with albumin measured, used for the same-rows crude refit); Discussion (albumin missing in 15.4% of records, platelets in 4.4%); Supplementary Table S1 gives every analyte; n is stated for every model in Table 3 and Supplementary Table S6

#	Item	Recommendation	Where reported (section pointer — layout-independent)
15	Outcome data	Report numbers of outcome events or summary measures	Results §Cohort (4,641 of 30,232 records, 15.4%, exceeded FIB-4 2.67); Results §Adjustment removes... (13,736 of 31,607, 43.5%, flagged as elevated ALT); Table 1; Table 3 Panel B gives events per model
16a	Main results	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision. Make clear which confounders were adjusted for and why they were included	Table 3 (both panels) reports all five models side by side with 95% CIs and per-step attenuation; Figure 4 legend; Results §Adjustment removes most of the FIB-4 association. The confounder set and the reason for each covariate are stated in Methods §Statistical analysis
16b	Main results	Report category boundaries when continuous variables were categorised	Methods §Variables and units (25(OH)D <30 / 30–<50 / 50–<75 / ≥75 nmol/L; FIB-4 > 2.67; ALT > 25 U/L women, > 33 U/L men, with the prespecified higher-threshold sensitivity analysis; eGFR < 60; seasons; eras)

#	Item	Recommendation	Where reported (section pointer — layout-independent)
16c	Main results	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Results §Sensitivity and missing-data analyses, final paragraphs — the adjusted estimate corresponds to a shift of −0.013 FIB-4 units at the cohort median of 1.22, and only 138 of 30,232 records (0.46%) lie close enough to 2.67 for that to change classification; the E-value is 1.31 (1.21 at the confidence limit); Supplementary Table S12, Panel C
17	Other analyses	Report other analyses done — e.g. analyses of subgroups and interactions, and sensitivity analyses	Results §The FIB-4 signal is largely an age and health-status signal; §Sex and age modify...; §Season, era and secular trend; §Sensitivity and missing-data analyses; §How large a residual does this adjustment set produce where no effect can exist? (negative-control calibration, Table 3 Panel C); Supplementary Tables S2, S6–S12, S14
Discussion 18	Key results	Summarise key results with reference to study objectives	Discussion, opening paragraphs

#	Item	Recommendation	Where reported (section pointer — layout-independent)
19	Limitations	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion, “The albumin adjustment is a bound, not a control” (over-adjustment, adjustment-for-the-exposure and collider bias, each with direction and, where possible, magnitude — including the E-value of 1.31) and “Other limitations” (cross-sectional design; single-centre vitamin D-request cohort; absent PTH, BMI, alcohol, medication, comorbidity, viral serology and UV exposure; FIB-4 specificity above 65; missing data; scale-dependence of the sex finding; the meaninglessness of p values at this n)
20	Interpretation	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion, opening paragraphs and the closing “Clinical bottom line” paragraph — placed against the supplementation trials, the Cochrane review, the Mendelian randomisation and the two large general-population trials; no causal language is used anywhere

#	Item	Recommendation	Where reported (section pointer — layout-independent)
21	Generalisability	Discuss the generalisability (external validity) of the study results	Discussion, “Other limitations” — “a vitamin D-request cohort from a single tertiary centre in a single country, in which the indication for testing is unknown and is certainly not random; generalisation beyond similar laboratory populations is not warranted”
Other information			
22	Funding	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Declarations §Funding — no specific grant from any funding agency in the public, commercial or not-for-profit sectors

Additional declarations required by the journal but not by STROBE

Item	Where reported
Ethics approval, named committee, named guideline (Declaration of Helsinki), informed-consent waiver	Methods §Ethics (protocol KSEV-5-111125, approval date 11 November 2025; consent waived) and Declarations §Ethics approval and consent to participate
Data availability (restricted access, explicit conditions, aggregate data in SI, code deposit DOI)	Declarations §Availability of data and materials
Competing interests, stated for every author Author contributions (CRediT-style, by initials) AI / LLM use	Declarations §Competing interests Declarations §Authors’ contributions Methods §Use of artificial intelligence; Declarations §Acknowledgements — both confirm that all figures are programmatic (matplotlib) output and that no generative-AI image tool was used
Overlap with companion work from the same database	Methods §Overlap with companion work from the same database; cover letter
STROBE statement cited in Methods	Methods §Study design, final sentence (cites this checklist as Supplementary Table S13)

Item	Where reported
Supplementary Information contents	Declarations §Availability of data and materials (Additional file 1); contents list on the first page of this file

Supplementary Table S14. Negative-control outcome analysis

What this table calibrates. The fully adjusted 25(OH)D–FIB-4 association is small but not zero, and a residual is interpretable only against a calibration: how large a residual does *this* adjustment set, applied to *this* testing stream, produce for analytes whose concentration vitamin D does not set? A **negative-control outcome** is such an analyte — one for which the true exposure effect is taken to be zero, so that whatever the models return measures what the design and the adjustment procedure generate on their own [39,40]. **This analysis was added on 1 August 2026, after the primary analysis was complete, and is not prespecified.** What was fixed before any control was fitted is the control set, the four classification criteria and the falsification test.

Criteria for admission to the clean band. (i) measured in the same specimen stream; (ii) subject to the same co-testing selection and confounding structure — a *requirement* rather than a defect, because a negative control is meant to be confounded, and that is precisely what makes it a calibration of this design; (iii) no credible causal path from 25(OH)D to the analyte; (iv) not an outcome, nor an arithmetic component of an outcome, in this paper or in the companion analyses of the same database.

Estimator. Identical to the study outcomes: 25(OH)D per +25 nmol/L; cascade M0 crude → M1 + age, sex → M2 + serum albumin → M3 + season, era; all four models fitted on the same complete-case sample so the cascade is internally comparable; patient-clustered (CR0) robust standard errors as the primary inference. **M3 is exactly this paper’s model 4.** The paper’s model 5 adds eGFR, which the controls do not carry; its residual ($|\text{OR} - 1| = 0.065$) lies closer to the null than model 4’s (0.069), so the model-4 comparison reported throughout is the conservative one, and no percentile is computed for model 5 because that would mix two adjustment sets.

Falsification test, fixed in advance. The calibration argument fails for this paper if its residual is larger than *every* clean negative control. It is not — in either analysis frame, on either metric — and the test is reported here whichever way it falls.

Panel A — the control set, its tiers and its cut-points

Availability is the percentage of the 31,607-record primary cohort in which the analyte was co-ordered.

Analyte	Tier	Binary control	Cut-point source	Avail. (%)	Reservations recorded at the outset
Sodium	CLEAN	Hyponatraemia, Na < 135 mmol/L	Lower reference limit 135–145 mmol/L; the threshold of the 2014 European hyponatraemia guideline	91.8	Set by water balance through vasopressin, thirst and the renin–angiotensin–aldosterone system; no vitamin D pathway determines it
Chloride	CLEAN	Hypochloraemia, Cl < 98 mmol/L	Lower reference limit 98–107 mmol/L; the threshold used in the heart-failure hypochloraemia literature	64.4	Analyser-dependent cut-point; chloride moves with sodium and acid–base status

Analyte	Tier	Binary control	Cut-point source	Avail. (%)	Reservations recorded at the outset
Potassium	CLEAN	Hyperkalaemia, K > 5.0 mmol/L	Upper reference limit 3.5–5.0 mmol/L	96.8	Caveat stated because this is the largest survivor: 1,25(OH) ₂ D suppresses renin in animal models, so a distal aldosterone → potassium path is not logically impossible. It concerns the active hormone, not 25(OH)D, and no human evidence places plasma potassium under vitamin D control. Retained as CLEAN with the caveat on the record
Total CO ₂	CLEAN	Low bicarbonate, < 22 mmol/L	KDIGO alkali-therapy threshold; the standard definition of metabolic acidosis	15.1	Availability caveat: co-ordered far less often than the other controls, so this control rests on a small and differently selected subsample. Never quotable without that figure

Analyte	Tier	Binary control	Cut-point source	Avail. (%)	Reservations recorded at the outset
Mean platelet volume	CLEAN	MPV > 11 fL	Upper reference limit 7.5–11.5 fL, rounded	95.5	Weakest-sourced cut-point in the set: no consensus clinical threshold exists, and MPV is analyser-dependent and drifts with time from venepuncture in EDTA Criterion (iii) fails: adiposity and dysglycaemia are documented shared causes of low 25(OH)D and the vitamin D receptor is expressed in β -cells. Fasting status is unknown in these records Criterion (iii) fails: immunomodulation is among the best-documented non-skeletal actions of vitamin D
Glucose	CONTAMINATED	≥ 7.0 mmol/L	ADA/WHO fasting threshold	93.2	
White-cell count	CONTAMINATED	$\geq 10 \times 10^9/L$	Upper reference limit $4\text{--}10 \times 10^9/L$	95.7	

Analyte	Tier	Binary control	Cut-point source	Avail. (%)	Reservations recorded at the outset
Urea	CONTAMINATED	7.5 mmol/L	Upper reference limit 2.5–7.5 mmol/L	98.5	Tracks glomerular filtration and protein catabolism, i.e. the same nutritional and illness axis that albumin proxies Renally excreted; tracks eGFR, adiposity, alcohol and diuretics. A companion analysis reports urate as its own outcome A companion paper's own outcome, and macrocytosis is itself a marker of liver disease Circular for this paper: the platelet count is the denominator of FIB-4. No mechanism problem — a pure criterion (iv) failure
Urate	CONTAMINATED	420 µmol/L (men), > 360 (women)	Conventional sex-specific definition	44.5	
Mean cell volume	CONTAMINATED	80 fL	Lower reference limit 80–100 fL	95.7	
Platelet count	CONTAMINATED	150 × 10 ⁹ /L	Lower reference limit 150–400 × 10 ⁹ /L	95.6	

Excluded outright, not down-weighted. *Calcium* and *phosphate*: vitamin D directly regulates intestinal absorption of both (1,25(OH)₂D → TRPV6/calbindin; NaPi-2b), and both are primary outcomes of a companion analysis of this database — criteria (iii) and (iv) both fail, so neither is a negative control under any reading. *Corrected calcium*: an arithmetic function of calcium and albumin, carrying no information beyond them. *Total protein*: serum albumin is an arithmetic component of it, so in an albumin-adjusted model the outcome would be regressed partly on itself. That is verified rather than asserted — on the parallel haematological cohort (n = 31,868) total protein minus albumin is positive in 100.00% of records with a median of 26.0 g/L, the globulin range, and $r(\text{total protein, albumin}) = 0.7008$ ($R^2 = 0.4911$); in

the present cohort the albumin step removes 104.2% of total protein's crude association and drives it to the null (standardised β +0.10381 crude $\rightarrow$ -0.00721 at model 4, $p = 0.29$). *Serum albumin* itself is the adjustment variable. Every analyte that is an outcome of this paper or of a companion analysis (AST, ALT, GGT, ALP, bilirubins; haemoglobin, red-cell indices, iron studies; creatinine, eGFR, calcium, phosphate, urate) is excluded under criterion (iv).

Panel B — clean negative-control band at model 4, primary (record-level) cohort

Rows ordered by $|\text{OR} - 1|$. All estimates use patient-clustered robust confidence intervals, applied identically to the controls and to this paper's own outcome; the interval for FIB-4 > 2.67 is therefore wider than the model-based interval in main-text Table 3, Panel B, and n differs by two records because every row is fitted on the common complete-case sample.

Negative control	n	Events	Prev. (%)	OR per +25 nmol/L [95% CI]	p	$ \text{OR} - 1 $	$p < 0.05$
Low bicarbonate (total CO ₂ < 22 mmol/L)	4,745	569	12.0	1.039 [0.963, 1.122]	0.318	0.039	no
High mean platelet volume (MPV > 11 fL)	25,553	2,225	8.7	0.953 [0.917, 0.991]	0.015	0.047	yes
Hyponatraemia (Na < 135 mmol/L)	25,641	1,845	7.2	1.050 [1.009, 1.092]	0.016	0.050	yes
Hypochloreaemia (Cl < 98 mmol/L)	18,038	1,261	7.0	1.114 [1.069, 1.162]	3.6×10^{-7}	0.114	yes
Hyperkalaemia (K > 5.0 mmol/L)	26,168	1,940	7.4	1.202 [1.154, 1.251]	4.6×10^{-19}	0.202	yes
FIB-4 > 2.67 — this paper's outcome	25,586	4,123	16.1	0.931 [0.896, 0.968]	3.4×10^{-4}	0.069	yes

Band: 4 of 5 controls survive at $p < 0.05$; median $|\text{OR} - 1| = 0.050$; interquartile range 0.047–0.114; full range 0.039–0.202; sorted values 0.039, 0.047, 0.050, 0.114, 0.202. **This paper's residual** (0.069) lies inside that range at the **60th percentile** — larger than three controls, smaller than two, and 1.38 times the control median.

The band is small, and that is a limitation. With five controls a percentile can take only the values 0, 20, 40, 60, 80 and 100, and the median rests on a single control. The sorted values and the full range

are therefore printed above rather than the percentile alone, and Panel C gives a nine-control sensitivity band.

Metric caveat. $|\text{OR} - 1|$ is not prevalence-free: the same underlying effect produces a larger $|\text{OR} - 1|$ in a rarer outcome. The clean controls are rarer (7.0–12.0% of records) than high-risk FIB-4 (16.1%), which flatters this paper’s position on that metric. The comparison is therefore repeated on $|\log \text{OR}|$, where the residual is 0.071 against a control median of 0.048 and a range of 0.039–0.184 — the same 60th percentile. Neither metric removes the prevalence dependence entirely and we do not claim that it does.

Panel C — extended nine-control band (both tails of each clean analyte), record level

A sensitivity band only. The two tails of one analyte are **not independent tests**, so this band is wider but is not nine independent replicates.

Negative control	n	Events	Prev. (%)	OR per +25 nmol/L [95% CI]	<i>p</i>	$ \text{OR} - 1 $	<i>p</i> < 0.05
High bi-carbonate (total CO ₂ > 29 mmol/L)	4,745	809	17.0	1.016 [0.941, 1.097]	0.683	0.016	no
Low bicarbonate (total CO ₂ < 22 mmol/L)	4,745	569	12.0	1.039 [0.963, 1.122]	0.318	0.039	no
Hypernatraemia (Na > 145 mmol/L)	25,641	592	2.3	0.954 [0.880, 1.034]	0.254	0.046	no
High mean platelet volume (MPV > 11 fL)	25,553	2,225	8.7	0.953 [0.917, 0.991]	0.015	0.047	yes
Hyponatraemia (Na < 135 mmol/L)	25,641	1,845	7.2	1.050 [1.009, 1.092]	0.016	0.050	yes
Hyperchloraemia (Cl > 107 mmol/L)	18,038	2,564	14.2	0.912 [0.877, 0.948]	4.3×10^{-6}	0.088	yes
Hypochloraemia (Cl < 98 mmol/L)	18,038	1,261	7.0	1.114 [1.069, 1.162]	3.6×10^{-7}	0.114	yes
Hypokalaemia (K < 3.5 mmol/L)	26,168	587	2.2	0.884 [0.809, 0.967]	0.007	0.116	yes

Negative control	n	Events	Prev. (%)	OR per +25 nmol/L [95% CI]	<i>p</i>	OR – 1	<i>p</i> < 0.05
Hyperkalaemia (K > 5.0 mmol/L)	26,168	1,940	7.4	1.202 [1.154, 1.251]	4.6×10^{-19}	0.202	yes

Band: 6 of 9 survive; median 0.050; range 0.016–0.202. This paper’s residual sits at the **56th percentile**, with four controls above it.

Panel D — contaminated controls, reported separately and never inside the band

These analytes fail criterion (iii) or criterion (iv) and are shown so that the contamination is visible rather than asserted. **They are not part of any band and no percentile is computed against them.**

Contaminated control	n	Events	Prev. (%)	OR per +25 nmol/L [95% CI]	<i>p</i>	OR – 1	<i>p</i> < 0.05
Hyperuricaemia (urate > 420 µmol/L men, > 360 µmol/L women)	12,921	3,586	27.8	0.981 [0.943, 1.021]	0.344	0.019	no
Leucocytosis (WBC > $10 \times 10^9/L$)	25,602	3,291	12.9	0.921 [0.888, 0.955]	9.0×10^{-6}	0.079	yes
Raised urea (> 7.5 mmol/L)	26,553	8,009	30.2	1.090 [1.060, 1.122]	2.5×10^{-9}	0.090	yes
Microcytosis (MCV < 80 fL)	25,608	867	3.4	0.866 [0.806, 0.931]	8.9×10^{-5}	0.134	yes
Hyperglycaemia (glucose ≥ 7.0 mmol/L)	25,162	4,087	16.2	0.864 [0.833, 0.897]	1.0×10^{-14}	0.136	yes

Four of these five also survive, with a median |OR – 1| of 0.090 — larger than the clean band, as expected of analytes that share a causal or an arithmetic path with the exposure or the outcome. Thrombocytopenia is omitted here because the platelet count is the denominator of FIB-4, so its inclusion would be circular for this paper in either direction.

Panel E — full adjustment cascade, M0 to M3, odds ratio per +25 nmol/L

Every row is fitted on its own common complete-case sample, so the four models within a row are directly comparable. Percentage attenuation is deliberately **not** tabulated: for controls whose crude estimate sits near the null (low bicarbonate, crude OR 1.009) a percentage of the crude coefficient is numerically unstable and conveys nothing.

Outcome	n	M0 crude	M1 + age, sex	M2 + albumin	M3 + season, era
<i>CLEAN, primary band</i>					
Hyponatraemia (Na < 135 mmol/L)	25,641	0.945	0.969	1.084	1.050
Hypochloraemia (Cl < 98 mmol/L)	18,038	1.048	1.067	1.134	1.114
Hyperkalaemia (K > 5.0 mmol/L)	26,168	1.099	1.134	1.163	1.202
Low bicarbonate (total CO ₂ < 22 mmol/L)	4,745	1.009	1.008	1.111	1.039
High mean platelet volume (MPV > 11 fL)	25,553	0.956	0.965	1.014	0.953
<i>CLEAN, opposite tail (sensitivity only)</i>					
Hypernatraemia (Na > 145 mmol/L)	25,641	0.778	0.827	0.943	0.954
Hyperchloraemia (Cl > 107 mmol/L)	18,038	0.814	0.825	0.900	0.912
Hypokalaemia (K < 3.5 mmol/L)	26,168	0.715	0.737	0.908	0.884
High bicarbonate (total CO ₂ > 29 mmol/L)	4,745	0.956	0.956	0.925	1.016
<i>CONTAMINATED (never in the band)</i>					
Hyperuricaemia	12,921	0.942	0.949	0.973	0.981

Outcome	n	M0 crude	M1 + age, sex	M2 + albumin	M3 + season, era
Leucocytosis (WBC > 10 × 10 ⁹ /L)	25,602	0.853	0.867	0.935	0.921
Raised urea (> 7.5 mmol/L)	26,553	0.937	0.991	1.079	1.090
Microcytosis (MCV < 80 fL)	25,608	0.842	0.820	0.858	0.866
Hyperglycaemia (glucose ≥ 7.0 mmol/L)	25,162	0.796	0.835	0.891	0.864
<i>THIS PAPER</i>					
FIB-4 > 2.67	25,586	0.775	0.814	0.960	0.931

The adjustment set manufactures associations as well as removing them. Four of the five clean controls end further from the null than they began (hyperkalaemia 1.099 → 1.202; hypochloraemia 1.048 → 1.114; low bicarbonate 1.009 → 1.039; MPV 0.956 → 0.953), and sodium **changes direction** (hyponatraemia 0.945 → 1.050). In three of the five — sodium, chloride and bicarbonate — the albumin step (M1 → M2) is the largest single move on the log-odds scale, exactly as it is for this paper's own outcome (0.814 → 0.960, a step of +0.166 in log-odds against +0.049 for age and sex and −0.030 for season and era). In the remaining two, potassium and MPV, the season-and-era step is marginally the larger (+0.033 versus +0.025, and −0.062 versus +0.049). Whatever the albumin adjustment is doing in this cohort, it is not doing it only to FIB-4.

Panel F — continuous forms of the five clean controls at model 4

The binary controls dichotomise; the continuous forms do not, and they are reported so that the calibration does not rest on the cut-points in Panel A.

Analyte	n	β per +25 nmol/L [95% CI]	Standardised β	p	standardised β
Sodium (mmol/L)	25,641	−0.0575 [−0.0974, −0.0176]	−0.02078	0.005	0.02078
Total CO ₂ (mmol/L)	4,745	−0.0622 [−0.1654, 0.0410]	−0.02085	0.238	0.02085
Mean platelet volume (fL)	25,553	−0.0217 [−0.0332, −0.0101]	−0.02400	2.4 × 10 ^{−4}	0.02400
Chloride (mmol/L)	18,038	−0.2245 [−0.2817, −0.1673]	−0.06863	1.4 × 10 ^{−14}	0.06863
Potassium (mmol/L)	26,168	0.0293 [0.0220, 0.0365]	0.07697	2.9 × 10 ^{−15}	0.07697

Analyte	n	β per +25 nmol/L [95% CI]	Standardised β	p	standardised β
log FIB-4 — this paper's outcome	25,586	-0.0122 [-0.0188, -0.0055]	-0.01996	3.2×10^{-4}	0.01996

Band: 4 of 5 survive; median |standardised β | = 0.02400; range 0.02078–0.07697. This paper's log FIB-4 residual (0.01996) is **smaller than all five** clean controls — the 0th percentile. On the continuous scale, the scale on which the paper's own headline estimate is expressed, the residual lies below the entire noise band.

Panel G — one record per patient

The patient-level frame eliminates within-patient clustering entirely; every cluster is a singleton, so the clustered estimator reduces to the heteroscedasticity-robust one and the two frames remain comparable.

Negative control	n	Events	Prev. (%)	OR per +25 nmol/L [95% CI]	p	OR – 1	$p < 0.05$
Hyperkalaemia (K > 5.0 mmol/L)	17,782	1,138	6.4	1.037 [0.991, 1.085]	0.116	0.037	no
Low bicar- bonate (total CO ₂ < 22 mmol/L)	3,392	327	9.6	0.952 [0.860, 1.055]	0.348	0.048	no
High mean platelet volume (MPV > 11 fL)	17,377	1,520	8.7	0.949 [0.906, 0.994]	0.027	0.051	yes
Hyponatraemia (Na < 135 mmol/L)	17,391	1,427	8.2	1.068 [1.025, 1.113]	0.002	0.068	yes
Hypochloaemia (Cl < 98 mmol/L)	17,409	988	8.0	1.104 [1.055, 1.154]	1.7×10^{-5}	0.104	yes
FIB-4 > 2.67 — this paper's outcome	17,395	3,404	19.6	0.968 [0.932, 1.005]	0.092	0.032	no

Band: 3 of 5 survive; median 0.051; range 0.037–0.104. This paper's residual (0.032) is **smaller than every clean control** — the 0th percentile. Moving from the record-level frame to the patient-level frame

leaves the control band essentially unchanged (median ratio 1.03) while the paper’s own residual falls to 0.47 of its record-level value, so the margin improves from 1.38 times the control median to 0.63 times it. For this paper the patient-level frame is the more favourable of the two, and both are reported.

Panel H — positive control: the calibration has a ceiling

Total calcium is **excluded** from the control set precisely because vitamin D regulates it. It is fitted here for the opposite purpose: to establish that the procedure can still register an association that really exists, and then to show what the same adjustment does to it.

Model	standardised β for calcium	p	Clean-control band	Ratio to largest clean control	Inside the band?
M0 crude	0.16474	1.7×10^{-77}	0.00407– 0.07732	2.13	no
M3 (= this paper’s model 4)	0.03655	2.8×10^{-8}	0.02078– 0.07697	0.47	YES

The interpretive limit this imposes, stated plainly. Crudely, calcium’s association is 2.13 times the largest clean control, so the method has power. After the same adjustment it falls **inside** the clean band. It follows that a residual inside this band is **not distinguishable from** the residual the adjustment set produces on analytes vitamin D does not set — and that it is **not thereby shown to be zero**. The band bounds what adjustment can achieve here; it does not refute what remains inside the bound, and nothing in this paper claims that it does.

For contrast, phosphate — also directly vitamin-D-regulated, and also excluded — moves in the opposite direction: its standardised association *rises* across the cascade (+0.07134 crude $\rightarrow$ +0.10147 at model 4). That is a further demonstration that this adjustment set moves coefficients both ways and is not a one-directional attenuator.

Provenance. Every value in this table is read from `corrected/results.json  $\rightarrow$  negative_controls`, which `run_corrected_analysis.py` writes from `_shared/negative_controls.json` (sha256 a7e31d31ca4f6040b78073a7f619524ebf3e800bd84da1678d6dfc4c7c3230b2), produced by `_shared/negative_controls.py` reading the cohorts through the same shared data layer `_shared/vitd_common.py` that builds the cohort for this paper. No random numbers are used and no timestamp is written, so the output is byte-reproducible. The embedding step asserts that the control cohort is this paper’s cohort (31,607 records from 22,103 patients; index analytes AST, ALT, GGT, ALP) and that the control script’s M3 estimate for FIB-4 > 2.67 reproduces this paper’s published model 4, and it fails loudly if either check breaks. Software: Python 3.13.1, pandas 3.0.3, NumPy 2.4.6, statsmodels 0.14.6.