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.

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
Ljubljana University Medical Centre
Uroš Godnov
University of Primorska
Aleš Jerin
Ljubljana University Medical Centre
Darko Siuka
Ljubljana University Medical Centre
David Drobne
david.drobne@kclj.si

Ljubljana University Medical Centre

Research Article

Keywords: 25-hydroxyvitamin D, FIB-4, liver enzymes, gamma-glutamyltransferase, confounding, laboratory epidemiology

Posted Date: September 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11032434/v1

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

Additional Declarations: No competing interests reported.

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[1,2], Uroš Godnov[3], Aleš Jerin[1], Darko Siuka[4], David Drobne[4,5,*]

Affiliations

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

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

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

[4]Department of Gastroenterology, University Medical Centre Ljubljana, Zaloška cesta 2, 1000 Ljubljana, Slovenia

[5]Faculty of Medicine, University of Ljubljana, Vrazov trg 2, 1000 Ljubljana, Slovenia

Corresponding author

David Drobne, Department of Gastroenterology, University Medical Centre Ljubljana, Zaloška cesta 2, 1000 Ljubljana, Slovenia. E-mail: david.drobne@kclj.si

Email addresses

JO: josko.osredkar@kclj.si; UG: uros.godnov@upr.si; AJ: ales.jerin@kclj.si; DS: darko.siuka@kclj.si; DD: david.drobne@kclj.si

Abstract

Low serum 25-hydroxyvitamin D is repeatedly reported alongside abnormal liver chemistry and higher fibrosis scores, but its size, and how much survives adjustment for general health, are rarely quantified. We analysed 31,607 adult records from 22,103 patients with co-ordered 25-hydroxyvitamin D and a complete liver panel at a single Slovenian laboratory, 2014–2023. Correlations with every hepatobiliary analyte were negative but

small: gamma-glutamyltransferase was strongest (rho -0.156, 95% CI -0.167 to -0.145; 2.43% of rank variance), FIB-4 0.65% (rho -0.081, -0.092 to -0.069). Each 25 nmol/L higher 25-hydroxyvitamin D was crudely associated with 5.21% lower FIB-4 and 23% lower odds of FIB-4 above 2.67 (odds ratio 0.765, 0.743–0.787). Adjustment for age, sex, albumin, season, era and kidney function removed 80% of the continuous signal (-1.06%, -1.62 to -0.49) and 75% of the log-odds (0.935, 0.906–0.966). The one-record-per-patient adjusted association was null (0.967, 0.933–1.002). Five negative-control analytes vitamin D does not regulate, identically adjusted (model 4), showed residuals of similar size (median |OR-1| 0.050). The association is statistically unambiguous, clinically negligible, and would bring fewer than 0.5% of records within reach of the high-risk threshold. We present crude and adjusted estimates as bounds, not causal effects.

Keywords: 25-hydroxyvitamin D; FIB-4; liver enzymes; gamma-glutamyltransferase; confounding; laboratory epidemiology

---

Introduction

Vitamin D deficiency and chronic liver disease travel together. Low serum 25-hydroxyvitamin D [25(OH)D] is found in the great majority of patients with established cirrhosis [13,14], and cross-sectional surveys repeatedly report inverse associations between 25(OH)D and hepatic steatosis, aminotransferase activity and non-invasive fibrosis scores [10–12]. The mechanistic case is attractive: the vitamin D receptor is expressed in hepatic stellate cells, and receptor ligands gate the transforming-growth-factor-β/SMAD programme that drives fibrogenesis in mice [15]. From these two observations a therapeutic hypothesis has been built and, repeatedly, tested.

The fibrosis scores in question are now part of routine practice rather than a research instrument. FIB-4, a four-variable index of age, aspartate aminotransferase, alanine aminotransferase and platelet count, is recommended as the initial risk-stratification step in steatotic liver disease both by the AASLD practice guidance [3] and by the joint EASL–EASD–EASO clinical practice guidelines [4], and the disease itself was renamed by multisociety Delphi consensus from non-alcoholic fatty liver disease to metabolic dysfunction-associated steatotic liver disease during the period of this analysis [5]. We retain the older terminology where it is the terminology of the studies we cite. Anything that shifts FIB-4 therefore has a route into clinical decisions, which is why the size of such a shift, and not only its statistical significance, is the quantity of interest.

It has not survived testing. Randomised supplementation in type 2 diabetes with fatty liver produced no change in hepatic fat or enzymes [16]; a Cochrane review of 27 trials in chronic liver disease found the evidence insufficient and inconsistent [17]; bi-directional Mendelian randomisation found no causal relation between genetically instrumented 25(OH)D and non-alcoholic fatty liver disease [18]; and the two largest general-population trials of vitamin D found no effect on the non-skeletal endpoints they were designed to test [19,20]. The most parsimonious reading of the discordance is the one Autier and

colleagues proposed from a wider literature: that low 25(OH)D is largely a *consequence* of ill health rather than a cause of it [21,22].

That reading has a testable corollary in laboratory data. If low 25(OH)D marks general poor health rather than hepatic biology specifically, then (i) the crude association with any liver marker should be small, (ii) it should be broadly similar across analytes rather than concentrated on a fibrosis-specific one, and (iii) it should attenuate substantially once age and a proxy for inflammatory and nutritional status are entered into the model. 25(OH)D behaves as a negative acute-phase reactant — it falls by roughly 40% within 48 h of elective arthroplasty, independently of any change in vitamin D stores [23,24] — and it circulates almost entirely bound to vitamin D-binding protein and albumin [25], both of which are hepatically synthesised and both of which fall in inflammation [26].

Large routine-laboratory datasets are well suited to this question and badly suited to the usual way of reporting it. With tens of thousands of observations, essentially every association reaches conventional statistical significance, so a p value carries almost no information; what matters is the effect size, its confidence interval, and how much of it survives adjustment. Such datasets also carry a structural hazard that is rarely acknowledged: patients enter the analysis only because a clinician ordered both tests, so the sample is conditioned on a collider of illness and vitamin D status [33].

We therefore asked three deliberately modest questions of a ten-year, single-centre laboratory database. First, how large — in variance-explained terms — is the association between 25(OH)D and each of the standard hepatobiliary analytes and the FIB-4 index? Second, how much of that association remains after adjustment for age, sex, serum albumin, season, calendar era and estimated glomerular filtration rate? Third, is what remains large enough to matter to a patient? We state at the outset that the answer to the third question is no, and that we regard the honest quantification of a widely over-claimed association as the contribution of this work.

Results

Cohort

Of 107,900 laboratory records from 2014–2023, 107,856 carried a valid serum 25(OH)D result and 82,235 belonged to adults aged 18–100 years. Complete and physiologically plausible co-measurement of aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT) and alkaline phosphatase (ALP) was available in 37,456 records; taking the first record per patient per calendar year yielded the primary cohort of **31,607 records from 22,103 adults** (Fig. 1). Most patients (78.3%) contributed a single record and the maximum was ten.

Figure 1. STROBE flow diagram, hepatobiliary cohort (2014-2023)

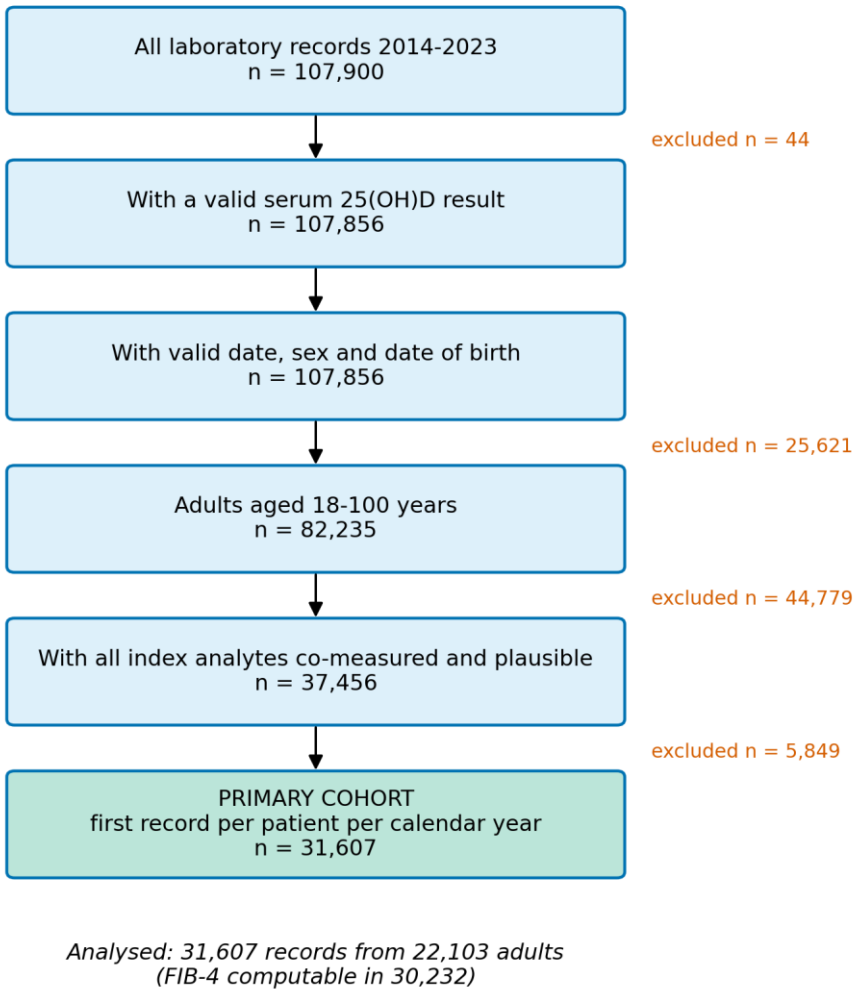


Figure 1. STROBE participant flow for the hepatobiliary cohort, 2014–2023. Sequential application of the prespecified cohort rule to the source laboratory extract. From 107,900 records, 107,856 carried a valid serum 25-hydroxyvitamin D result, 107,856 also had a valid sampling date, sex and date of birth, and 82,235 belonged to adults aged 18–100 years. Requiring aspartate aminotransferase, alanine aminotransferase, gamma-glutamyltransferase and alkaline phosphatase to be co-measured and physiologically plausible left 37,456 records, and retaining the first record per patient per calendar year gave the primary analysis cohort of 31,607 records from 22,103 adults. Boxes show n at each step and the number excluded. The FIB-4 index was computable in 30,232 of these records (95.6%); the 1,375 records without a platelet count are treated as missing for every FIB-4 analysis and are never counted as low risk.

The cohort was 58.3% female with a median age of 60.8 years (interquartile range, IQR 45.0–73.2) and a median 25(OH)D of 59 nmol/L (IQR 38–79); 16.1% were deficient (<30

nmol/L). Median GGT was 22.2 U/L (0.37 µkat/L) and median ALP 72.0 U/L (1.20 µkat/L). FIB-4 was computable in 30,232 records (95.6%; platelet counts were missing in the remainder), with a median of 1.22 (IQR 0.76–1.99); 4,641 of those 30,232 (15.4%) exceeded the 2.67 high-risk threshold. Table 1 gives the full baseline description by 25(OH)D category.

Two features of Table 1 deserve emphasis because they set up everything that follows. First, the four 25(OH)D categories are not exchangeable groups: the deficient group was six years older than the insufficient group (median 65.2 vs 58.9 years), had a 3.0 g/L lower median albumin (41.0 vs 44.0 g/L), and was sampled disproportionately in winter (40.3% vs 26.1% of the optimal group). Second, the gradient in FIB-4 across 25(OH)D categories is **not monotone** (median 1.56, 1.17, 1.13, 1.22 from deficient to optimal; Kruskal–Wallis H = 602.4, df = 3, $p = 3.1 \times 10^{-130}$, $\varepsilon^2 = 0.020$). The "optimal" group is older than the two middle groups and contains supplemented patients, so its slightly higher FIB-4 is expected and is a first warning against reading the exposure gradient as a dose–response curve.

Every hepatobiliary association is negative and small

All ten prespecified Spearman correlations with 25(OH)D were negative and all were statistically unambiguous, but none explained more than 2.5% of rank variance (Table 2; Fig. 5a). GGT was the strongest (rho = -0.156, 95% CI -0.167 to -0.145; n = 31,601; rho^2 = 2.43%; $p = 6.4 \times 10^{-171}$), followed by ALP (rho = -0.131, -0.142 to -0.120; 1.71%), AST (rho = -0.090, -0.101 to -0.079; 0.82%), FIB-4 (rho = -0.081, -0.092 to -0.069; n = 30,226; 0.65%; $p = 7.3 \times 10^{-45}$) and ALT (rho = -0.071, -0.082 to -0.060; 0.50%). The De Ritis ratio was essentially unrelated (rho = -0.023; 0.05%). The largest correlate of 25(OH)D in the entire panel was not a liver enzyme at all but serum albumin (rho = +0.189, +0.178 to +0.201; 3.59%) — a stronger association than any hepatobiliary analyte.

Figure 5. Effect sizes for 25(OH)D and the hepatobiliary panel

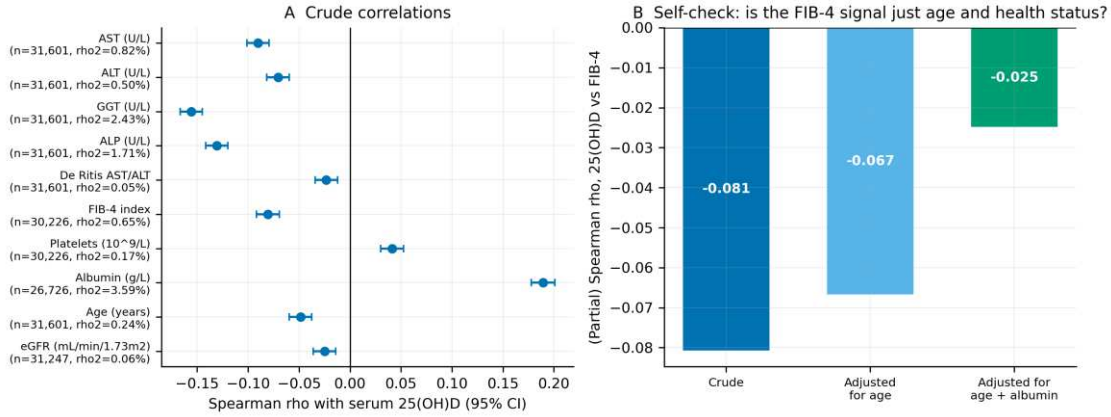


Figure 5. Effect sizes for 25-hydroxyvitamin D and the hepatobiliary panel, with the FIB-4 self-check. (a) Forest plot of crude Spearman rank correlations between 25-hydroxyvitamin D and each analyte, with Fisher z-transformed 95% confidence intervals and n annotated. All hepatobiliary correlations are negative and small (maximum |rho| = 0.156 for gamma-

glutamyltransferase, 2.43% of rank variance); the largest correlation in the whole panel is with serum albumin (+0.189, 3.59%). (b) Self-check for a derived-variable artefact: the 25-hydroxyvitamin D–FIB-4 partial Spearman correlation falls from -0.081 (crude) to -0.067 after residualising on age and to -0.025 after residualising on age and albumin, and the same cascade computed on the age-free surrogate AST/(platelets × √ALT) reproduces it (crude rho -0.065). FIB-4 contains age and platelets by construction (rho with age +0.687, with platelets -0.618), whereas 25-hydroxyvitamin D correlates only weakly with either (-0.049 and +0.041), so the association is not arithmetically forced — but neither is it larger than the crude derived variable implies.

Comparing the extremes of exposure gave the same ordering of magnitudes (Fig. 2): Cliff's delta between deficient and optimal patients was 0.265 for GGT, 0.228 for ALP, 0.193 for FIB-4, 0.173 for AST and 0.099 for ALT, but -0.338 for albumin (Supplementary Table S5). The strongest contrast between vitamin D-deficient and vitamin D-replete patients in this cohort concerns a nutrition and inflammation marker, not a hepatic one.

Figure 2. Hepatobiliary outcomes across 25(OH)D status categories (n = 31,607 records)

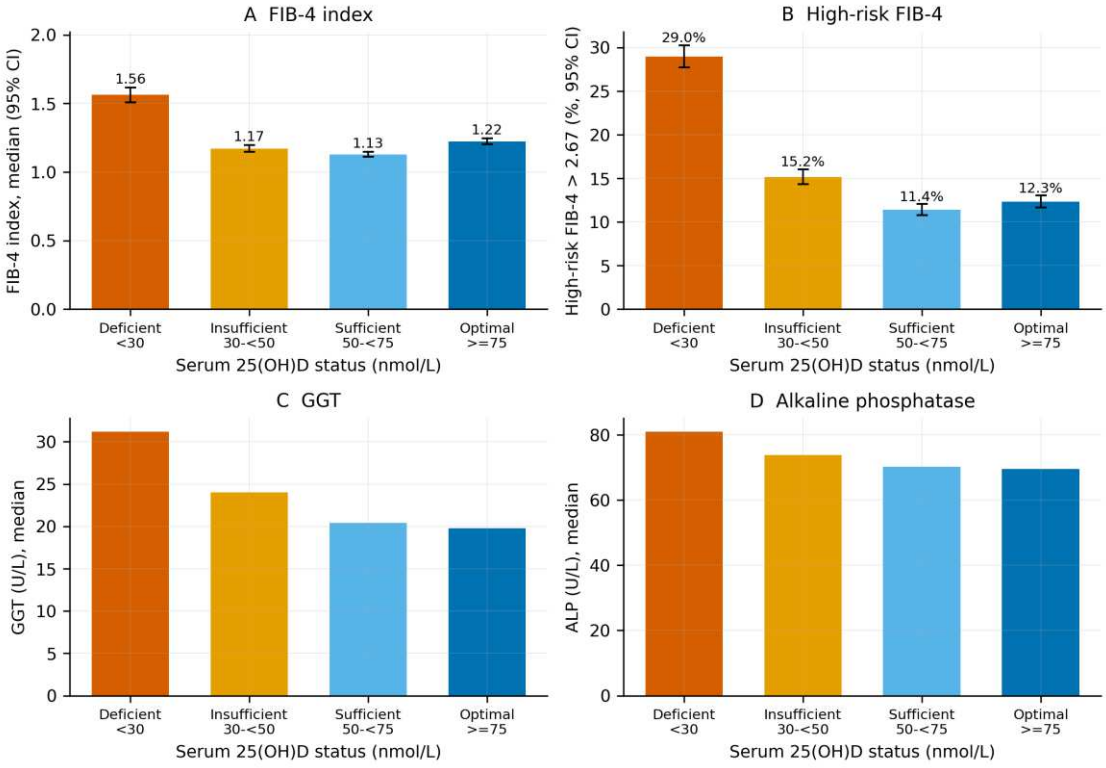


Figure 2. Hepatobiliary outcomes across 25-hydroxyvitamin D status categories. Four panels for the primary cohort, plotted against the Institute of Medicine / Endocrine Society categories deficient (<30 nmol/L, n = 5,080), insufficient (30 to <50, n = 7,055), sufficient (50 to <75, n = 10,045) and optimal (≥75, n = 9,421). (a) Median FIB-4 index with bootstrap 95% confidence interval; note the non-monotone pattern (1.56, 1.17, 1.13, 1.22), which tracks the median age of each group (65.2, 58.9, 59.0, 61.3 years) rather than the exposure. (b)

Percentage with high-risk FIB-4 > 2.67 with Wilson 95% confidence intervals, computed only among records in which FIB-4 was computable (29.0%, 15.2%, 11.4%, 12.3%). (c) Median gamma-glutamyltransferase and (d) median alkaline phosphatase, both in U/L (divide by 60 for µkat/L). Kruskal–Wallis tests across the four categories with epsilon-squared effect sizes are given in Supplementary Table S4, and the corresponding Dunn post-hoc comparisons in Supplementary Table S3; all epsilon-squared values are below 0.03, i.e. less than 3% of the variance in any outcome is shared with the exposure category.

Adjustment removes most of the FIB-4 association

Table 3 and Fig. 3 report the prespecified five-step adjustment cascade. Crudely, each +25 nmol/L of 25(OH)D was associated with a **5.21% lower FIB-4** (95% CI -5.90% to -4.52%; n = 30,226; $p = 9.7 \times 10^{-47}$). Adding age and sex reduced this to -2.78% (-3.32% to -2.23%; 47.3% attenuation). Adding serum albumin reduced it to -0.39% (-0.94% to +0.17%; $p = 0.168$; 92.7% attenuation) — that is, to a value indistinguishable from zero. Adding season and era moved the coefficient partly back, to -1.21% (-1.77% to -0.65%; 77.2% attenuation), and the full model additionally including estimated glomerular filtration rate (eGFR) gave **-1.06% (-1.62% to -0.49%; $p = 2.4 \times 10^{-4}$; n = 25,504; 80.2% attenuation)**.

Figure 3. Crude versus progressively adjusted 25(OH)D effect sizes

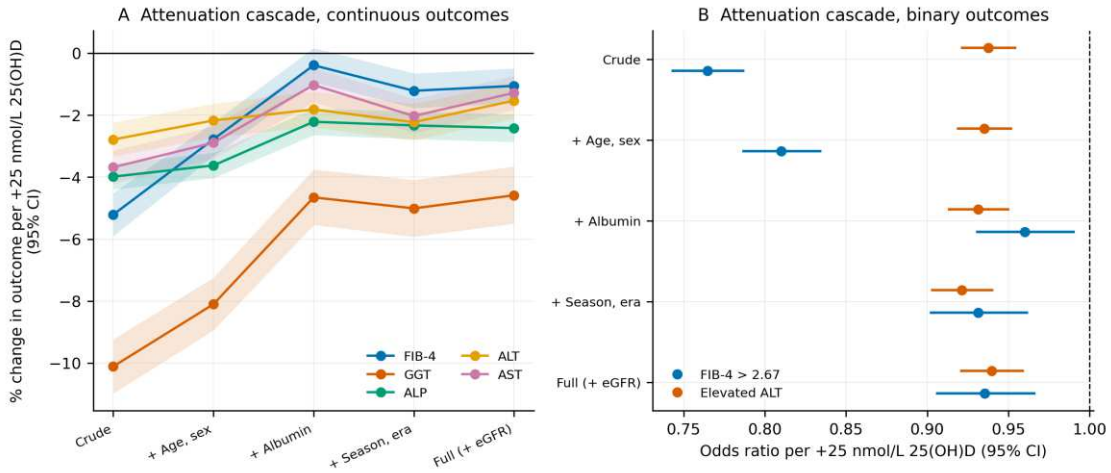


Figure 3. Crude versus progressively adjusted 25-hydroxyvitamin D effect sizes. The prespecified five-step adjustment cascade: crude; + age and sex; + serum albumin; + season and era; + estimated glomerular filtration rate. (a) Continuous outcomes, plotted as the percentage change in the log-transformed outcome per +25 nmol/L of 25-hydroxyvitamin D with 95% confidence intervals from HC3 robust standard errors. For FIB-4 the coefficient moves from -5.21% (crude) to -1.06% (full model), an 80.2% attenuation, and the intermediate albumin-only step over-shoots to -0.39%. (b) Binary outcomes as odds ratios per +25 nmol/L with 95% Wald confidence intervals; for FIB-4 > 2.67 the odds ratio moves from 0.765 to 0.935 (75.1% attenuation of the log-odds ratio), while elevated alanine aminotransferase is essentially unattenuated (0.938 to 0.940). The cascade is deliberately shown step by step because it is not monotone; reporting only the endpoints

would conceal that the adjusted estimate depends on where the model stops. n differs between steps because of complete-case analysis and is annotated on each row.

We report all five steps rather than the endpoints because **the cascade is not monotone**: albumin alone over-shoots (92.7%) and the seasonal and era terms partially restore the coefficient. A reader given only "crude -5.2%, adjusted -1.1%, attenuation 80%" would not know that the adjusted estimate depends on which step the model stops at, and we regard that as material.

The attenuation is not an artefact of the changing sample size. Restricting the crude model to the 25,588 rows in which albumin was measured gave a crude coefficient of -5.12% — essentially identical to the full-cohort crude value — against a fully adjusted -1.06% on almost the same rows, i.e. **79.8% attenuation on identical data**.

For the binary outcome the pattern was the same. Crude odds of FIB-4 > 2.67 fell by 23.5% per +25 nmol/L (OR 0.765, 95% CI 0.743–0.787; 4,641 events in 30,226 records). Adjustment for age and sex gave OR 0.810; adding albumin gave OR 0.960 (0.931–0.990); adding season and era gave OR 0.931; the full model gave **OR 0.935 (0.906–0.966), $p = 4.2 \times 10^{-5}$, n = 25,504**, corresponding to 75.1% attenuation of the log-odds ratio. For comparison, in that same model the odds ratio per year of age was 1.061, per g/L of albumin 0.874, and for the 2020–2023 era 1.87 — each far larger than the exposure of interest.

Elevated ALT behaved differently. Using the ACG "true-normal" upper limits of 25 U/L in women and 33 U/L in men [8,9], 13,736 of 31,607 records (43.5%) were flagged. The crude odds ratio was 0.938 (0.922–0.954) and the fully adjusted 0.940 (0.921–0.959) — essentially unattenuated overall, but again not monotonically so, since the season-and-era step *strengthened* the association by 27% before eGFR adjustment reversed it. Because a 43.5% prevalence is a consequence of the deliberately low "true-normal" thresholds rather than of the population, we repeated the analysis at the routine clinical-chemistry limits of 0.58/0.74 µkat/L (34.8/44.4 U/L), which flagged 22.5% of records: crude OR 0.896 (0.877–0.916), fully adjusted 0.940 (0.917–0.964). The conclusion is unchanged by threshold; the full threshold series, including a single 40 U/L cut-off in both sexes, is in Supplementary Table S12, Panel A.

For GGT, the analyte with the strongest crude association, the same cascade gave -10.11% per +25 nmol/L crudely and -4.59% fully adjusted (56.0% attenuation); for ALP, -3.98% to -2.42% (39.8%). Full model output for all six continuous outcomes is in Supplementary Table S6.

The FIB-4 signal is largely an age and health-status signal

FIB-4 contains age in its numerator and platelets in its denominator, so an apparent 25(OH)D–FIB-4 association could be arithmetically forced. We tested this explicitly (Fig. 5b). FIB-4 correlated with age at rho = +0.687 and with platelets at rho = -0.618, whereas 25(OH)D correlated with age at only rho = -0.049 and with platelets at rho = +0.041; the 25(OH)D–FIB-4 association is therefore not simply age leaking through. Partial Spearman

correlation, residualising both ranks, fell from -0.081 crude to -0.067 adjusting for age and to **-0.025 adjusting for age and albumin** ($p = 7.2 \times 10^{-5}$, n = 25,588) — 0.06% of rank variance. Repeating the entire analysis with an age-free surrogate, AST/(platelets × √ALT), gave the same picture (rho = -0.065; crude -3.94% per +25 nmol/L, fully adjusted -1.13%, 71.7% attenuation), confirming that the residual association is not a derived-variable artefact but is also not larger than the derived variable suggested.

Sex and age modify the association; era does not

Formal likelihood-ratio interaction tests, not subgroup comparisons, were used throughout (Supplementary Table S7). The 25(OH)D × sex interaction on log FIB-4 was supported (LR $\chi^2 = 19.42$, df = 1, $p = 1.0 \times 10^{-5}$; n = 25,588), as was 25(OH)D × age ($\chi^2 = 11.97$, $p = 5.4 \times 10^{-4}$); the interaction with era was not ($\chi^2 = 0.25$, $p = 0.616$). Sex-stratified correlations were rho = -0.031 in women (n = 17,576) versus -0.132 in men (n = 12,650), a difference confirmed by Fisher z (z = 8.77, $p < 1 \times 10^{-16}$). In the fully adjusted models the association was confined to men (OR 0.904, 0.862–0.947) and to patients over 60 (OR 0.920, 0.889–0.953), and was absent in women (OR 0.966, 0.925–1.009) and in those aged ≤60 (OR 0.973, 0.903–1.049) (Supplementary Table S8). On the binary outcome, however, the sex interaction was *not* supported ($\chi^2 = 1.72$, $p = 0.190$) while the age interaction was ($\chi^2 = 13.25$, $p = 2.7 \times 10^{-4}$), so the sex-specific claim should be regarded as scale-dependent and provisional.

Season, era and secular trend

25(OH)D showed the expected seasonal cycle (Fig. 4; median 52 nmol/L in winter and 53 in spring versus 65 in both summer and autumn; deficiency 21.3% versus 8.3%; Kruskal–Wallis H = 1,258.8, $p = 1.3 \times 10^{-272}$, $\varepsilon^2 = 0.040$; Fig. 4a), consistent in direction with the pronounced seasonal variation previously described for the Slovenian general adult and elderly population in the Nutrihealth survey [29]. FIB-4 varied far less across seasons ($\varepsilon^2 = 0.005$). Median 25(OH)D rose across the decade (rho with calendar year = +0.107, 0.096–0.118) while FIB-4 rose only marginally (rho = +0.041). The 2020–2023 era differed materially from 2014–2019 in both exposure (Cliff's delta -0.126) and outcome (delta -0.113), and carried a large independent effect on high-risk FIB-4 (OR 1.87), which is why era was retained as a covariate rather than asserted to be irrelevant.

Figure 4. Seasonal and secular patterns in 25(OH)D and FIB-4

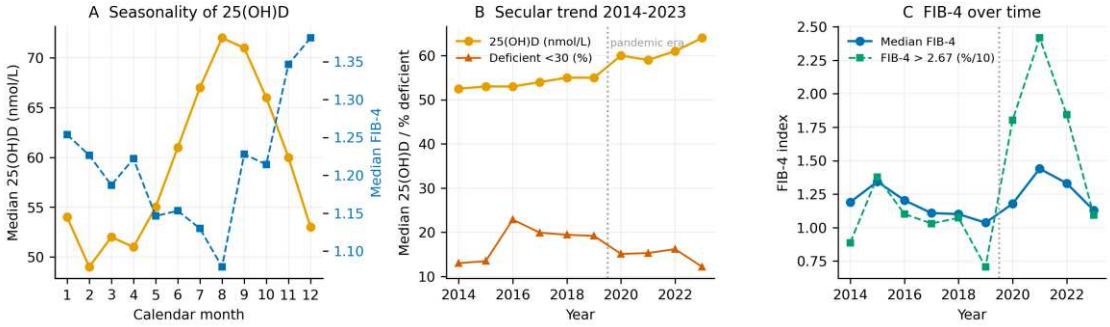


Figure 4. Seasonal and secular patterns in 25-hydroxyvitamin D and FIB-4, 2014–2023. (a) Median 25-hydroxyvitamin D by calendar month (left axis) with median FIB-4 overlaid (right axis), showing the expected late-winter trough and late-summer peak in the exposure and a far flatter outcome curve. (b) Median 25-hydroxyvitamin D and the percentage of records below 30 nmol/L by calendar year; the exposure rose across the decade (Spearman rho with year = +0.107, 95% CI 0.096 to 0.118). (c) FIB-4 by calendar year, showing the marked 2020–2021 excursion that motivated retaining calendar era (pre-pandemic 2014–2019 versus 2020–2023) as a model covariate rather than asserting it to be unimportant; in the fully adjusted model the pandemic era carried an odds ratio of 1.87 for high-risk FIB-4, far larger than the exposure of interest. Numeric values underlying all three panels are in Supplementary Tables S9, S9b and S10.

Sensitivity and missing-data analyses

Restricting to one record per patient (n = 22,103, the earliest record) gave a *larger* crude correlation with FIB-4 (rho = -0.126, -0.139 to -0.113; n = 21,058), but the association attenuated to the null on adjustment: the fully adjusted odds ratio for FIB-4 > 2.67 was **0.9670 (95% CI 0.9334–1.0018), *p* = 0.063**, and the fully adjusted continuous estimate -0.40% per +25 nmol/L (*p* = 0.261) (Supplementary Table S11). At the patient level, therefore, the adjusted association is not distinguishable from the null, and we do not present this analysis as confirmation of the primary one.

We record explicitly why this number changed, because it changed in our own favour. An earlier version of this analysis returned a fully adjusted odds ratio of 0.963 (0.929–0.997), *p* = 0.033, and we had reported the patient-level result as concordant with the primary analysis. That significance rested on 208 records whose FIB-4 was computed from a platelet count that had been carried over from a different visit: the sensitivity cohort was assembled with a de-duplication rule that took the first non-missing value in each column rather than the earliest record as a whole, so patients whose earliest visit had no platelet count were nonetheless assigned one. Removing those 208 records — they have no measured FIB-4 and should never have had one — reduces the evaluable sample from 21,266 to 21,058 and removes the significance with them. The defect and its correction are documented in Supplementary Methods S.3.

Two readings of the resulting discordance are compatible with the data and we cannot separate them here. The patient-level analysis discards about 30% of the records and is correspondingly less precise, so an effect of this size would not reliably reach conventional significance in it; equally, the primary adjusted estimate may be partly sustained by repeated sampling of the same patients. The point estimate is the same negligible magnitude in both, which is the basis on which we draw our conclusion — but the statistical significance of the primary estimate should be read with this discordance in mind.

Missingness of co-ordered analytes was not random with respect to the exposure, which is the empirical signature of the co-testing selection problem. Patients in whom glucose was *not* co-ordered had a median 25(OH)D of 67 versus 58 nmol/L (Cliff's delta -0.149, $p = 1.0 \times 10^{-30}$), and those without a urate result had a median of 57 versus 61 nmol/L (delta +0.083, $p = 7.1 \times 10^{-37}$). Albumin missingness was much less differential (delta -0.023, $p = 0.011$) but not null (Supplementary Table S1).

Finally, we quantified what the adjusted estimate would mean for a patient. A +25 nmol/L difference in 25(OH)D corresponds, in the fully adjusted model, to a FIB-4 shift of -0.013 units at the cohort median of 1.22. Only 138 of 30,232 records (0.46%) lie close enough to the 2.67 threshold for a shift of that size to cross it. The E-value for the adjusted odds ratio is 1.31 (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 would suffice to produce the entire residual association. The reclassification arithmetic, the distribution of records per patient and the E-value calculation are set out in Supplementary Table S12, Panels B–D.

How large a residual does this adjustment set produce where no effect can exist?

The adjusted association is small but not zero, and the obvious objection to a deflationary reading is that something survived adjustment. We therefore calibrated it against negative-control outcomes — analytes co-ordered in the same specimen and not under vitamin D control (Methods; Table 3, Panel C).

Under the identical adjustment set, which is this paper's model 4, four of the five clean negative controls "survived" at $p < 0.05$, with a median |OR – 1| of 0.050 and a range of 0.039 to 0.202. Two were substantially larger than this paper's own residual: hypochloraemia (OR 1.114, 95% CI 1.069–1.162) and hyperkalaemia (1.202, 1.154–1.251). The adjustment set does not only remove associations, it also produces them: four of the five controls moved further from the null on adjustment (hyperkalaemia 1.099 → 1.202), and sodium changed direction (0.945 → 1.050).

This paper's residual at that same step is OR 0.931 (0.896–0.968), |OR – 1| = 0.069; the interval is wider than Table 3's because every estimate in this calibration carries patient-clustered robust standard errors. It is **inside** the control range and sits at its **60th percentile** — three controls smaller, two larger, 1.38 times the control median. Because |OR – 1| is not prevalence-free (the same effect gives a larger value in a rarer outcome) and

the clean controls are rarer (7.0–12.0% of records) than high-risk FIB-4 (16.1%), we also report the comparison on the |log OR| scale, where the position is unchanged (0.071 against a control median of 0.048, range 0.039–0.184; 60th percentile). A nine-control band adding the opposite tail of each analyte — two tails of one analyte are not independent — puts it at the 56th percentile, range 0.016–0.202.

On the continuous outcome and in the patient-level analysis the residual sits lower still. For log FIB-4 the standardised coefficient at model 4 is -0.0200, **below all five** clean controls (band 0.0208 to 0.0770, median 0.0240). Restricting to one record per patient, the residual falls to |OR – 1| = 0.032 while the control band barely moves (median 0.051, range 0.037–0.104), so it too is **smaller than every clean control**. The comparison has a stated failure mode — the argument fails for this paper if its residual exceeds *every* clean control — and it was not triggered in either frame or on either metric.

The calibration has a ceiling, which we measured rather than assumed. Total calcium — which vitamin D genuinely regulates, and which is excluded from the control set for that reason — has a crude standardised association 2.13 times the largest clean control, so the procedure has power; but after the same adjustment calcium too falls **inside** the clean band (|standardised β| 0.037 against 0.021 to 0.077). A residual inside this band is therefore not distinguishable from what the adjustment set returns on analytes vitamin D does not set — it is not thereby shown to be absent. The full control set, all four cascade steps, the continuous forms, the contaminated controls kept outside the band and the analytes excluded outright are in Supplementary Table S14.

Discussion

In 31,607 records from 22,103 adults, serum 25(OH)D was inversely associated with every hepatobiliary analyte we examined and with the FIB-4 fibrosis index. Every one of those associations was statistically overwhelming and every one was small: the largest, with GGT, explained 2.4% of rank variance, and the FIB-4 association explained 0.65%. Adjustment for age, sex, serum albumin, season, era and eGFR removed 80% of the continuous FIB-4 association and 75% of the log-odds of high-risk FIB-4, leaving a residual of about -1% in FIB-4 per 25 nmol/L. That residual would shift a median patient's FIB-4 by 0.013 units and would reclassify essentially nobody. We therefore report an association that is real in the statistical sense and negligible in the clinical one.

Two internal patterns argue that what we measured is general health status rather than hepatic biology. First, the analyte most strongly associated with 25(OH)D in the whole panel was serum albumin (rho = +0.19), not any liver enzyme — and the largest deficient-versus-optimal contrast was likewise albumin (Cliff's delta -0.33). Second, adding albumin to the model was, by a wide margin, the single most consequential adjustment step, moving the FIB-4 coefficient from -2.78% to -0.39%. If 25(OH)D were tracking fibrogenesis, one would not expect a nutrition-and-inflammation marker to absorb almost the whole signal.

This is consistent with the wider literature, in which the observational and interventional evidence have diverged persistently. Supplementation trials in fatty liver disease [16], the Cochrane synthesis in chronic liver disease [17], Mendelian randomisation against NAFLD [18] and the two largest general-population vitamin D trials [19,20] all failed to reproduce the benefit predicted by cross-sectional associations of the kind we report here. Our findings do not adjudicate that question — a cross-sectional laboratory study cannot — but they do quantify how little there is to explain: an association of 0.65% of rank variance does not require a large biological effect to produce it, and the physiology of 25(OH)D as a negative acute-phase reactant [23,24] bound to hepatically synthesised carrier proteins [25,26] supplies an unforced explanation. The observation that GGT, an oxidative-stress and induction marker rather than a fibrosis marker [37], showed the strongest association fits the same reading.

The albumin adjustment is a bound, not a control. Serum albumin was the only available proxy for systemic inflammation and nutritional status, and it is not a clean confounder. Albumin is hepatically synthesised and therefore lies partly downstream of the hepatic process that FIB-4 is intended to capture (Spearman rho = -0.39 with FIB-4 in this cohort); it is simultaneously a negative acute-phase reactant that covaries with 25(OH)D (rho = +0.19), and a carrier compartment for circulating 25-hydroxyvitamin D, alongside the quantitatively dominant vitamin D-binding protein [25]. Adjustment for albumin is therefore expected to remove genuine confounding, some mediated pathway, and some of the exposure itself, in unknown proportions [32]. We consequently present the crude (-5.2% in FIB-4 per 25 nmol/L; OR 0.765 for FIB-4 > 2.67) and the fully adjusted (-1.1%; OR 0.935, 95% CI 0.906–0.966) estimates as **bounds on a partially confounded, partially over-adjusted association rather than as competing estimates of a causal effect**, and we make no claim as to where within that interval any true association lies. This is reinforced by the sampling frame: the cohort exists only where a clinician co-ordered 25(OH)D and a liver panel, and we show that the probability of co-ordering additional analytes varies systematically with 25(OH)D concentration (Cliff's delta = -0.149 for glucose co-testing, $p = 1.0 \times 10^{-30}$). Conditioning on having been tested is conditioning on a collider of ill health and vitamin D status [33], so neither the magnitude nor the direction of the residual association is formally identified. Even taken at face value, the fully adjusted association corresponds to 0.65% of the rank variance in FIB-4 and would bring fewer than 0.5% of records within reach of the high-risk fibrosis threshold; we therefore report it as a population-level marker of general health status and explicitly not as evidence that vitamin D status influences hepatic fibrosis risk. The E-value of 1.31 makes the same point quantitatively [34]: a modest unmeasured confounder suffices.

"But something survived adjustment." That is the natural objection to a deflationary reading, and we answered it by measuring rather than arguing. The identical adjustment set applied to five analytes whose concentration vitamin D does not set returns residuals of the same order as ours: four reach $p < 0.05$, their median $|OR - 1|$ is 0.050, and two exceed our own 0.069, which sits at the 60th percentile of that distribution; on the continuous outcome and at one record per patient our residual is smaller than every clean control. The inference we draw is deliberately narrow: **a residual that lies inside the range this**

adjustment set produces on analytes vitamin D does not control cannot be read as a specific vitamin D effect on hepatic fibrosis risk, because a residual of that size is what the procedure returns where no such effect can exist.

We are equally explicit about what the controls do not establish. They do not prove that our residual is confounded — a negative control cannot identify the mechanism by which an association arises, and in a single cross-section neither can anything else. Nor do they show the residual to be zero, and we quantified that limit rather than asserting it: total calcium, which vitamin D demonstrably does regulate, has a crude association 2.13 times larger than the largest clean control, yet after the same adjustment it too lies inside the clean band. The controls therefore **bound what this adjustment set can achieve; they do not adjudicate what remains inside that bound**. Two caveats belong with them: the band is built from five controls, so its median is fragile and the sorted values are reported alongside every percentile; and the controls share this laboratory's testing stream and its co-testing selection — a requirement of the design rather than a defect, since a negative control is meant to be confounded, but one that limits the calibration to this stream. Read together with the E-value of 1.31, the reclassification arithmetic and the patient-level null, the controls say that what we are reporting is of the size this design generates as background.

Other limitations. The design is cross-sectional, so no temporal ordering and no causal inference is available; all language in this paper is deliberately associational. The cohort is 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. We lack parathyroid hormone, body-mass index, alcohol intake, medication (including vitamin D supplementation, which is very likely to contribute to the non-monotone FIB-4 gradient in the "optimal" group), comorbidity coding, viral hepatitis serology and any measure of ultraviolet exposure other than sampling month; residual confounding by all of these is likely and is not bounded by our models. FIB-4 is a surrogate: it was derived in HIV/HCV co-infection [1] and validated in fatty liver disease [2], it predicts liver-related mortality at population scale [7], but its specificity above 65 years of age falls sharply [6], and 31.2% of our cohort was over 70, so some of the 15.4% classified as high-risk are false positives. Albumin was missing in 15.4% of records and platelets in 4.4%, and we used complete-case analysis; the missing-data analysis (Supplementary Table S1) shows the missingness is mildly differential in the exposure. The sex-specific finding was supported on the continuous scale but not on the binary one, and should be treated as hypothesis-generating. Finally, with n above 30,000 essentially every comparison attains significance, which is why we have reported an effect size with a confidence interval for every claim and have avoided any inference resting on a p value alone.

Clinical bottom line. There is no clinically actionable association here. Nothing in these data supports measuring 25(OH)D to refine hepatic fibrosis risk, adding it to a non-invasive fibrosis score, or supplementing vitamin D with a hepatic endpoint in mind. Conversely, a low 25(OH)D in a patient with abnormal liver chemistry should be read as what it most

likely is — a marker of poor general health, inflammation, low albumin, indoor living or advanced age — and investigated as such rather than as a hepatic finding. We consider the calibration of this association to be the useful result: it converts a frequently over-interpreted correlation into a bounded, negligible quantity.

Methods

Study design, setting and data source

This was a retrospective cross-sectional analysis of routine clinical laboratory data generated at the Institute of Clinical Chemistry and Biochemistry, University Medical Centre Ljubljana, Slovenia, between 2 January 2014 and 29 December 2023. The database comprises 107,900 records; 107,856 (99.96%) carry a serum 25(OH)D result, so the source population is a vitamin D-request cohort with co-ordered panels attached rather than a general-population sample. Reporting follows the STROBE statement for cross-sectional studies [31] (completed checklist: Supplementary Table S13).

Ethics

The study used pre-existing, pseudonymised routine laboratory data and involved no intervention and no contact with patients. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of University Medical Center Ljubljana (protocol code KSEV-5-111125; approval date 11 November 2025), which authorized retrospective use of anonymized laboratory data from 2014 to 2023 for this study. Patient consent was waived because the study is a retrospective analysis of fully de-identified routine laboratory data with no possibility of re-identification and no contact with participants, in accordance with. Patient names were irreversibly hashed to a pseudonymous identifier at the point of loading and the name field was discarded; no identifier, and no individual-level data, appear in this article, its figures, its tables or its Supplementary Information.

Variables and units

Analytes are reported in the units delivered by the laboratory (Slovenian convention): 25(OH)D in nmol/L; creatinine in µmol/L; albumin in g/L; platelets in 10^9/L; AST, ALT, GGT and ALP in µkat/L. Catalytic activities are additionally given in U/L throughout (µkat/L × 60) so that an international readership can interpret them. 25(OH)D was categorised using the Institute of Medicine and Endocrine Society cut-points [27,28] as deficient (<30 nmol/L), insufficient (30 to <50), sufficient (50 to <75) and optimal (≥75). Effect sizes for 25(OH)D are expressed per +25 nmol/L, a clinically interpretable increment corresponding to roughly one interquartile step in this cohort.

Derived variables: FIB-4 = age (years) × AST (U/L) / [platelets (10^9/L) × √ALT (U/L)] [1], with the conventional high-risk threshold of 2.67 adopted from its validation in fatty liver disease [2]; the De Ritis ratio as AST/ALT; eGFR by the race-free CKD-EPI 2021 creatinine equation

[30]; and elevated ALT by the sex-specific ACG “true-normal” upper limits of 25 U/L (women) and 33 U/L (men) [8,9]. These are the low “healthy-range” limits derived from screened blood donors [9] and not the reporting laboratory’s own reference interval; they flag 43.5% of this cohort, and we therefore report a prespecified sensitivity analysis at the routine clinical-chemistry limits of 0.58 µkat/L (women) and 0.74 µkat/L (men), i.e. 34.8 and 44.4 U/L, and at 40 U/L in both sexes. Season was assigned from month of sampling (winter = December–February) and era as pre-pandemic (2014–2019) versus pandemic (2020–2023). Values outside prespecified physiological plausibility ranges were set to missing rather than deleted row-wise.

Cohort definition

The cohort rule was fixed in advance and is shared across this laboratory’s paper series: (i) valid 25(OH)D; (ii) valid sampling date, sex and date of birth; (iii) age 18–100 years; (iv) all four index analytes (AST, ALT, GGT, ALP) co-measured and physiologically plausible; (v) primary analysis unit = the first record per patient per calendar year. A one-record-per-patient sensitivity analysis is reported.

Analysis unit and estimand. Retaining the first qualifying record per patient per calendar year, so that a patient may contribute up to ten records, is a deliberate estimand choice, and we state its consequences rather than treat it as a technicality. The estimand is a **record-level** association within a laboratory testing stream and not a patient-level association: 78.3% of patients contribute a single record and supply 54.7% of the analysed records, while the remaining 21.7% of patients supply 45.3% of them. Patient-clustered dependence is why a one-record-per-patient analysis is reported alongside the primary estimates; it does not alter the weighting of the primary analysis and we do not claim that it does. Because the rule retains the earliest qualifying record within each patient-year by construction, records sampled early in the calendar year are over-represented relative to the eligible pool, which is one of the reasons season is a covariate in every adjusted model. Where the primary and the one-record analyses disagree we report both and draw our conclusion from the magnitude of the point estimates rather than from the statistical significance of either. The one-record-per-patient cohort is formed by taking each patient’s earliest qualifying record **as a whole row** — not the first non-missing value in each column, which would splice values recorded at different visits into a profile that no visit produced — with complete-case restriction applied afterwards within each model.

Note that the plausibility filter is applied after the 25(OH)D-present filter, so **6 of the 31,607 cohort records have a 25(OH)D value that was subsequently set to missing as implausible** (outside 1–500 nmol/L); analyses of the exposure therefore use n = 31,601 while cohort-level counts use n = 31,607. This discrepancy is intentional and is documented in Supplementary Methods.

Statistical analysis

All analyses were performed in Python 3 with pandas, NumPy, SciPy and statsmodels. Alpha was set at 0.05 and all tests were two-tailed. Exact p values are reported throughout, and the word “significant” is used only when accompanied by a p value.

Continuous variables were markedly right-skewed (e.g. GGT median 22.2 U/L, IQR 13.8–45.0, with a maximum two orders of magnitude above the median) and Shapiro–Wilk and visual assessment rejected normality for every analyte; **non-parametric methods were therefore used for all univariable comparisons**, and all regression models with a continuous outcome were fitted on the natural-log scale, in which the residuals were approximately symmetric. Datasets are summarised as median (a labelled measure of centre) with interquartile range (a labelled measure of variability) and n.

Univariable association. Spearman rank correlation with Fisher z-transformed 95% confidence intervals, reported with n, rho, its confidence interval, rho^2 as the percentage of rank variance explained, and the exact p value. Fisher z-transformed confidence intervals were verified against 1,000-replicate bootstrap intervals during independent verification and agreed to within 0.002 in rho.

Group comparison. Kruskal–Wallis across the four 25(OH)D categories with epsilon-squared as the effect size, followed by Dunn post-hoc tests with **Bonferroni** correction across the six pairwise comparisons within each outcome — the conservative alternative to false-discovery-rate control [36] — with both raw and corrected p values reported (Supplementary Table S3). Two-group contrasts (deficient versus optimal; pandemic versus pre-pandemic era) used the Mann–Whitney U test with Cliff’s delta as the effect size [35].

Multivariable models. A prespecified five-step adjustment cascade was applied to every outcome: (1) crude; (2) + age and sex; (3) + serum albumin; (4) + season and era; (5) + eGFR. Continuous outcomes were modelled by ordinary least squares on the log scale with HC3 heteroscedasticity-consistent standard errors, and coefficients are reported as the percentage change in the outcome per +25 nmol/L, with standardised beta and partial R^2. Binary outcomes (FIB-4 > 2.67; elevated ALT) were modelled by logistic regression, reporting odds ratios per +25 nmol/L with 95% Wald confidence intervals and McFadden pseudo-R^2. Attenuation is reported as the percentage reduction in the exposure coefficient (or log-odds ratio) relative to the crude model **at every step, not only at the endpoint**, because the cascade is not monotone. To exclude changing sample size as the explanation for attenuation, the crude model was refitted on the complete-case rows of the fully adjusted model. Complete-case analysis was used; no imputation was performed, and a missing-data analysis comparing the exposure distribution by missingness of each co-ordered analyte is reported (Supplementary Table S1).

Effect modification. Subgroup estimates are reported only alongside a formal likelihood-ratio test of the exposure × modifier interaction term in the fully adjusted model; no claim is made from a comparison of two separately significant subgroups.

Derived-variable artefact check. Because FIB-4 contains age and platelets by construction, we (a) report correlations of 25(OH)D with each FIB-4 component separately, (b) compute partial Spearman correlations residualising both ranks on age and on age plus albumin, and (c) repeat the whole cascade using the age-free surrogate AST/(platelets × √ALT).

Sensitivity analysis for unmeasured confounding. The E-value was computed for the fully adjusted odds ratio on the approximate risk-ratio scale, converting the odds ratio using the marginal outcome prevalence [34].

Negative-control outcomes. A residual that survives adjustment 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 any association the models return is a measure of what the design and the adjustment procedure generate on their own [39,40]. **This analysis was added after the primary analysis was complete and is not prespecified**; what was fixed before any control was fitted is the control set, the classification criteria and the falsification test stated below.

Four criteria were required: (i) measured in the same specimen stream; (ii) subject to the same co-testing selection and confounding structure — a *requirement* rather than a defect, since a negative control is meant to be confounded, and that is what makes it a calibration; (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 (Methods, "Overlap"). Five analytes met all four and form the **clean** control set, with cut-points taken from conventional adult reference limits or standard clinical thresholds: sodium (hyponatraemia, Na < 135 mmol/L), chloride (hypochloraemia, Cl < 98 mmol/L), potassium (hyperkalaemia, K > 5.0 mmol/L), total CO_2 (low bicarbonate, < 22 mmol/L) and mean platelet volume (MPV > 11 fL). Each cut-point, and the reasoning and the reservations recorded for each control, are given in Supplementary Table S14; potassium and total CO_2 carry explicit caveats there, and total CO_2 was available in only 15.1% of the cohort, so its control estimate rests on a differently selected subsample.

Calcium and phosphate were excluded outright rather than down-weighted: vitamin D directly regulates intestinal absorption of both, so neither can serve as a negative control at any level of adjustment, and both are primary outcomes of a companion analysis of this database. Total protein was excluded because serum albumin — the pivotal covariate here — is an arithmetic component of it, so an albumin-adjusted model would regress the outcome partly on itself. Serum albumin itself, and every analyte that is an outcome of this paper or its companions, were excluded under criterion (iv). Six further analytes are reported separately as **contaminated** controls and are never admitted to the control band: glucose and white-cell count because a vitamin D pathway is plausible (adiposity and dysglycaemia are documented shared causes of low 25(OH)D; immunomodulation is among the best-documented non-skeletal actions of vitamin D), urea and urate because they track the same nutritional and renal axis that albumin proxies, mean cell volume

because it is a companion paper's own outcome, and the platelet count because it is the denominator of FIB-4 and therefore circular for this paper.

Every control was fitted with the same estimator and the same cascade as the study outcomes — 25(OH)D per +25 nmol/L, crude → + age and sex → + serum albumin → + season and era — with all four models on the same complete-case sample and with patient-clustered robust standard errors. That set is exactly this paper's model 4, so the comparison with the controls is made at model 4, where the adjustment sets match; model 5 adds eGFR, which the controls do not carry, and its residual lies closer to the null, so the model-4 comparison is the conservative one. Binary and continuous forms of every control are reported. Total calcium was additionally fitted as a **positive** control, to establish that the procedure retains power to register an association vitamin D really does produce. The falsification test was fixed before any control was fitted: the calibration argument fails for this paper if its residual is larger than *every* clean control, and that test is reported whichever way it falls.

Overlap with companion work from the same database

This laboratory has published, and is publishing, other analyses of the same source database, and we declare the overlap explicitly. (i) A companion paper on vitamin D status in this laboratory population appeared in *Nutrients* [38] in 2026. (ii) A manuscript on the relationship between 25(OH)D and indirect versus direct bilirubin, by Osredkar, Godnov, Jerin and Žiberna — the authors of the present paper — is being finalised for submission to *Antioxidants*. That manuscript is unpublished and has no DOI, accession number or preprint identifier at the time of this submission; it therefore carries no reference number here, in accordance with this journal's requirement that the reference list contain only work that has been published or accepted by a named publication, recognised preprint server or data repository. A clearly marked copy is uploaded with this submission and is available to the editors and reviewers. **Bilirubin is not an exposure, outcome or covariate in the present analysis**, and no bilirubin result appears anywhere in this paper. (iii) Two further analyses of the same database, addressing haematological indices and renal-mineral indices respectively, are in preparation by the same authors; they use different index analytes, different cohort selections (48,402 and 33,852 records respectively) and different research questions. The present paper is the only one of the series in which hepatobiliary enzymes and FIB-4 are the outcomes.

Because all four analyses draw on the same testing stream, the patient overlap is substantial and we state it numerically rather than qualitatively: 96.7% of the patients analysed here also appear in the haematological cohort and 61.4% in the renal-mineral cohort, and 13,043 patients appear in all three of the parallel analyses. No hepatobiliary or FIB-4 result reported here appears in any companion work. Two descriptive quantities do recur across the series — the Spearman correlations of 25(OH)D with serum albumin and with eGFR — and their values differ between cohorts (rho with eGFR -0.025 here, against +0.072 and +0.076 in the haematological and renal-mineral cohorts respectively). We regard that difference as informative rather than contradictory: each cohort is defined by

which analytes a clinician co-ordered, so the co-ordering pattern itself determines the sign of a marginal correlation. This is the selection mechanism we describe as a collider in the Discussion, made visible.

Use of artificial intelligence

Large language models were used as a drafting and code-review aid during preparation of the analysis scripts and the manuscript text, under the direct supervision and full responsibility of the named authors; all statistical results were produced by the deposited, re-runnable analysis code and were independently re-derived by a second code path before submission. No large language model is an author or satisfies authorship criteria. **All five figures were generated with Python/matplotlib directly from the study data; no generative-AI image tool was used at any point.**

References

1. Sterling, R. K. et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* **43**, 1317–1325 (2006).
2. Shah, A. G. et al. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin. Gastroenterol. Hepatol.* **7**, 1104–1112 (2009).
3. Rinella, M. E. et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology* **77**, 1797–1835 (2023).
4. European Association for the Study of the Liver, European Association for the Study of Diabetes & European Association for the Study of Obesity. EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *J. Hepatol.* **81**, 492–542 (2024).
5. Rinella, M. E. et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Hepatology* **78**, 1966–1986 (2023).
6. McPherson, S. et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. *Am. J. Gastroenterol.* **112**, 740–751 (2017).
7. Unalp-Arida, A. & Ruhl, C. E. Liver fibrosis scores predict liver disease mortality in the United States population. *Hepatology* **66**, 84–95 (2017).
8. Kwo, P. Y., Cohen, S. M. & Lim, J. K. ACG Clinical Guideline: evaluation of abnormal liver chemistries. *Am. J. Gastroenterol.* **112**, 18–35 (2017).
9. Prati, D. et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. *Ann. Intern. Med.* **137**, 1–10 (2002).
10. Eliades, M. et al. Meta-analysis: vitamin D and non-alcoholic fatty liver disease. *Aliment. Pharmacol. Ther.* **38**, 246–254 (2013).
11. Ji, Y., Wei, C.-B., Gu, W. & Hou, L.-L. Relevance of vitamin D on NAFLD and liver fibrosis detected by vibration controlled transient elastography in US adults: a cross-sectional analysis of NHANES 2017–2018. *Ann. Med.* **55**, 2209335 (2023).

12. Ciardullo, S., Muraca, E., Cannistraci, R., Perra, S. & Perseghin, G. Low 25 (OH) vitamin D levels are associated with increased prevalence of nonalcoholic fatty liver disease and significant liver fibrosis. *Diabetes Metab. Res. Rev.* **39**, e3628 (2023).
13. Arteh, J., Narra, S. & Nair, S. Prevalence of vitamin D deficiency in chronic liver disease. *Dig. Dis. Sci.* **55**, 2624–2628 (2010).
14. Kitson, M. T. & Roberts, S. K. D-livering the message: the importance of vitamin D status in chronic liver disease. *J. Hepatol.* **57**, 897–909 (2012).
15. Ding, N. et al. A vitamin D receptor/SMAD genomic circuit gates hepatic fibrotic response. *Cell* **153**, 601–613 (2013).
16. Barchetta, I. et al. No effects of oral vitamin D supplementation on non-alcoholic fatty liver disease in patients with type 2 diabetes: a randomized, double-blind, placebo-controlled trial. *BMC Med.* **14**, 92 (2016).
17. Bjelakovic, M., Nikolova, D., Bjelakovic, G. & Gluud, C. Vitamin D supplementation for chronic liver diseases in adults. *Cochrane Database Syst. Rev.* **8**, CD011564 (2021).
18. Wang, N. et al. Vitamin D and nonalcoholic fatty liver disease: bi-directional Mendelian randomization analysis. *EBioMedicine* **28**, 187–193 (2018).
19. Manson, J. E. et al. Vitamin D supplements and prevention of cancer and cardiovascular disease. *N. Engl. J. Med.* **380**, 33–44 (2019).
20. Neale, R. E. et al. The D-Health Trial: a randomised controlled trial of the effect of vitamin D on mortality. *Lancet Diabetes Endocrinol.* **10**, 120–128 (2022).
21. Autier, P., Boniol, M., Pizot, C. & Mullie, P. Vitamin D status and ill health: a systematic review. *Lancet Diabetes Endocrinol.* **2**, 76–89 (2014).
22. Autier, P., Mullie, P., Macacu, A. et al. Effect of vitamin D supplementation on non-skeletal disorders: a systematic review of meta-analyses and randomised trials. *Lancet Diabetes Endocrinol.* **5**, 986–1004 (2017).
23. Waldron, J. L. et al. Vitamin D: a negative acute phase reactant. *J. Clin. Pathol.* **66**, 620–622 (2013).
24. Reid, D. et al. The relation between acute changes in the systemic inflammatory response and plasma 25-hydroxyvitamin D concentrations after elective knee arthroplasty. *Am. J. Clin. Nutr.* **93**, 1006–1011 (2011).
25. Chun, R. F., Peercy, B. E., Orwoll, E. S., Nielson, C. M., Adams, J. S. & Hewison, M. Vitamin D and DBP: the free hormone hypothesis revisited. *J. Steroid Biochem. Mol. Biol.* **144**, 132–137 (2014).
26. Gabay, C. & Kushner, I. Acute-phase proteins and other systemic responses to inflammation. *N. Engl. J. Med.* **340**, 448–454 (1999).
27. Ross, A. C. et al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J. Clin. Endocrinol. Metab.* **96**, 53–58 (2011).
28. Holick, M. F. et al. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J. Clin. Endocrinol. Metab.* **96**, 1911–1930 (2011).

29. Hribar, M. et al. Nutrihealth Study: seasonal variation in vitamin D status among the Slovenian adult and elderly population. *Nutrients* **12**, 1838 (2020).
30. Inker, L. A. et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N. Engl. J. Med.* **385**, 1737–1749 (2021).
31. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).
32. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488–495 (2009).
33. Griffith, G. J. et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat. Commun.* **11**, 5749 (2020).
34. VanderWeele, T. J. & Ding, P. Sensitivity analysis in observational research: introducing the E-value. *Ann. Intern. Med.* **167**, 268–274 (2017).
35. Cliff, N. Dominance statistics: ordinal analyses to answer ordinal questions. *Psychol. Bull.* **114**, 494–509 (1993).
36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
37. Lee, D. H., Blomhoff, R. & Jacobs, D. R. Is serum gamma glutamyltransferase a marker of oxidative stress? *Free Radic. Res.* **38**, 535–539 (2004).
38. Osredkar J, Siuka D, Jerin A, Štabuc B, Godnov U. Ten-Year Trends in Serum 25-Hydroxyvitamin D in Slovenia (2014-2023): Laboratory-Based Data from Tested Individuals and COVID-19-Period Changes. Nutrients. 2026;18(7):1168. https://doi.org/10.3390/nu18071168.
39. Lipsitch, M., Tchetgen Tchetgen, E. & Cohen, T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* **21**, 383–388 (2010).
40. Shi, X., Miao, W. & Tchetgen Tchetgen, E. A selective review of negative control methods in epidemiology. *Curr. Epidemiol. Rep.* **7**, 190–202 (2020).

Declarations

Ethics approval and consent to participate

This study was approved by the National Medical Ethics Committee of the Republic of Slovenia (approval number KSEV-5-111125). The committee waived the requirement for individual informed consent because the analysis used pseudonymised, routinely collected clinical laboratory data with no identifiers returned to the research team. All research was conducted in accordance with the Declaration of Helsinki and applicable Slovenian and European Union data-protection legislation.

Consent for publication

Not applicable. The manuscript contains no individual person's data in any form (including images, videos, or personal or clinical details).

Availability of data and materials

The aggregate datasets supporting the conclusions of this article — all baseline, correlation, model, interaction, subgroup, seasonal, secular and sensitivity tables — are provided in full in the Supplementary Information (Additional file 1). Individual-level clinical laboratory data cannot be made publicly available, because they are pseudonymised routine healthcare records held by the Institute of Clinical Chemistry and Biochemistry, University Medical Centre Ljubljana, and are governed by Slovenian and European Union data-protection law. De-identified individual-level data are available from the corresponding author (D.D.) on reasonable request, subject to (i) a written research protocol, (ii) written permission from the Institute of Clinical Chemistry and Biochemistry, University Medical Centre Ljubljana, (iii) a data-transfer agreement, and (iv) any approval required by the National Medical Ethics Committee of the Republic of Slovenia. Requests will be answered within 30 days. The complete, re-runnable analysis code (analysis.py, the shared data layer vitd_common.py, the correction script run_corrected_analysis.py, the negative-control script negative_controls.py and the independent verification script verify_stats.py) is archived on Zenodo with DOI [TO CONFIRM prior to acceptance], from which every number in this article can be regenerated from the source extract.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The article-processing charge will be covered by the Springer Nature transformative agreement with the Slovenian Consortium of Academic and Research Institutions through the corresponding author's affiliation with the University of Ljubljana.

Authors' contributions

J.O. conceived the study and provided institutional resources. U.G. and A.J. designed the methodology. U.G. wrote the analysis software. U.G., A.J. and D.D. validated the analysis. U.G. performed the formal statistical analysis. J.O., A.J. and D.D. conducted the investigation. J.O. and U.G. curated the data. J.O. and U.G. wrote the original draft. J.O., D.S., A.J. and D.D. reviewed and edited the manuscript. U.G. produced all figures. D.D. supervised the project. J.O. administered the project. All authors have read and approved the final manuscript.

Acknowledgements

The authors thank the staff of the Institute of Clinical Chemistry and Biochemistry, University Medical Centre Ljubljana, for maintaining the laboratory information system from which these data were extracted. No medical writer was employed. No co-author drew any figure by hand; all figures are programmatic output. Generative artificial-intelligence tools were used only for language polishing and code review; all scientific content, statistical analysis and interpretation were performed by the authors.

Figure legends

Figure 1. STROBE participant flow for the hepatobiliary cohort, 2014–2023. Sequential application of the prespecified cohort rule to the source laboratory extract. From 107,900 records, 107,856 carried a valid serum 25-hydroxyvitamin D result, 107,856 also had a valid sampling date, sex and date of birth, and 82,235 belonged to adults aged 18–100 years. Requiring aspartate aminotransferase, alanine aminotransferase, gamma-glutamyltransferase and alkaline phosphatase to be co-measured and physiologically plausible left 37,456 records, and retaining the first record per patient per calendar year gave the primary analysis cohort of 31,607 records from 22,103 adults. Boxes show n at each step and the number excluded. The FIB-4 index was computable in 30,232 of these records (95.6%); the 1,375 records without a platelet count are treated as missing for every FIB-4 analysis and are never counted as low risk.

Figure 2. Hepatobiliary outcomes across 25-hydroxyvitamin D status categories. Four panels for the primary cohort, plotted against the Institute of Medicine / Endocrine Society categories deficient (<30 nmol/L, n = 5,080), insufficient (30 to <50, n = 7,055), sufficient (50 to <75, n = 10,045) and optimal (≥75, n = 9,421). (**a**) Median FIB-4 index with bootstrap 95% confidence interval; note the non-monotone pattern (1.56, 1.17, 1.13, 1.22), which tracks the median age of each group (65.2, 58.9, 59.0, 61.3 years) rather than the exposure. (**b**) Percentage with high-risk FIB-4 > 2.67 with Wilson 95% confidence intervals, computed only among records in which FIB-4 was computable (29.0%, 15.2%, 11.4%, 12.3%). (**c**) Median gamma-glutamyltransferase and (**d**) median alkaline phosphatase, both in U/L (divide by 60 for µkat/L). Kruskal–Wallis tests across the four categories with epsilon-squared effect sizes are given in Supplementary Table S4, and the corresponding Dunn post-hoc comparisons in Supplementary Table S3; all epsilon-squared values are below 0.03, i.e. less than 3% of the variance in any outcome is shared with the exposure category.

Figure 3. Crude versus progressively adjusted 25-hydroxyvitamin D effect sizes. The prespecified five-step adjustment cascade: crude; + age and sex; + serum albumin; + season and era; + estimated glomerular filtration rate. (**a**) Continuous outcomes, plotted as the percentage change in the log-transformed outcome per +25 nmol/L of 25-hydroxyvitamin D with 95% confidence intervals from HC3 robust standard errors. For FIB-4 the coefficient moves from -5.21% (crude) to -1.06% (full model), an 80.2% attenuation, and the intermediate albumin-only step over-shoots to -0.39%. (**b**) Binary outcomes as odds ratios per +25 nmol/L with 95% Wald confidence intervals; for FIB-4 > 2.67 the odds ratio moves from 0.765 to 0.935 (75.1% attenuation of the log-odds ratio), while elevated alanine aminotransferase is essentially unattenuated (0.938 to 0.940). The cascade is deliberately shown step by step because it is not monotone; reporting only the endpoints would conceal that the adjusted estimate depends on where the model stops. n differs between steps because of complete-case analysis and is annotated on each row.

Figure 4. Seasonal and secular patterns in 25-hydroxyvitamin D and FIB-4, 2014–2023. (**a**) Median 25-hydroxyvitamin D by calendar month (left axis) with median FIB-4 overlaid (right axis), showing the expected late-winter trough and late-summer peak in the exposure

and a far flatter outcome curve. (**b**) Median 25-hydroxyvitamin D and the percentage of records below 30 nmol/L by calendar year; the exposure rose across the decade (Spearman rho with year = +0.107, 95% CI 0.096 to 0.118). (**c**) FIB-4 by calendar year, showing the marked 2020–2021 excursion that motivated retaining calendar era (pre-pandemic 2014–2019 versus 2020–2023) as a model covariate rather than asserting it to be unimportant; in the fully adjusted model the pandemic era carried an odds ratio of 1.87 for high-risk FIB-4, far larger than the exposure of interest. Numeric values underlying all three panels are in Supplementary Table S9 (seasonal), Table S9b (monthly) and Table S10 (annual).

Figure 5. Effect sizes for 25-hydroxyvitamin D and the hepatobiliary panel, with the FIB-4 self-check. (**a**) Forest plot of crude Spearman rank correlations between 25-hydroxyvitamin D and each analyte, with Fisher z-transformed 95% confidence intervals and n annotated. All hepatobiliary correlations are negative and small (maximum |rho| = 0.156 for gamma-glutamyltransferase, 2.43% of rank variance); the largest correlation in the whole panel is with serum albumin (+0.189, 3.59%). (**b**) Self-check for a derived-variable artefact: the 25-hydroxyvitamin D–FIB-4 partial Spearman correlation falls from -0.081 (crude) to -0.067 after residualising on age and to -0.025 after residualising on age and albumin, and the same cascade computed on the age-free surrogate AST/(platelets × √ALT) reproduces it (crude rho -0.065). FIB-4 contains age and platelets by construction (rho with age +0.687, with platelets -0.618), whereas 25-hydroxyvitamin D correlates only weakly with either (-0.049 and +0.041), so the association is not arithmetically forced — but neither is it larger than the crude derived variable implies.

Tables

Table 1. Baseline characteristics of the hepatobiliary cohort by serum 25-hydroxyvitamin D category Values are median [interquartile range] for continuous variables and n (%) for categorical variables. Group comparisons are Kruskal–Wallis tests across the four categories with epsilon-squared effect sizes for continuous variables, and chi-squared tests with Cramér's V for categorical variables; exact p values are given. Enzyme activities are shown in U/L with the delivered μkat/L values beneath (U/L = μkat/L × 60). The FIB-4 > 2.67 row uses the 30,232 records in which FIB-4 was computable as denominator; the eGFR < 60 row uses the 31,253 records with a creatinine result.

Characteristic	Overall (n = 31,607)	Deficient <30 (n = 5,080)	Insufficient 30–<50 (n = 7,055)	Sufficient 50–<75 (n = 10,045)	Optimal ≥75 (n = 9,421)	*p*	Effect size
Age (years)	60.8 [45.0–73.2]	65.2 [52.1–77.8]	58.9 [43.6–71.0]	59.0 [42.8–72.0]	61.3 [45.1–73.5]	1.27×10^{-94}	ε^2 = 0.0138

Characteristic	Overall (n = 31,607)	Deficient <30 (n = 5,080)	Insufficient 30–<50 (n = 7,055)	Sufficient 50–<75 (n = 10,045)	Optimal ≥75 (n = 9,421)	*p*	Effect size
Female, n (%)	18,435 (58.3)	2,680 (52.8)	3,794 (53.8)	5,969 (59.4)	5,988 (63.6)	6.83×10^{-51}	V = 0.086
25(OH)D (nmol/L)	59 [38–79]	21 [16–25]	40 [35–45]	62 [56–68]	91 [82–106]	$<1 \times 10^{-300}$	ε^2 = 0.9262
AST (U/L)	25.2 [19.8–34.2]	27.6 [21.0–45.0]	25.2 [19.8–35.4]	24.6 [19.2–31.8]	24.6 [19.8–31.8]	6.22×10^{-85}	ε^2 = 0.0124
AST (µkat/L)	0.42 [0.33–0.57]	0.46 [0.35–0.75]	0.42 [0.33–0.59]	0.41 [0.32–0.53]	0.41 [0.33–0.53]	—	—
ALT (U/L)	25.8 [19.2–37.2]	27.0 [19.2–42.0]	27.0 [19.2–40.2]	25.2 [19.2–36.0]	24.6 [18.6–34.2]	3.02×10^{-35}	ε^2 = 0.0051
ALT (µkat/L)	0.43 [0.32–0.62]	0.45 [0.32–0.70]	0.45 [0.32–0.67]	0.42 [0.32–0.60]	0.41 [0.31–0.57]	—	—
GGT (U/L)	22.2 [13.8–45.0]	31.2 [16.8–78.0]	24.0 [15.0–49.2]	20.4 [13.2–39.0]	19.8 [13.2–36.6]	9.51×10^{-187}	ε^2 = 0.0272
GGT (µkat/L)	0.37 [0.23–0.75]	0.52 [0.28–1.30]	0.40 [0.25–0.82]	0.34 [0.22–0.65]	0.33 [0.22–0.61]	—	—
ALP (U/L)	72.0 [57.6–93.6]	81.0 [63.0–112.2]	73.8 [58.8–96.0]	70.2 [56.4–88.8]	69.6 [55.8–88.8]	7.16×10^{-138}	ε^2 = 0.0201
ALP (µkat/L)	1.20 [0.96–1.56]	1.35 [1.05–1.87]	1.23 [0.98–1.60]	1.17 [0.94–1.48]	1.16 [0.93–1.48]	—	—
De Ritis ratio (AST/ALT)	0.98 [0.76–1.27]	1.05 [0.79–1.47]	0.94 [0.73–1.24]	0.95 [0.75–1.22]	1.00 [0.79–1.26]	3.20×10^{-65}	ε^2 = 0.0095
Platelets (10^9/L)	233 [187–285]	224 [168–284]	235 [188–286]	236 [193–285]	233 [191–284]	8.46×10^{-23}	ε^2 = 0.0034

Characteristic	Overall (n = 31,607)	Deficient <30 (n = 5,080)	Insufficient 30–<50 (n = 7,055)	Sufficient 50–<75 (n = 10,045)	Optimal ≥75 (n = 9,421)	*p*	Effect size
FIB-4 index	1.22 [0.76–1.99]	1.56 [0.89–3.14]	1.17 [0.72–1.96]	1.13 [0.72–1.78]	1.22 [0.79–1.88]	3.10×10^{-130}	$\varepsilon^2 = 0.0198$
FIB-4 > 2.67, n (%)[a]	4,641 (15.4)	1,410 (29.0)	1,028 (15.2)	1,092 (11.4)	1,111 (12.3)	4.59×10^{-189}	V = 0.170
Elevated ALT, n (%)[b]	13,736 (43.5)	2,394 (47.1)	3,260 (46.2)	4,246 (42.3)	3,836 (40.7)	4.06×10^{-18}	V = 0.052
Albumin (g/L)	44.0 [40.0–46.0]	41.0 [35.0–45.0]	44.0 [39.0–46.0]	44.0 [41.0–47.0]	44.0 [42.0–47.0]	5.64×10^{-245}	$\varepsilon^2 = 0.0422$
Creatinine (μmol/L)	74 [63–92]	75 [61–101]	74 [62–91]	73 [62–89]	75 [64–93]	1.14×10^{-14}	$\varepsilon^2 = 0.0021$
eGFR (mL/min/1.73 m^2)	90.5 [68.8–104.0]	87.8 [59.3–103.6]	92.9 [71.5–106.0]	92.0 [73.3–104.8]	88.0 [66.7–101.7]	8.75×10^{-54}	$\varepsilon^2 = 0.0079$
eGFR < 60, n (%)[c]	5,924 (19.0)	1,287 (25.6)	1,208 (17.3)	1,549 (15.6)	1,880 (20.2)	4.38×10^{-51}	V = 0.087
Season: winter, n (%)	9,600 (30.4)	2,045 (40.3)	2,418 (34.3)	2,681 (26.7)	2,455 (26.1)	—	—
Season: spring, n (%)	8,377 (26.5)	1,744 (34.3)	2,110 (29.9)	2,564 (25.5)	1,958 (20.8)	—	—
Season: summer, n (%)	5,517 (17.5)	459 (9.0)	1,037 (14.7)	2,012 (20.0)	2,006 (21.3)	—	—
Season: autumn, n (%)	8,113 (25.7)	832 (16.4)	1,490 (21.1)	2,788 (27.8)	3,002 (31.9)	—	—
Era: pandemic 2020–	20,475 (64.8)	3,010 (59.3)	4,205 (59.6)	6,572 (65.4)	6,685 (71.0)	—	—

Characteristic	Overall (n = 31,607)	Deficient <30 (n = 5,080)	Insufficient 30–<50 (n = 7,055)	Sufficient 50–<75 (n = 10,045)	Optimal ≥75 (n = 9,421)	*p*	Effect size
2023, n (%)							
Age > 70 years, n (%)	9,869 (31.2)	2,017 (39.7)	1,905 (27.0)	2,913 (29.0)	3,034 (32.2)	—	—

[a] Denominator = 30,232 records with computable FIB-4 (by category, among the 31,601 records with an analysable 25(OH)D value: 4,866 / 6,770 / 9,582 / 9,008). [b] Sex-specific ACG “true-normal” thresholds, ALT > 25 U/L in women and > 33 U/L in men [8]; see Methods for the threshold sensitivity analysis. [c] Denominator = 31,253 records with a creatinine result.

Table 2. Spearman rank correlations of serum 25-hydroxyvitamin D with hepatobiliary and reference analytes Spearman rho with Fisher z-transformed 95% confidence intervals; rho^2 is expressed as the percentage of rank variance explained. Exact two-tailed p values. n differs by analyte because of complete-case analysis. Sex-stratified correlations and formal Fisher z tests of the sex difference are in Supplementary Table S2.

Analyte	n	rho	95% CI	Variance explained (rho^2, %)	*p*
Gamma-glutamyltransferase (U/L)	31,601	−0.156	−0.167 to -0.145	2.43	6.4×10^{-171}
Alkaline phosphatase (U/L)	31,601	−0.131	−0.142 to -0.120	1.71	1.8×10^{-120}
Aspartate aminotransferase (U/L)	31,601	−0.090	−0.101 to -0.079	0.82	2.3×10^{-58}
FIB-4 index	30,226	−0.081	−0.092 to -0.069	0.65	7.3×10^{-45}
Alanine aminotransferase (U/L)	31,601	−0.071	−0.082 to -0.060	0.50	2.9×10^{-36}
De Ritis ratio (AST/ALT)	31,601	−0.023	−0.034 to -0.012	0.05	3.6×10^{-5}

Analyte	n	rho	95% CI	Variance explained (rho^2, %)	*p*
Platelets (10^9/L)	30,226	+0.041	+0.030 to +0.052	0.17	8.1×10^{-13}
Age (years)	31,601	−0.049	−0.060 to -0.038	0.24	5.8×10^{-18}
eGFR (mL/min/1.73 m^2)	31,247	−0.025	−0.036 to -0.014	0.06	1.0×10^{-5}
Albumin (g/L)	**26,726**	**+0.189**	**+0.178 to +0.201**	**3.59**	**1.7×10^{-214}**

The largest single association of 25-hydroxyvitamin D in this panel is with serum albumin, a nutrition and inflammation marker, not with any hepatobiliary analyte.

Table 3. Crude and progressively adjusted associations of serum 25-hydroxyvitamin D (per +25 nmol/L) with FIB-4, elevated ALT and the two strongest enzyme outcomes
Continuous outcomes are natural-log transformed and fitted by ordinary least squares with HC3 robust standard errors; coefficients are back-transformed to the percentage change in the outcome per +25 nmol/L. Binary outcomes are fitted by logistic regression; effects are odds ratios per +25 nmol/L with 95% Wald confidence intervals. Attenuation is the percentage reduction in the exposure coefficient (log-odds ratio for binary outcomes) relative to the crude model at that step. n differs between steps because of complete-case analysis. Model 1 = crude; 2 = + age, sex; 3 = + serum albumin; 4 = + season, era; 5 = + eGFR (full model). Panel C calibrates the residual in Panel B against negative-control outcomes at model 4 and uses patient-clustered robust confidence intervals throughout.

Panel A — continuous outcomes, % change per +25 nmol/L (95% CI)

Model	FIB-4 (n)	% change [95% CI]	Atten.	GGT (n)	% change [95% CI]	Atten.	ALP (n)	% change [95% CI]	Atten.
1 Crude	30,226	−5.21 [-5.90, -4.52]	—	31,601	−10.11 [-10.98, -9.23]	—	31,601	−3.98 [-4.40, -3.57]	—
2 + Age, sex	30,226	−2.78 [-3.32, -2.23]	47.3%	31,601	−8.10 [-8.94, -7.25]	20.7%	31,601	−3.62 [-4.03, -3.22]	9.2%

Model	FIB-4 (n)	% change [95% CI]	Atten.	GGT (n)	% change [95% CI]	Atten.	ALP (n)	% change [95% CI]	Atten.
3 + Albumin	25,588	−0.39 [-0.94, +0.17]	92.7%	26,726	−4.65 [-5.55, -3.75]	55.3%	26,726	−2.21 [-2.64, -1.78]	45.0%
4 + Season, era	25,588	−1.21 [-1.77, -0.65]	77.2%	26,726	−5.01 [-5.92, -4.09]	51.8%	26,726	−2.33 [-2.77, -1.89]	42.0%
5 Full (+ eGFR)	**25,504**	**−1.06 [-1.62, -0.49]**	**80.2%**	**26,570**	**−4.59 [-5.50, -3.66]**	**56.0%**	**26,570**	**−2.42 [-2.87, -1.97]**	**39.8%**

Exact p values, model 5: FIB-4 $p = 2.4 \times 10^{-4}$; GGT $p = 1.8 \times 10^{-21}$; ALP $p = 1.3 \times 10^{-25}$. Model 3 for FIB-4 is the only step at which the exposure coefficient is not distinguishable from zero ($p = 0.168$).

Panel B — binary outcomes, odds ratio per +25 nmol/L (95% CI)

Model	FIB-4 > 2.67: n / events	OR [95% CI]	*p*	Atten.[a]	Elevated ALT: n / events	OR [95% CI]	*p*	Atten.[a]
1 Crude	30,226 / 4,641	0.765 [0.743, 0.787]	4.1×10^{-77}	—	31,601 / 13,736	0.938 [0.922, 0.954]	3.0×10^{-13}	—
2 + Age, sex	30,226 / 4,641	0.810 [0.787, 0.834]	1.8×10^{-45}	21.5%	31,601 / 13,736	0.935 [0.919, 0.952]	4.1×10^{-14}	−4.2%
3 + Albumin	25,588 / 4,125	0.960 [0.931, 0.990]	9.6×10^{-3}	84.8%	26,726 / 11,630	0.931 [0.914, 0.950]	7.8×10^{-13}	−10.4%
4 + Season, era	25,588 / 4,125	0.931 [0.902, 0.961]	1.1×10^{-5}	73.5%	26,726 / 11,630	0.921 [0.903, 0.940]	9.7×10^{-16}	−27.3%
5 Full (+ eGFR)	**25,504 / 4,112**	**0.935 [0.906, 0.966]**	**4.2×10^{-5}**	**75.1%**	**26,570 / 11,576**	**0.940 [0.921, 0.959]**	**1.5×10^{-9}**	**3.5%**

[a] Attenuation of the log-odds ratio relative to model 1; negative values indicate that adjustment *strengthened* the association. For context, in model 5 for FIB-4 > 2.67 the odds ratios for the covariates were 1.061 per year of age, 0.632 for female sex, 0.874 per g/L of albumin and 1.872 for the 2020–2023 era — each substantially larger than the exposure

effect. Sensitivity analysis at higher ALT thresholds (0.58/0.74 μkat/L, i.e. 34.8/44.4 U/L; 22.5% of records flagged) gave crude OR 0.896 [0.877, 0.916] and fully adjusted OR 0.940 [0.917, 0.964].

Panel C — negative-control outcomes at the same adjustment step (model 4), odds ratio per +25 nmol/L (95% CI)

Calibration of the residual in Panel B. Each row is a routine analyte co-ordered in the same specimen whose concentration vitamin D does not set (Methods, "Negative-control outcomes"), fitted with the identical exposure scaling, the identical adjustment set — age, sex, serum albumin, season and era, i.e. model 4 — and the identical estimator. Rows are ordered by |OR – 1|. Model 5 is not shown because it adds eGFR, which the controls do not carry.

Outcome	n / events	OR [95% CI][b]	*p*	|OR – 1|
Low bicarbonate (total CO_2 < 22 mmol/L)[c]	4,745 / 569	1.039 [0.963, 1.122]	0.318	0.039
High mean platelet volume (> 11 fL)	25,553 / 2,225	0.953 [0.917, 0.991]	0.015	0.047
Hyponatraemia (Na < 135 mmol/L)	25,641 / 1,845	1.050 [1.009, 1.092]	0.016	0.050
Hypochloraemia (Cl < 98 mmol/L)	18,038 / 1,261	1.114 [1.069, 1.162]	3.6×10^{-7}	0.114
Hyperkalaemia (K > 5.0 mmol/L)	26,168 / 1,940	1.202 [1.154, 1.251]	4.6×10^{-19}	0.202
FIB-4 > 2.67 (this paper, model 4)	**25,586 / 4,123**	**0.931 [0.896, 0.968]**	**3.4×10^{-4}**	**0.069**

Four of the five controls reach $p < 0.05$; their median |OR – 1| is 0.050 and the range is 0.039 to 0.202. This paper's own residual lies **inside** that range at its **60th percentile** — larger than three controls and smaller than two. The falsification test set in advance (residual larger than *every* clean control) is **not** met. The band bounds the residual; it does not show it to be zero — total calcium, which vitamin D genuinely regulates, also falls inside the band after the same adjustment (Results; Supplementary Table S14).

[b] Patient-clustered (CR0) robust confidence intervals throughout this panel, applied identically to the controls and to the study outcome; the interval for FIB-4 > 2.67 is

therefore wider than the model-based interval given for the same odds ratio in Panel B, and n differs by two records because all rows are fitted on the common complete-case sample. [c] Total CO_2 was measured in only 15.1% of the cohort, so this control rests on a differently selected subsample and its estimate is the least comparable of the five. The contaminated controls deliberately excluded from this band, the extended nine-control band, the continuous forms and all four cascade steps are in Supplementary Table S14.

Supplementary Files

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

- SupplementaryInformationREVISED.pdf