

Journal Pre-proof

Fluctuating biomarkers in primary sclerosing cholangitis: a longitudinal comparison of alkaline phosphatase, liver stiffness, and ELF

Guri Fossdal, Anders Batman Mjelle, Kristine Wiencke, Ida Bjørk, Odd Helge Gilja, Trine Folseraas, Tom Hemming Karlsen, William Rosenberg, Lasse Melvær Giil, Mette Vesterhus

PII: S2589-5559(21)00104-X

DOI: <https://doi.org/10.1016/j.jhepr.2021.100328>

Reference: JHEPR 100328

To appear in: *JHEP Reports*

Received Date: 18 December 2020

Revised Date: 25 May 2021

Accepted Date: 16 June 2021

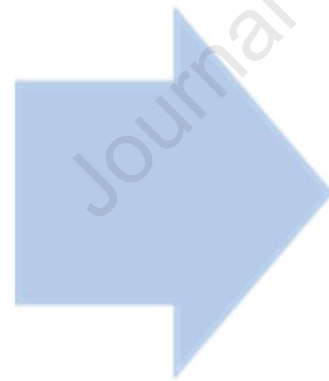
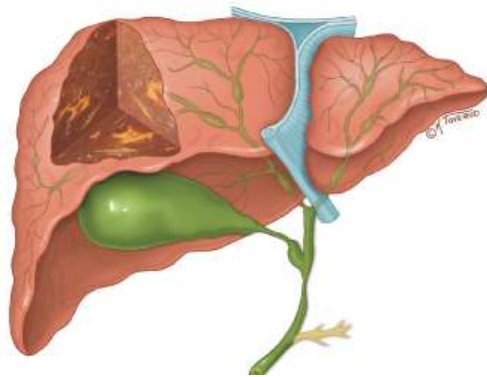
Please cite this article as: Fossdal G, Mjelle AB, Wiencke K, Bjørk I, Gilja OH, Folseraas T, Karlsen TH, Rosenberg W, Giil LM, Vesterhus M, Fluctuating biomarkers in primary sclerosing cholangitis: a longitudinal comparison of alkaline phosphatase, liver stiffness, and ELF, *JHEP Reports* (2021), doi: <https://doi.org/10.1016/j.jhepr.2021.100328>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

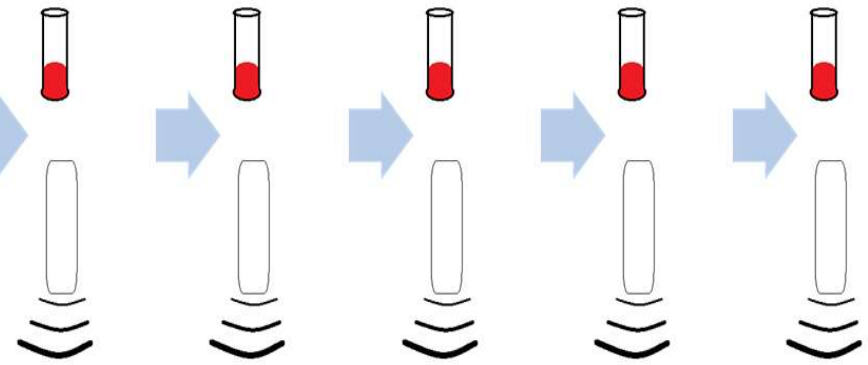
© 2021 Published by Elsevier B.V. on behalf of European Association for the Study of the Liver (EASL).



Graphical abstract



ELF & ALP: between-patient variation



Liver stiffness: within- & between-patient variation

Fluctuating biomarkers in primary sclerosing cholangitis: a longitudinal comparison of alkaline phosphatase, liver stiffness, and ELF

Guri Fossdal^{1,2,3}, Anders Batman Mjelle⁴, Kristine Wiencke^{1,5,6}, Ida Bjørk⁷, Odd Helge Gilja^{4,8}, Trine Folseraas^{1,5,6}, Tom Hemming Karlsen^{1,5,6,9}, William Rosenberg¹⁰, Lasse Melvær Giil³, Mette Vesterhus^{1,2,3}

1. Norwegian PSC Research Centre, Department of Transplantation Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway;
2. Department of Clinical Science, University of Bergen, Bergen, Norway;
3. Department of Medicine, Haraldsplass Deaconess Hospital, Bergen, Norway;
4. Department of Clinical Medicine, University of Bergen, Bergen, Norway;
5. Section of Gastroenterology, Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway;
6. Research Institute of Internal Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway;
7. Department of Radiology, Oslo University Hospital Rikshospitalet, Oslo, Norway;
8. National Centre for Ultrasound in Gastroenterology, Department of Medicine, Haukeland University Hospital, Bergen, Norway
9. Institute of Clinical Medicine, University of Oslo, Oslo, Norway;
10. UCL Institute for Liver and Digestive Health, University College London & Royal Free London NHS Foundation Trust, London, UK

Corresponding Author: Ass. Prof. Mette Vesterhus, Department of Medicine, Haraldsplass

Deaconess Hospital, Ulriksdal 8, 5009 Bergen, Norway.

Tel.: +47-90748643 E-mail: vm@ihelse.net

Word count: 4872

Total number of figures and tables: 7

Abbreviations:

ALP:	Alkaline phosphatase;
ALT:	Alanine aminotransferase;
APRI:	The Aspartate Aminotransferase-to-Platelet Ratio Index;
AST:	Aspartate aminotransferase;
BMI:	Body mass index;
CRP:	C-reactive protein;
ELF:	Enhanced liver fibrosis;
ERCP:	Endoscopic retrograde cholangiopancreatography;
FIB-4:	Fibrosis-4 score;
GGT:	Gamma-glutamyl transferase;
HA:	Hyaluronic acid;
IBD:	Inflammatory bowel disease;
ICC:	Intraclass correlation;
INR:	International normalized ratio;
IQR/M:	Interquartile range divided by the median;
LSM:	Liver stiffness measurement;
MRCP:	Magnetic resonance cholangiopancreatography;
PIIINP:	Propeptide of type III procollagen;

pSWE:	Point shear wave elastography;
PSC:	Primary sclerosing cholangitis;
TE:	Transient elastography;
TIMP-1:	Tissue inhibitor of metalloproteinases-1;
UDCA:	Ursodeoxycholic acid

Keywords: Primary sclerosing cholangitis; alkaline phosphatase; elastography; liver stiffness; enhanced liver fibrosis test; biomarker; risk stratification

Conflict of interest: W. Rosenberg is one of the inventors and patent holders of the ELF test.

Financial support: The work is part of the PhD program for GF funded by the Western Norway Regional Health Authority.

Authors' contributions: MV is the guarantor of the article and supervised the project. MV, LMG, WR, and THK conceived and designed the study. MV, KW, and TF collected the biological samples and clinical data. MV, ABM, and IB performed the ultrasound scans and liver stiffness measurements. OHG contributed to the liver stiffness measurements. WR contributed to the ELF test laboratory analyses. GF and LMG designed and performed the statistical analyses. GF, ABM, TF, THK, LMG, and MV contributed to the interpretation of the data. GF, LMG, and MV drafted the manuscript. All authors reviewed the manuscript for critical content and approved the final version of the manuscript.

ABSTRACT

Background & Aims: Primary sclerosing cholangitis (PSC) is a progressive liver disease characterized by fluctuating liver biochemistries and highly variable disease progression. The Enhanced Liver Fibrosis (ELF) test and liver stiffness measurements (LSM) reflect fibrosis and predict clinical outcomes in PSC; however, longitudinal assessments are missing. We aimed to characterize the systematic change in ELF and LSM over time in a prospective PSC cohort, along with their longitudinal relationship to alkaline phosphatase (ALP) and bilirubin.

Methods: We included 113 non-transplant PSC patients (86 males [76.1%]; mean age 43.3 ± 15.7 years) with annual study visits between 2013 and 2019 at two Norwegian centers. ELF test, LSM, clinical data, liver biochemistries, and revised Mayo risk score were measured. We used linear mixed-effects models to estimate change over time, intraclass correlations (ICCs), as well as their relationship with ALP and bilirubin.

Results: At baseline, the median (range) ELF test was 9.3 (7.5-12.9) and median LSM 1.26 m/s (0.66-3.04 m/s). ELF and LSM increased over time (0.09 points/year, 95% CI [0.03-0.15], $p=0.005$, vs. 0.12 points/year, 95% CI [0.03-0.21], $p=0.009$). Between-patient effects explained 78% of ELF variation (ICC 0.78), and 56% of LSM variation (ICC 0.56). ALP also increased and showed the highest ICC (0.86).

Conclusions: ELF and LSM increased over a 5-year period. Longitudinal analyses demonstrated differences regarding within- and between-patient effects suggesting that the ELF test may have superior reliability for risk stratification compared to LSM in PSC.

LAY SUMMARY/ KEY POINTS

Primary sclerosing cholangitis is characterized by substantial disease variability between patients and fluctuating liver biochemistries. Hence, new biomarkers are needed to identify individuals with an increased risk of developing end-stage liver disease. We explore the change over time of two putative prognostic biomarkers in PSC, the serum ELF test and liver stiffness measurements by ultrasound, demonstrating differences that may reflect differing abilities to discriminate risk.

INTRODUCTION

Primary sclerosing cholangitis (PSC) is characterized by multifocal strictures and dilatations of the biliary tree due to inflammation and biliary fibrosis, ultimately progressing to end-stage liver disease¹⁻³. The natural course of PSC is highly variable, with median transplant-free survival ranging from 13 to 20 years^{2,4,5}. A major unmet need is the lack of established biomarkers to (a) gauge changes in disease activity that reflect the pathophysiological processes involved in PSC, (b) identify high-risk patients for risk stratification and prognostication, and (c) evaluate treatment effects before reaching clinical endpoints.

Alkaline phosphatase (ALP) has been applied widely to predict clinical disease progression, select patients for clinical trials, and as a surrogate outcome marker in treatment studies.

Elevated ALP is a consistent marker of poor outcomes at the group level across several studies⁶⁻⁹. However, longitudinal fluctuation in ALP limits its use at the individual level.

Thus, there is a need to identify more accurate biomarkers with less fluctuation over time.

The Enhanced Liver Fibrosis (ELF) test and liver stiffness measurements (LSM) are emerging biomarkers for risk prediction and evaluation of treatment effects in clinical trials in PSC^{10,11}.

They both reflect fibrosis severity but are based on different approaches. The ELF test is a serum-based biomarker panel measuring three direct markers of extracellular matrix remodelling and fibrosis^{12,13}. In contrast, LSM assesses the physical, viscoelastic properties of the liver using ultrasound-based elastography methods¹⁴. Both the ELF test and LSM have been shown to predict transplant-free survival in PSC across independent studies¹⁵⁻¹⁹.

However, studies assessing repeated measurements are limited and have not established whether ELF or LSM change systematically over time in a similar fashion to each other or similar to ALP. Furthermore, it is not known whether ELF or LSM fluctuate together with ALP.

Therefore, we aimed to characterize the longitudinal change in ELF and LSM compared to ALP in a prospective PSC cohort. We also aimed to evaluate the relative contributions of intra- and inter-individual variation for each of these variables using repeated measurements. Finally, we sought to establish the longitudinal associations between ELF, LSM, ALP, and bilirubin.

METHODS

Study design

We prospectively included 113 non-transplant PSC patients during 2013-2018 from two Norwegian centers: Haukeland University Hospital, Bergen and Oslo University Hospital, Rikshospitalet, Oslo. The diagnosis of PSC was based on characteristic findings on magnetic resonance cholangiography or endoscopic retrograde cholangiopancreatography according to established diagnostic criteria²⁰. The first pathological radiologic finding defined the time of PSC diagnosis. Eight patients with PSC and features of autoimmune hepatitis were included. Patients with small duct PSC were excluded. Inflammatory bowel disease was diagnosed based on endoscopy and histological findings according to accepted criteria²¹. Clinical and demographic information, including laboratory data, was acquired from patient records and research databases. Liver biochemistry, ELF test, and elastography were sampled annually (\pm 1 month from study visit) from the baseline visit. All patients provided informed written consent. The study was in accordance with the Declaration of Helsinki and approved by the Regional committees for medical and health research ethics of Western and South-Eastern Norway (Reference 2012/2214/REK VEST and 2008/8670, respectively).

Laboratory analyses

Biochemical analyses were performed by standard laboratory protocols, including hemoglobin, leukocytes, platelets, international normalized ratio (INR), aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), total bilirubin, albumin, creatinine, immunoglobulin 4 (IgG4), C-reactive protein (CRP). The Mayo risk score and the Fibrosis-4 Index for Liver Fibrosis (FIB-4 score) were calculated using published algorithms²²⁻²⁴.

ELF test

Frozen serum samples were collected from the 113 patients from two biobanks in Bergen and Oslo, following an identical protocol. The ELF test was analyzed using the commercially available kit, Siemens ELF®Test, performed on an ADVIA Centaur XP analyzer (Siemens Medical Solutions Inc., Tarrytown, NY, USA). The ELF test was calculated according to the published algorithm, including the levels of hyaluronic acid (HA), the propeptide of procollagen type III (PIIINP), and tissue inhibitor of matrix metalloproteinases-1 (TIMP-1) using the formula: $\text{ELF test} = 2.278 + 0.851 \ln (C_{\text{HA}}) + 0.751 \ln (C_{\text{PIIINP}}) + 0.394 \ln (C_{\text{TIMP-1}})$.

Elastography

Point shear wave elastography (pSWE) was performed using an ElastPQ® Philips iU22 (Philips Healthcare, Andover, MA, USA) scanner (software version 6.3.2.2, convex C5-1 probe) and ARFI®, Siemens Acuson S3000 (Siemens Medical Solutions USA, Inc., Malvern, PA), in the Bergen and Oslo cohorts, respectively. The examination was performed following international guidelines, including at least three hours of fasting before examination²⁶. Following a B-mode ultrasound scan of the liver and spleen, LSM was measured using a right

intercostal approach during relaxed mid-respiration breath-hold with patients in the supine position, with their right hand beneath the head.

A region of interest (ROI) representing a 0.5 x 1.5 cm sample volume was placed two to six cm below the liver capsule in an area where homogenous liver parenchyma could be visualized, avoiding large vessels and bile ducts. LSM was based on the median of ten acquisitions and considered valid when the success rate was equal to or above sixty percent. LSM was measured in meters per second (m/s). The published cut-off value of 4.9 kPa (~1.28 m/s) was used to stratify patients for subgroup analyses²⁵. Liver stiffness is expressed as shear wave speed (m/s) or converted into Young's modulus using the equation $\text{kPa} = 3[(\text{ms}^{-1})^2]$ ²⁶. Each patient was followed by a single elastography platform.

Statistics

P-values < 0.05 were considered statistically significant. Continuous variables were evaluated for approximate normality using Q-Q plots and presented as means and standard deviations (SDs) or medians and interquartile ranges (IQRs) as appropriate. Due to significant right skewness, logarithmic transformations were applied to liver biochemistries, ELF, and LSM. Transformation resulted in approximate normality as assessed by Q-Q plots, in line with the assumptions of parametric statistical models. Mann-Whitney U test, Student's t-test, and Chi-squared test were applied as appropriate. Correlations at study baseline were tested by the Spearman rank correlation due to the non-normality of variables and illustrated graphically as a correlation network.

We used a linear mixed model with an unstructured covariance structure for repeated measurement analyses with random intercept and random slope. Intraclass correlation

coefficients (ICCs) were estimated from an empty-means linear mixed-effect model. We used a two-step approach to characterize the associations between LSM, ELF, ALP, and bilirubin in a multilevel context. First, the random intercepts, slopes, and residuals from a multilevel model, either ALP or bilirubin, were estimated and scaled to z-scores. By standardizing the variables to a mean of zero and a standard deviation of one, the biomarkers are on the same scale with comparable effect sizes. The resulting positive or negative z-score will represent the magnitude of increase or decrease, respectively, in the effect size for all variables. The z-scores were subsequently entered as predictors in a second multilevel model, where they represent between-person differences (random intercepts), between-person linear rate of change (random slopes), and fluctuations (the remaining residuals) ²⁷. For the relationship between LSM and ELF, we were able to fit a multilevel structural equation model with random intercepts only using both LSM and ELF as separate outcomes. We estimated the correlation between the intercepts and residuals, representing the between-person and within-person correlations. The model was adjusted for time in study. Missing values were assumed to be missing at random. Data were pooled for the two different elastography modalities as individual patient trajectories were followed longitudinally using a single platform; there were no significant differences between the two cohorts ($p=0.39$).

Post hoc analyses were performed for defined subgroups. Subgroups for liver fibrosis stages F0-2 and F3-4 were defined using the published cut-off value of 4.9 kPa (~ 1.28 m/s) for pSWE in PSC ²⁵. For further subgroup analyses, the cohort was divided according to presumed high-risk profiles at baseline ^{8-10,13,15,28,29}, i.e., ALP $\geq 1.5 \times$ (ULN), ELF level ≥ 9.8 , and for discrimination between mild and advanced fibrosis corresponding to METAVIR score F0-2 vs. F3-4; LSM ≥ 1.28 m/s, as outlined in **Table 1**. The analyses were conducted using

SPSS version 26 (SPSS Inc., 2016, Armonk, NY) and STATA 16 (StataCorp. 2019, Stata

Statistical Software: Release 16.1. College Station, Tx: StataCorp LP) for all analyses. The correlation network was generated using the qgraph package in R (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Patient characteristics are outlined in **Table 1**. We included 113 PSC patients (86 males; 76.1%). Their mean age at baseline was 43 years (15.7), with a 4-year median duration of PSC and a median follow-up time of 4.5 years. Median time from study visit to LSM was 0 months (SD 1.33 and 2.33 for the two cohorts, respectively). Clinical events are listed in **Suppl. Table S1**.

Baseline ELF test, liver stiffness, and ALP values

At baseline, the patients had median (IQR) ELF 9.3 (1.34), LSM 1.26 m/s (0.52) and ALP 151.5 U/l (197). There was no significant difference between males and females. There were 37 (33%), 50 (45%), and 52 (46%) high-risk patients defined by ELF test, LSM, or ALP, respectively. Correlation analysis showed a strong correlation of liver parameters, as illustrated by a network diagram (**Figure 1**). The liver enzymes ALT, AST, GT, and ALP were strongly correlated; ELF and LSM showed moderate correlation with each other (ρ 0.483, $p < 0.001$), and both were correlated with ALP, other liver enzymes as well as bilirubin and (negatively) albumin.

Longitudinal change and intraclass correlation coefficients

The development over time for the ELF test, LSM, ALP, and bilirubin is illustrated in **Figure 2**. Using a linear mixed-effects model, we demonstrated a small, but significant increase over

five years for ELF (0.09 points/year, 95% CI [0.03-0.15], $p = 0.005$) and LSM (0.12 points/year, 95% CI [0.03-0.21], $p = 0.009$). Scaling of the outcome variables to z-scores demonstrated a slightly larger increase in LSM (0.07 SD per year, 95% CI [0.02, 0.13]) compared to ELF (0.06 SD per year, 95% CI [0.03, 0.20]). By comparison, ALP increased 0.04 SD per year (95% CI [0.01, 0.07], $p = 0.011$) and bilirubin increased 0.07 SD per year (95% CI [0.02, 0.12], $p = 0.007$). The ICC was highest for ALP (0.86) and ELF (0.78), with lower ICCs for bilirubin (0.64) and LSM (0.56). The results are summarized in **Table 2**.

Longitudinal change over time in high-risk subgroups

Post hoc subgroup analyses of predefined high-risk groups, i.e., ELF test ≥ 9.8 , LSM ≥ 1.28 m/s, and ALP $\geq 1.5 \times$ ULN at baseline, demonstrated a significantly higher baseline ELF level among the high-ALP group compared to the low-ALP group ($p = 0.001$) and a similar trend for LSM ($p = 0.06$). Both ELF and LSM increased significantly over time in the high-ALP group ($p = 0.014$ and 0.022 , respectively), whereas they showed no significant increase in the low-ALP group (**Figure 3**). However, the interaction between time and ALP subgroup did not reach significance. There were no significant differences in the change in ELF or LSM over time, according to the baseline risk groups defined by ELF or LSM (data not shown).

UDCA treatment was received by 35% of the patients at any time during the study with a median duration of 3.4 (range 1-6 years) of treatment. Subgroup analysis indicated that ELF and ALP increased significantly over time in UDCA naïve but not UDCA-treated patients (ELF: $p = 0.009$ vs 0.803 ; ALP: $p = 0.008$ vs $p = 0.883$), with a similar trend for LSM ($p = 0.057$ vs 0.125); however, data were insufficient to adjust analyses for the biomarker*treatment interaction. Endoscopic interventions ($n = 10$ in six patients) during the study were not associated with consistent changes in ELF at subsequent visits.

Longitudinal association between ELF and LSM

Using a multi-outcome multilevel structural equation model adjusted for time, the correlation between the random intercepts of ELF and LSM was good (0.79, $p < 0.001$), representing the between-person association between LSM and ELF. In contrast, the correlation coefficient of the residuals was weak (0.24, $p = 0.007$), representing the within-person association between LSM and ELF.

Longitudinal association between ELF test or LSM and liver biochemistries and Mayo Risk Score

Over time, liver biochemistries and Mayo risk score were significantly associated with LSM and ELF outcomes (**Table 3**). ALP showed stronger association with ELF (standardized fixed effect (sFE) 0.47) than with LSM (sFE 0.28). Similarly, ELF showed a stronger association than LSM with Mayo risk score (sFE 0.48 vs. 0.37) and the FIB-4 score (sFE 0.56 vs. 0.42). LSM was more associated with bilirubin (sFE 0.29) than ELF (sFE 0.20), while ELF and LSM showed similar associations with albumin. The effect size sFE can be interpreted similarly in magnitude as correlation coefficients.

Between- and within-person associations between ALP, bilirubin, LSM, and ELF

Variation in the individual means of ALP and bilirubin accounted for most of the association between ALP, bilirubin, and ELF (**Table 4**). By comparison, variation in the annual rate of change in ALP and bilirubin was not associated with ELF. However, we identified a smaller but significant association between fluctuations in ALP and ELF. For LSM, variation in individual means accounted for most of the association between ALP, bilirubin, and LSM,

whereas there was no association with fluctuations in ALP or bilirubin. However, a higher annual rate of change in bilirubin was associated with higher LSM scores.

Spontaneous reductions in ELF, LSM, and ALP

The subpopulation with ALP ≥ 1.5 x ULN accounted for all of the patients with $\geq 40\%$ ALP reduction at each of the visits in our study. Out of the high-ALP group, a total of 13, 13, 10 and 6% experienced $\geq 40\%$ ALP reduction at visits 1, 2, 3, and 5 years from baseline, respectively.

In 40% of the total patient cohort, ELF levels decreased from baseline to 5 years, with a mean value of -0.67. A similar proportion of patients (44.7 and 42.2%) showed a reduction in ELF levels within the same range (mean change -0.51 and -0.54) at 1 and 2 years from baseline. Reduction in LSM was shown in 34% of the patients at 5 years (mean change -0.29 m/s); similar proportions of patients demonstrated LSM reduction at 1 and 2 years from baseline (42.7 and 36.7%, respectively; mean change of -0.33 to -0.38 m/s). Among the patients with 5 years follow-up time, all remained in the same category concerning low or high levels of ELF or LSM, whereas 16% of the patients moved between categories of low to high ALP as defined by ALP ≥ 1.5 x ULN at baseline). At each follow-up visit (1-5 years from baseline), about 10% of patients featured a concomitant reduction in all of ELF, LSM, and ALP (**Suppl. Table S3**) out of which only 25% received UDCA. Six patients received a total of ten endoscopic treatments during the study period, of which only two procedures were followed by significant ALP reductions.

DISCUSSION

To our knowledge, this is the first study to provide an in-depth characterization of the variation over time in ELF and LSM as well as ALP in a prospective PSC cohort, allowing differentiation of “background noise” (random variation) from biological significant variation. ELF and LSM demonstrated a significant but minor increase over 5-years, in line with previous reports in PSC patients with mild fibrosis^{9,17,28}. Using standardized z-scores in a linear mixed model, our results suggest that LSM increased more than ELF and ALP over time. We demonstrated a strong between-person association between LSM and ELF but a weak association for individual fluctuations over time. Overall, our study indicates that ELF and LSM may stratify similar patients to high-risk groups at baseline whereas there may be different effects driving change in ELF and liver stiffness over time.

Using ICC analyses yielded by the mixed model, we demonstrated essential differences between ELF and LSM regarding between- and within-person effects influencing variation in these parameters. Whereas ELF showed high ICC, suggesting predominant between-person variation, between- and within-person variations contributed relatively equally for LSM. The relatively stable values within individual patients at repeated measurements for ELF support ELF as a reliable risk stratification marker and may imply that the ELF test is superior over LSM for risk stratification purposes when measured at a single time point. Biologically this is plausible, as ELF test reflects three direct markers of extracellular matrix remodelling, providing a biological link to disease severity, in contrast to LSM which represents the sum of several factors affecting liver stiffness.

For a test to be useful for monitoring purposes, the “noise-to-signal” should be low; i.e., any change should reflect a biological difference. Establishment of the magnitude of variation

between and within patients is, therefore, a key factor for assessing the qualities of biomarkers. The ICC from the mixed model represents a measure of within- and between-variation in a test at a single time point and longitudinally. In general, a higher ICC-value represents a lower degree of variation²⁹, reflecting a stronger ability to stratify risk between individuals at a single time point, whereas a lower ICC suggests higher sensitivity to biological variation over time, relevant for monitoring and assessment of treatment effect. However, interobserver variation and other factors may also contribute to lower ICC. Our findings are in line with quality assessments of ELF, which have shown good stability and a low coefficient of variation¹². The lower within-person variation for ELF compared to LSM may partly reflect the inherent differences between patented laboratory assays such as the ELF test compared to ultrasound-based LSM.

As a small note of caution, the ICC of ALP was higher than that of ELF, yet ALP is notoriously fluctuating over time in PSC patients. This trait is a major challenge limiting the use of ALP in individual prognostication and monitoring of disease activity. In the decomposed mixed model analysis, we identified concurrent fluctuations in ALP and ELF, which might suggest similar underlying mechanisms behind fluctuations in both parameters. Possibly, ELF may not overcome the problems of individual fluctuation typical for ALP. In favor of ELF towards LSM, we demonstrated stronger associations for ELF with ALP and other liver biochemistries, as well as the Mayo risk score and FIB-4 score.

For LSM, a lower ICC indicated that within-person variation explained a larger proportion of the variability compared to the ELF test, reflecting either improved sensitivity to detect biologically relevant changes or increased sampling variability. LSM has previously demonstrated good agreement towards histological stages of fibrosis and clinical outcome in

PSC¹⁷⁻¹⁹, and a strong predictive ability for clinical outcomes in independent studies^{17,18}.

Moreover, the elastography modalities we used (pSWE and ARFI quantification) were reported to correlate well with histology^{19,31-33}, demonstrated high accuracy in discriminating between lower and higher degrees of fibrosis³²⁻³⁴, and excellent correlation to TE in PSC patients²⁵. Due to lack of power for end-point analyses, we cannot decipher whether the larger relative contribution of within-patient effects on variability is due to sampling variability or reflect biological variation over time. Inter- and intraobserver variability is an acknowledged possible bias in all ultrasound-based methods^{25,35-37}. Furthermore, the patchy disease distribution in PSC and variation in cholestasis may contribute to variations in LSM^{38,39}. Based on our results, we cannot rule out that the lower ICC for LSM results from increased measurement variability rather than reflecting a relevant change in fibrosis. The significant linear association between bilirubin levels and LSM over time but no association between their intermediate fluctuations indicates that limited segmental cholestasis in PSC does not severely affect LSM over time. This might suggest that ELF and LSM act as complementary biomarkers, indicative of slightly different aspects of the disease concerning fibrosis and cholestasis.

Interestingly, in a post hoc subgroup analysis, we found that patients with an ALP level ≥ 1.5 x ULN at baseline demonstrated elevated baseline levels as well as a significant increase in ELF over time in the high-ALP compared to the low-ALP group. These findings support previous reports proposing this ALP level as an appropriate cut-off level for risk stratification^{6,7,40}.

Clinical trials in patients with PSC are suffering from a lack of robust surrogate markers to reliably evaluate the effect of novel therapeutic agents. Reduction in ALP is commonly used as an outcome parameter in pharmacological studies; however, spontaneous reductions in

ALP challenge the use of ALP as a surrogate marker in PSC^{7,8,41,40}. While a reduction of ALP by 40% or more is a commonly applied primary outcome, this is questioned by reports of patients showing ALP reductions not supported by reductions in histological fibrosis⁹. In the present study, we found that about 8% of the patients experienced spontaneous ALP reductions of at least 40% at 1, 2, and 3 years of study follow-up, respectively. These time points are commonly applied when designing clinical trials, underscoring the challenges of using ALP reduction as a surrogate endpoint. Furthermore, we demonstrated that between one-third and nearly one-half of the patients showed spontaneous reductions in ELF test and LSM, respectively, during the same time frame. Moreover, we identified a subgroup of about 10% of patients at each follow-up visit showing a concomitant reduction in ALP, ELF, and LSM, raising the question of whether the fibrosis level or disease stage may actually regress in PSC. These findings warrant further investigation prior to considering these biomarkers as surrogate endpoints in clinical trials.

UDCA treatment has been associated with ALP reduction in PSC patients in clinical studies⁴²⁻⁴⁴. We did not demonstrate ALP, ELF, or LSM reduction associated with UDCA; however, subgroup analysis showed significant increases in ELF and ALP over time in UDCA naïve (65%) but not UDCA treated (35%) patients. Moreover, UDCA-users had higher levels of ELF, LSM, ALP, and bilirubin at baseline, suggesting a more advanced disease in this group. Unfortunately, our study was not powered to investigate biomarker*treatment interactions.

Limitations of the study

The major limitation of this study is the limited number of long-term clinical outcomes such as deaths and liver transplantations, precluding endpoint analyses. Liver biopsies allowing direct assessment of the degree of liver fibrosis were also not available. However, in PSC, liver biopsies are poorly representative due to the patchy disease distribution, and the procedure carries a risk of adverse outcomes. Current guidelines do not recommend liver biopsies; hence, this was considered unethical.

Conclusion

ELF test and LSM increased slightly but significantly over five years in a prospective PSC patient panel. Our longitudinal analyses demonstrated differences regarding within- and between-patient effects suggesting that ELF test may be more stable than LSM and is likely to perform better for risk stratification in PSC using single measurements. We advocate that the ELF test may hold practical utility for identification of PSC patients with a high risk of disease progression. ELF and LSM showed a significant increase over time only in patients with $ALP \geq 1.5 \times ULN$, supporting this as a relevant cut-off level for risk stratification. The significance of concomitant reductions in ELF, LSM, and ALP in a patient subgroup warrants further studies.

Acknowledgments: The authors thank Liv Wenche Thorbjørnsen for assistance in the collection of serum samples.

References

1. Broome U, Olsson R, Loof L, Bodemar G, Hulterantz R, Danielsson A, et al. Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. *Gut*. 1996;38(4):610-615.
2. Ponsioen CY, Vrouenraets SM, Prawirodirdjo W, Rajaram R, Rauws EA, Mulder CJ, et al. Natural history of primary sclerosing cholangitis and prognostic value of cholangiography in a Dutch population. *Gut*. 2002;51(4):562-566.
3. Eaton JE, McCauley BM, Atkinson EJ, Juran BD, Schlicht EM, de Andrade M, et al. Variations in primary sclerosing cholangitis across the age spectrum. *J Gastroenterol Hepatol*. 2017;32(10):1763-1768.
4. Aadland E, Schrumpf E, Fausa O, Elgjo K, Heilo A, Aakhus T, et al. Primary sclerosing cholangitis: a long-term follow-up study. *Scand J Gastroenterol*. 1987;22(6):655-664.
5. Boonstra K, Weersma RK, van Erpecum KJ, Rauws EA, Spanier BW, Poen AC, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. *Hepatology*. 2013;58(6):2045-2055.
6. Al Mamari S, Djordjevic J, Halliday JS, Chapman RW. Improvement of serum alkaline phosphatase to <1.5 upper limit of normal predicts better outcome and reduced risk of cholangiocarcinoma in primary sclerosing cholangitis. *J Hepatol*. 2013;58(2):329-334.
7. Rupp C, Rossler A, Halibasic E, Sauer P, Weiss KH, Friedrich K, et al. Reduction in alkaline phosphatase is associated with longer survival in primary sclerosing cholangitis, independent of dominant stenosis. *Aliment Pharmacol Ther*. 2014;40(11-12):1292-1301.
8. de Vries EM, Wang J, Leeflang MM, Boonstra K, Weersma RK, Beuers UH, et al. Alkaline phosphatase at diagnosis of primary sclerosing cholangitis and 1 year later: evaluation of prognostic value. *Liver Int*. 2016;36(12):1867-1875.

9. Trivedi PJ, Muir AJ, Levy C, Bowlus CL, Manns M, Lu X, et al. Inter- and Intra-individual Variation, and Limited Prognostic Utility, of Serum Alkaline Phosphatase in a Trial of Patients With Primary Sclerosing Cholangitis. *Clin Gastroenterol Hepatol*. 2020.
10. Ponsioen CY, Lindor KD, Mehta R, Dimick-Santos L. Design and Endpoints for Clinical Trials in Primary Sclerosing Cholangitis. *Hepatology*. 2018;68(3):1174-1188.
11. Ponsioen CY, Chapman RW, Chazouilleres O, Hirschfield GM, Karlsen TH, Lohse AW, et al. Surrogate endpoints for clinical trials in primary sclerosing cholangitis: Review and results from an International PSC Study Group consensus process. *Hepatology*. 2016;63(4):1357-1367.
12. Rosenberg WM, Voelker M, Thiel R, Becka M, Burt A, Schuppan D, et al. Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology*. 2004;127(6):1704-1713.
13. Parkes J, Roderick P, Harris S, Day C, Mutimer D, Collier J, et al. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. *Gut*. 2010;59(9):1245-1251.
14. Dietrich CF, Bamber J, Berzigotti A, Bota S, Cantisani V, Castera L, et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Liver Ultrasound Elastography, Update 2017 (Long Version). *Ultraschall Med*. 2017;38(4):e16-e47.
15. Vesterhus M, Hov JR, Holm A, Schrumpf E, Nygard S, Godang K, et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. *Hepatology*. 2015;62(1):188-197.
16. de Vries EMG, Färkkilä M, Milkiewicz P, Hov JR, Eksteen B, Thorburn D, et al. Enhanced liver fibrosis test predicts transplant-free survival in primary sclerosing cholangitis, a multi-centre study. *Liver Int*. 2017;37(10):1554-1561.

17. Corpechot C, Gaouar F, El Naggar A, Kemgang A, Wendum D, Poupon R, et al. Baseline values and changes in liver stiffness measured by transient elastography are associated with severity of fibrosis and outcomes of patients with primary sclerosing cholangitis. *Gastroenterology*. 2014;146(4):970-979; quiz e915-976.
18. Ehlken H, Wroblewski R, Corpechot C, Arrive L, Rieger T, Hartl J, et al. Validation of Transient Elastography and Comparison with Spleen Length Measurement for Staging of Fibrosis and Clinical Prognosis in Primary Sclerosing Cholangitis. *PLoS One*. 2016;11(10):e0164224.
19. Mjelle AB, Mulabecirovic A, Hausken T, Havre RF, Gilja OH, Vesterhus M. Ultrasound and Point Shear Wave Elastography in Livers of Patients with Primary Sclerosing Cholangitis. *Ultrasound Med Biol*. 2016;42(9):2146-2155.
20. European Association for the Study of the L. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. *J Hepatol*. 2009;51(2):237-267.
21. Podolsky DK. Inflammatory Bowel Disease. *New England Journal of Medicine*. 2002;347(6):417-429.
22. Kim WR, Therneau TM, Wiesner RH, Poterucha JJ, Benson JT, Malinchoc M, et al. A revised natural history model for primary sclerosing cholangitis. *Mayo Clin Proc*. 2000;75(7):688-694.
23. de Vries EM, Wang J, Williamson KD, Leeflang MM, Boonstra K, Weersma RK, et al. A novel prognostic model for transplant-free survival in primary sclerosing cholangitis. *Gut*. 2018;67(10):1864-1869.
24. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317-1325.

25. Mjelle AB, Fosssdal G, Gilja OH, Vesterhus M. Liver Elastography in Primary Sclerosing Cholangitis Patients Using Three Different Scanner Systems. *Ultrasound Med Biol.* 2020;46(8):1854-1864.
26. Dietrich CF, Bamber J, Berzigotti A, Bota S, Cantisani V, Castera L, et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Liver Ultrasound Elastography, Update 2017 (Long Version). *Ultraschall Med.* 2017;38(04):e16-e47.
27. Hoffman L. *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change.* Taylor & Francis; 2015.
28. Muir AJ, Levy C, Janssen HLA, Montano-Loza AJ, Shiffman ML, Caldwell S, et al. Simtuzumab for Primary Sclerosing Cholangitis: Phase 2 Study Results With Insights on the Natural History of the Disease. *Hepatology.* 2019;69(2):684-698.
29. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155-163.
30. Corpechot C, El Naggar A, Poujol-Robert A, Ziol M, Wendum D, Chazouillères O, et al. Assessment of biliary fibrosis by transient elastography in patients with PBC and PSC. *Hepatology.* 2006;43(5):1118-1124.
31. Goertz RS, GaBmann L, Strobel D, Wildner D, Schellhaas B, Neurath MF, et al. Acoustic Radiation Force Impulse (ARFI) Elastography in Autoimmune and Cholestatic Liver Diseases. *Ann Hepatol.* 2019;18(1):23-29.
32. Fraquelli M, Baccarin A, Casazza G, Conti CB, Giunta M, Massironi S, et al. Liver stiffness measurement reliability and main determinants of point shear-wave elastography in patients with chronic liver disease. *Aliment Pharmacol Ther.* 2016;44(4):356-365.
33. Zayed N, Darweesh SK, Mousa S, Atef M, Ramzy E, Yosry A. Liver stiffness measurement by acoustic radiation forced impulse and transient elastography in patients with intrahepatic cholestasis. *Eur J Gastroenterol Hepatol.* 2019;31(4):520-527.

34. Lefebvre T, Wartelle-Bladou C, Wong P, Sebastiani G, Giard JM, Castel H, et al. Prospective comparison of transient, point shear wave, and magnetic resonance elastography for staging liver fibrosis. *Eur Radiol.* 2019;29(12):6477-6488.
35. Ferraioli G, Tinelli C, Zicchetti M, Above E, Poma G, Di Gregorio M, et al. Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity. *Eur J Radiol.* 2012;81(11):3102-3106.
36. Hudson JM, Milot L, Parry C, Williams R, Burns PN. Inter- and intra-operator reliability and repeatability of shear wave elastography in the liver: a study in healthy volunteers. *Ultrasound Med Biol.* 2013;39(6):950-955.
37. Vignier N, Esmat G, Elsharkawy A, Hassany M, Bonnard P, Delarocque-Astagneau E, et al. Reproducibility of liver stiffness measurements in hepatitis C virus (HCV)-infected patients in Egypt. *J Viral Hepat.* 2011;18(7):e358-365.
38. Millonig G, Reimann FM, Friedrich S, Fonouni H, Mehrabi A, Buchler MW, et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. *Hepatology.* 2008;48(5):1718-1723.
39. Yashima Y, Tsujino T, Masuzaki R, Nakai Y, Hirano K, Tateishi R, et al. Increased liver elasticity in patients with biliary obstruction. *J Gastroenterol.* 2011;46(1):86-91.
40. Bakhshi Z, Hilscher MB, Gores GJ, Harmsen WS, Viehman JK, LaRusso NF, et al. An update on primary sclerosing cholangitis epidemiology, outcomes and quantification of alkaline phosphatase variability in a population-based cohort. *J Gastroenterol.* 2020;55(5):523-532.
41. Lindstrom L, Hultcrantz R, Boberg KM, Friis-Liby I, Bergquist A. Association between reduced levels of alkaline phosphatase and survival times of patients with primary sclerosing cholangitis. *Clin Gastroenterol Hepatol.* 2013;11(7):841-846.

42. Lindström L, Boberg KM, Wikman O, Friis-Liby I, Hultcrantz R, Prytz H, et al. High dose ursodeoxycholic acid in primary sclerosing cholangitis does not prevent colorectal neoplasia. *Aliment Pharmacol Ther.* 2012;35(4):451-457.
43. Lindström L, Hultcrantz R, Boberg KM, Friis-Liby I, Bergquist A. Association between reduced levels of alkaline phosphatase and survival times of patients with primary sclerosing cholangitis. *Clin Gastroenterol Hepatol.* 2013;11(7):841-846.
44. Olsson R, Boberg KM, de Muckadell OS, Lindgren S, Hultcrantz R, Folvik G, et al. High-dose ursodeoxycholic acid in primary sclerosing cholangitis: a 5-year multicenter, randomized, controlled study. *Gastroenterology.* 2005;129(5):1464-1472.

Table 1. Baseline characteristics of the primary sclerosing cholangitis cohorts. Reference values for laboratory parameters are equal for men and women and across study centers unless otherwise specified.

Demographics and clinical description	Total	Bergen	Oslo	Reference values	p-value
Age at study start, \bar{x} (SD)	43.3 (15.7)	44.6 (16.0)	40.1 (14.6)		0.209
Age at diagnosis, \bar{x} (SD)	35.3 (14.8)	37.0 (15.1)	31.0 (13.0)		0.045
Males, n (%)	86 (76.1)	58 (71.6)	28 (87.5)		<0.001
PSC duration in years, M (IQR)	4.0 (11)	3.0 (13)	7.0 (9)		0.093
Mayo risk score, \bar{x} (SD)	-0.5 (0.9)	-0.5 (0.9)	-0.4 (1.0)		0.430
FIB-4 score, M (IQR)	1.1 (1.2)	1.2 (1.5)	0.9 (0.9)		0.808
Decompensated liver disease, n	2	1	1		0.251
Any inflammatory bowel disease, n (%)	85 (75.2)	62 (76.5)	23 (71.9)		0.627
Ulcerative colitis, n (%)	64 (56.6)	45 (55.6)	23 (71.9)		
Crohn's disease, n (%)	12 (10.6)	10 (12.3)	2 (6.3)		
Indeterminate, n (%)	8 (7.1)	6 (7.4)	2 (6.3)		
UDCA treatment at any time, n (%)	39 (34.5)	25 (22.1)	14 (12.4)		<0.001
Patients with endoscopic intervention, n (%)	6 (5.3)	3 (3.7)	3 (9.3)		0.362
Prognostic biomarkers					
Participants above cut-off values					
ALP ^a , n (%)	52 (46%)	36 (44.4%)	16 (50%)		0.362
ELF ^b , n (%)	37 (32.7%)	22 (33.3%)	10 (31.3%)		0.428
LSM ^c , n (%)	50 (45%)	37 (45.7%)	13 (43.4%)		0.098
Levels, M (IQR)					
ALP (U/L)	151.5 (197)	149.0 (196)	165.0 (206)	35-105	0.871
ALP by ULN, M (range)	1.4 (0.4, 8.0)	1.4 (0.4, 8.0)	1.5 (0.5, 6.1)		
ELF	9.3 (1.34)	9.3 (1.32)	9.4 (1.45)		0.905
LSM (m/s)	1.26 (0.52)	1.26 (0.48)	1.17 (1.21)		0.373
Other blood tests, M (IQR)					
ALT (U/L)	53.0 (81)	52.0 (66)	74.0 (127)	10-70 (m) 10-45 (f)	0.241
AST (U/L)	48.0 (49)	47.0 (48)	51.5 (75)	15-45 (m) 15-35 (f)	0.633
GGT (U/L)	228.0 (597)	149.0 (565)	238.5 (753)	10-80 (m<40 y) ^d 10-45 (f<40 y) ^d	0.856
Bilirubin ($\mu\text{mol/L}$)	11.0 (10)	11.0 (9)	12.5 (16)	5-25 ^e	0.048
Thrombocytes ($\times 10^9$)	245.0 (105)	240.0 (102)	240.0 (111)	145-390 ^g	0.779
Albumin (g/L)	45.0 (5)	46.0 (5)	44.0 (5)	36-48 (<40 y) ^h	0.122

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ELF, enhanced liver fibrosis; f, females; GGT, Gamma-glutamyl transferase; IQR, interquartile range; LSM, Liver stiffness measurement, M, median; m, males; ULN, upper limit of normal; y, patient's years of age.

^a $\geq 1.5 \times \text{ULN}$, ^b ≥ 9.8 , ^c $\geq 1.28 \text{ m/s}$

^d GGT 15-115 U/L for males ≥ 40 years of age, 10-75 U/L for females ≥ 40 years of age.

^e Bilirubin $\leq 21 \mu\text{mol/L}$. ^f Thrombocytes $145\text{-}348 \times 10^9$ (m), $165\text{-}387 \times 10^9$ (f).

^g Albumin 39-50 g/L for patients <40 , 39-48 g/L for patients between 40-69, 36-48 g/L for patients ≥ 70 years of age in the Bergen cohort.

^h Albumin 36-45 g/L for patients 40-69, 34-45 g/L for patients ≥ 70 years of age in the Oslo cohort.

Table 2. Liver stiffness measures and liver parameters over time.

		Effect size	95% CI	p-value
ELF	Fixed intercept ^a	-0.11	[-0.29, -0.06]	.196
	Fixed slope ^b	0.06	[0.02, 0.09]	.005*
	Crude ICC ^c	0.78	[0.72, 0.83]	
	Adjusted ICC ^d	0.83	[0.77, 0.87]	
LSM	Fixed intercept ^a	-0.11	[-0.27, 0.06]	.199
	Fixed slope ^b	0.07	[0.02, 0.13]	.009*
	Crude ICC ^c	0.56	[0.47, 0.65]	
	Adjusted ICC ^d	0.59	[0.48, 0.70]	
ALP	Fixed intercept ^a	-0.03	[-0.21, 0.16]	.775
	Fixed slope ^b	0.04	[0.01, 0.07]	.011*
	Crude ICC ^c	0.86	[0.82, 0.89]	
	Adjusted ICC ^d	0.89	[0.85, 0.92]	
Bilirubin	Fixed intercept ^a	-0.09	[-0.26, -0.09]	.325
	Fixed slope ^b	0.07	[0.02, 0.12]	.007*
	Crude ICC ^c	0.64	[0.55, 0.72]	
	Adjusted ICC ^d	0.71	[0.62, 0.78]	

Abbreviations: ALP, alkaline phosphatase; ELF, enhanced liver fibrosis; LSM, liver stiffness measurement;

^a The fixed effect at baseline. All variables have been log-transformed and z-scored so that the mean represents the grand mean over 5-years. A negative fixed intercept indicates how much lower the variable is at baseline compared to the grand mean, in standard deviations.

^b The fixed slope indicates change in the outcome in standard deviations per year.

^c The ICC from an empty-means random intercept model.

^d The ICC from a random slope model adjusted for time-in-study.

* p-value < 0.05

Table 3. Associations of ELF and liver stiffness measurement with biochemical markers and clinical scores in a linear mixed-effects model^a.

Predictor	Outcome	sFE ^b	95% CI	P-value
ALP	ELF	0.47	[0.37, 0.56]	< 0.001
	LSM	0.28	[0.16, 0.39]	< 0.001
Albumin ^c	ELF	-0.39	[-0.47, -0.32]	< 0.001
	LSM	-0.35	[-0.44, -0.25]	< 0.001
Bilirubin	ELF	0.20	[0.11, 0.29]	< 0.001
	LSM	0.29	[0.18, 0.39]	< 0.001
Mayo risk score ^c	ELF	0.48	[0.40, 0.56]	< 0.001
	LSM	0.37	[0.26, 0.47]	< 0.001
FIB-4	ELF	0.56	[0.46, 0.65]	< 0.001
	LSM	0.42	[0.31, 0.53]	< 0.001

Abbreviations: ALP, alkaline phosphatase BR, bilirubin; ELF, enhanced liver fibrosis; LSM, liver stiffness measurement; FE, fixed effects, SD, standard deviation; sFE, standardized fixed effects.

^a Linear mixed-effects models as described under statistics.

^b Standardized fixed effect calculated as $sFE = (FE \times SD \text{ predictor variable}) / SD \text{ dependent variable}$.

^c Not log-transformed (all other log-transformed)

Table 4. Decomposition of longitudinal associations of ELF and LSM with liver biochemistries in PSC^a.

	Individual means		Linear change		Fluctuation	
	(Random intercepts)		(Random slopes)		(Residuals)	
ELF as the outcome						
	sFE (95% CI)	p	sFE (95% CI)	p	sFE (95% CI)	p
ALP	0.37 (0.21, 0.52)	<.001**	0.03 (-0.14, 0.19)	.768	0.15 (0.11, 0.18)	<.001**
Bilirubin	0.40 (0.26, 0.54)	<.001**	0.16 (-0.01, 0.31)	.052	0.03 (-0.01, 0.08)	.161
LSM as the outcome						
ALP	0.32 (0.18, 0.46)	<.001**	0.07 (-0.08, 0.21)	.384	0.05 (-0.01, 0.11)	.091
Bilirubin	0.42 (0.30, 0.54)	<.001**	0.23 (0.10, 0.35)	<.001**	0.03 (-0.04, 0.10)	.407

Abbreviations: ALP, alkaline phosphatase; ELF, enhanced liver fibrosis; LSM, liver stiffness measurement; PSC, primary sclerosing cholangitis; sFE, standardized fixed effects.

^a A two-step multilevel model where first the random intercepts, slopes, and residuals for the predictors ALP and bilirubin were estimated from separate models with time as the predictor. These now represent differences in individual means, individual linear rate of change, and the residuals represent fluctuating deviations from these. These were entered as predictors in a second multilevel model, with ELF or LSM as the outcome and time as the only covariate.

Figure legends

FIGURE 1 Correlation network for ELF, LSM, and relevant biochemistries. The strength of correlations is indicated by the widths of the connecting lines. Positive and negative correlations are represented by green and red color, respectively. The diagram highlights liver enzymes ALT, AST, ALP, and GT as a group with high correlation. ELF and LSM were most strongly correlated with each other and showed correlations with liver enzymes and negative correlations with albumin and platelets. ALB, albumin; ALT, alanine aminotransferase; ALP, alkaline phosphatase; AST, aspartate aminotransferase; BIL, bilirubin; DUR, PSC duration; ELF, enhanced liver fibrosis; GT, gamma-glutamyl transferase; LSM, liver stiffness measurement; MAY, Mayo risk score; PLT, platelets.

FIGURE 2 Development of ELF, LSM, ALP, and bilirubin over time in patients with PSC (n=113). Boxplot; whiskers represent the 1st and 3rd quartiles, respectively. Each box is represented by the number of measurements for each parameter per year in study. When applying a longitudinal mixed model analysis considering all available repeated measurements, there was a small but significant increase in ELF and LSM over time ($p = 0.005$ and 0.009 , respectively). ELF, enhanced liver fibrosis test; LSM, liver stiffness measurement; ALP, alkaline phosphatase.

FIGURE 3 Linear mixed model analysis of the longitudinal development of ELF and LSM in high and low-risk groups defined by ALP. The high-risk subgroup ($ALP \geq 1.5 \times ULN$ at baseline) showed significantly higher baseline ELF ($p = 0.001$) compared to the low-risk group, with a similar trend for LSM ($p = 0.06$). Both ELF and LSM increased significantly over time in the high-ALP group ($p = 0.014$ and 0.022 , respectively), whereas there was no

significant increase for ELF or LSM in the low-ALP group. For ELF, there was a trend towards interaction between ALP-defined risk group and time which did not reach significance ($p > 0.05$), whereas for LSM, there was no interaction between risk group and time ($p > 0.50$).

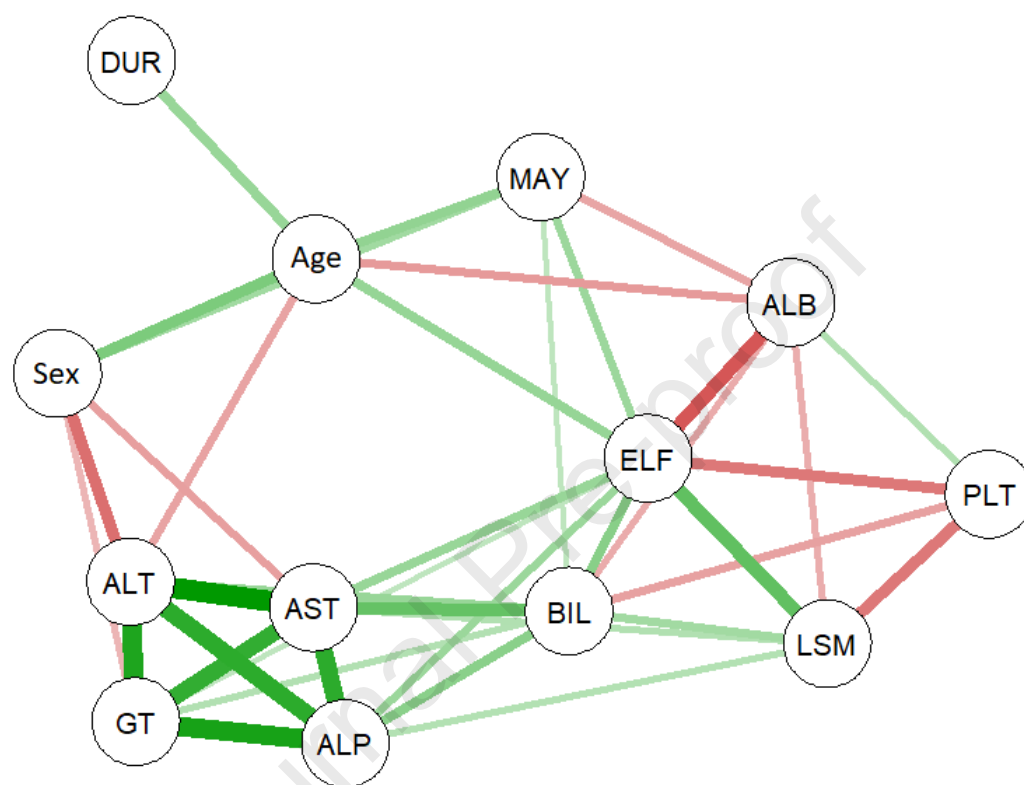
Figures**Figure 1**

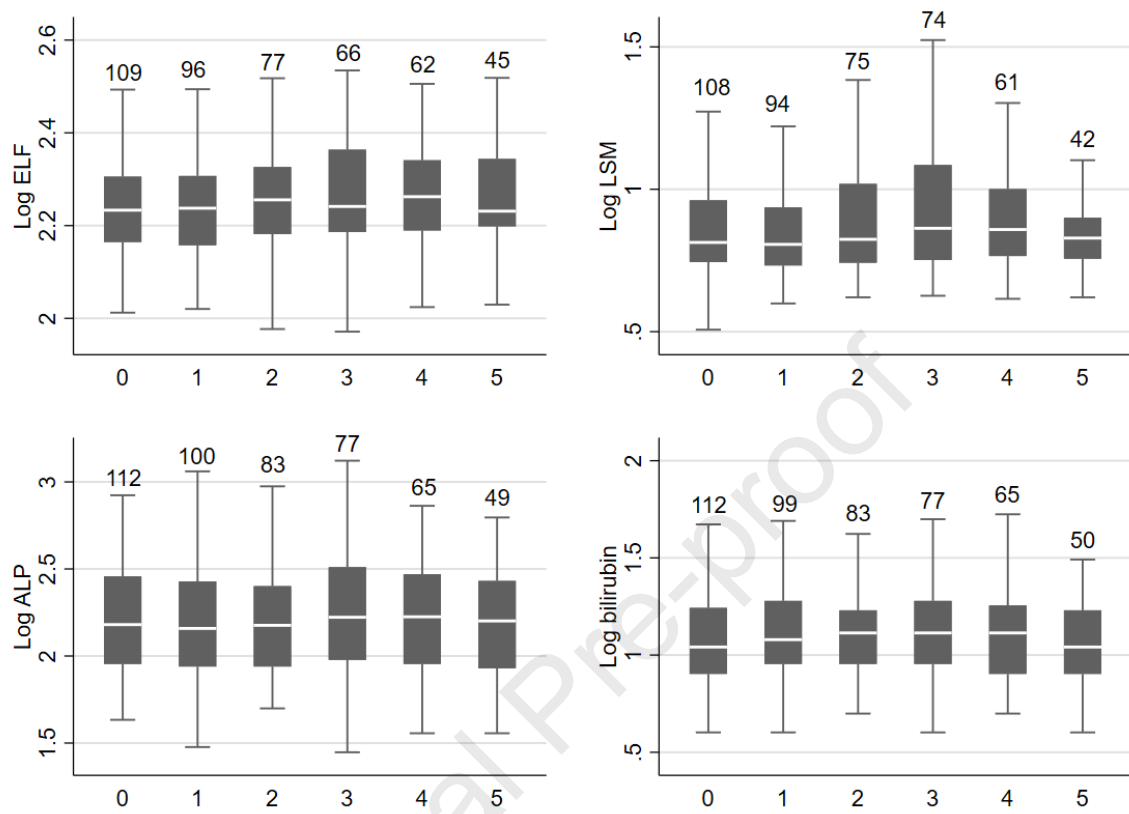
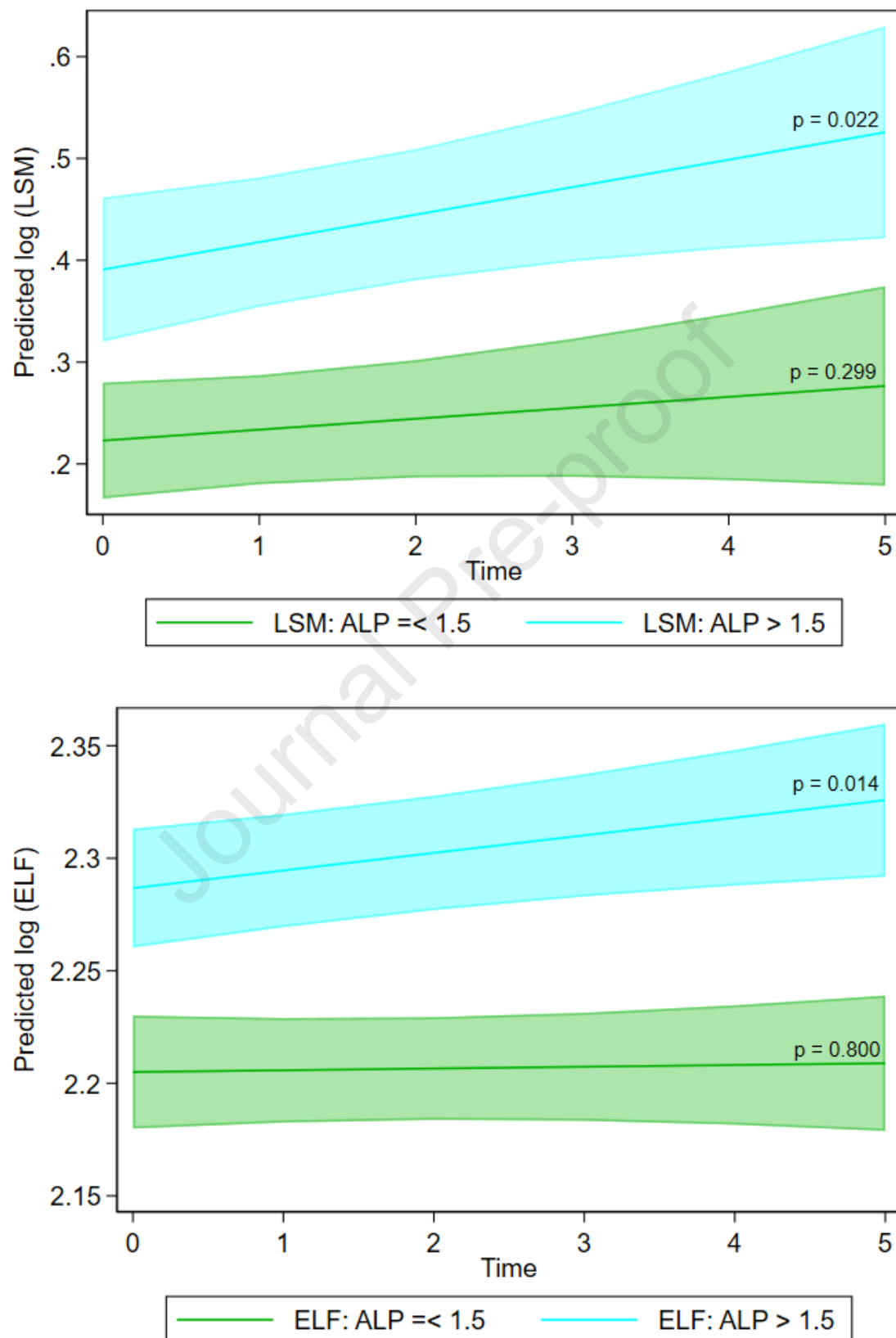
Figure 2

Figure 3



Highlights

- ELF and LSM increased in PSC patients, but only in patients with $ALP > 1.5 \times ULN$
- ELF may be more reliable for PSC risk stratification (low within-patient variation)
- A subgroup showed concomitant spontaneous reduction in ALP, ELF, and LSM