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


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


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\textsuperscript{1,2,3}, Anders Batman Mjelle\textsuperscript{4}, Kristine Wiencke\textsuperscript{1,5,6}, Ida Bjørk\textsuperscript{7}, Odd Helge Gilja\textsuperscript{4,8}, Trine Folseraas\textsuperscript{1,5,6}, Tom Hemming Karlsen\textsuperscript{1,5,6,9}, William Rosenberg\textsuperscript{10}, Lasse Melvær Giil\textsuperscript{3}, Mette Vesterhus\textsuperscript{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
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 disease. The natural course of PSC is highly variable, with median transplant-free survival ranging from 13 to 20 years. 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 studies. 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. 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. In contrast, LSM assesses the physical, viscoelastic properties of the liver using ultrasound-based elastography methods. Both the ELF test and LSM have been shown to predict transplant-free survival in PSC across independent studies. 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 criteria. 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 criteria. Clinical and demographic information, including laboratory data, was acquired from patient records and research databases. Liver biochemistry, ELF test, and elastography were sampled annually (± 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 algorithms.

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: ELF test = 2.278 + 0.851 ln (C_{HA}) + 0.751 ln (C_{PIIINP}) + 0.394 ln (C_{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 examination. 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 analyses. Liver stiffness is expressed as shear wave speed (m/s) or converted into Young’s modulus using the equation kPa = 3[(ms⁻¹)²]. 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 9
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). 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. For further subgroup analyses, the cohort was divided according to presumed high-risk profiles at baseline, i.e., ALP ≥1.5x (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 10

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 (rho 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 11
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 ≥1.5x 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 ≥1.5x ULN accounted for all of the patients with ≥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 ≥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 ≥1.5x 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. 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 variation\(^{29}\), 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 variation\(^{12}\). 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
PSC\textsuperscript{17-19}, and a strong predictive ability for clinical outcomes in independent studies\textsuperscript{17,18}. Moreover, the elastography modalities we used (pSWE and ARFI quantification) were reported to correlate well with histology\textsuperscript{19,31-33}, demonstrated high accuracy in discriminating between lower and higher degrees of fibrosis\textsuperscript{32-34}, and excellent correlation to TE in PSC patients\textsuperscript{25}. 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\textsuperscript{25,35-37}. Furthermore, the patchy disease distribution in PSC and variation in cholestasis may contribute to variations in LSM\textsuperscript{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 $\geq 1.5 \times$ 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\textsuperscript{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. 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 fibrosis. 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 studies. 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≥1.5xULN, 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


Prospective comparison of transient, point shear wave, and magnetic resonance elastography for staging liver fibrosis. *Eur Radiol.* 2019;29(12):6477-6488.


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

<table>
<thead>
<tr>
<th>Demographics and clinical description</th>
<th>Total</th>
<th>Bergen</th>
<th>Oslo</th>
<th>Reference values</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics and clinical description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>43.3</td>
<td>44.6</td>
<td>40.1</td>
<td></td>
<td>0.209</td>
</tr>
<tr>
<td>Age at study start, x (SD)</td>
<td>(15.7)</td>
<td>(16.0)</td>
<td>(14.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at diagnosis, x (SD)</td>
<td>(14.8)</td>
<td>(15.1)</td>
<td>(13.0)</td>
<td></td>
<td>0.045</td>
</tr>
<tr>
<td>Males, n (%)</td>
<td>86 (76.1)</td>
<td>58 (71.6)</td>
<td>28 (87.5)</td>
<td></td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>PSC duration in years, M (IQR)</td>
<td>4.0 (11)</td>
<td>3.0 (13)</td>
<td>7.0 (9)</td>
<td></td>
<td>0.093</td>
</tr>
<tr>
<td>Mayo risk score, x (SD)</td>
<td>-0.5 (0.9)</td>
<td>-0.5 (0.9)</td>
<td>-0.4 (1.0)</td>
<td></td>
<td>0.430</td>
</tr>
<tr>
<td>FIB-4 score, M (IQR)</td>
<td>1.1 (1.2)</td>
<td>1.2 (1.5)</td>
<td>0.9 (0.9)</td>
<td></td>
<td>0.808</td>
</tr>
<tr>
<td>Decompensated liver disease, n (%)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0.251</td>
</tr>
<tr>
<td>Any inflammatory bowel disease, n (%)</td>
<td>85 (75.2)</td>
<td>62 (76.5)</td>
<td>23 (71.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulcerative colitis, n (%)</td>
<td>64 (56.6)</td>
<td>45 (55.6)</td>
<td>23 (71.9)</td>
<td></td>
<td>0.627</td>
</tr>
<tr>
<td>Crohn’s disease, n (%)</td>
<td>12 (10.6)</td>
<td>10 (12.3)</td>
<td>2 (6.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indeterminate, n (%)</td>
<td>8 (7.1)</td>
<td>6 (7.4)</td>
<td>2 (6.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDCA treatment at any time, n (%)</td>
<td>39 (34.5)</td>
<td>25 (22.1)</td>
<td>14 (12.4)</td>
<td></td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>Patients with endoscopic intervention, n (%)</td>
<td>6 (5.3)</td>
<td>3 (3.7)</td>
<td>3 (9.3)</td>
<td></td>
<td>0.362</td>
</tr>
<tr>
<td>Prognostic biomarkers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants above cut-off values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALP a, n (%)</td>
<td>52 (46%)</td>
<td>(44.4%)</td>
<td>16 (50%)</td>
<td></td>
<td>0.362</td>
</tr>
<tr>
<td>ELF b, n (%)</td>
<td>37</td>
<td>22</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM c, n (%)</td>
<td>50 (45%)</td>
<td>(45.7%)</td>
<td>13</td>
<td></td>
<td>0.098</td>
</tr>
<tr>
<td>Levels, M (IQR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALP (U/L)</td>
<td>151.5</td>
<td>149.0</td>
<td>165.0</td>
<td>35-105</td>
<td>0.871</td>
</tr>
<tr>
<td>ALP by ULN, M (range)</td>
<td>1.4 (0.4, 8.0)</td>
<td>1.4 (0.4, 8.0)</td>
<td>1.5 (0.5, 6.1)</td>
<td>905</td>
<td></td>
</tr>
<tr>
<td>ELF</td>
<td>9.3 (1.34)</td>
<td>9.3 (1.32)</td>
<td>9.4 (1.45)</td>
<td>0.905</td>
<td></td>
</tr>
<tr>
<td>LSM (m/s)</td>
<td>1.26</td>
<td>1.26</td>
<td>1.17</td>
<td></td>
<td>0.373</td>
</tr>
<tr>
<td>Other blood tests, M (IQR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALT (U/L)</td>
<td>53.0 (81)</td>
<td>52.0 (66)</td>
<td>74.0 (127)</td>
<td></td>
<td>0.241</td>
</tr>
<tr>
<td>AST (U/L)</td>
<td>48.0 (49)</td>
<td>47.0 (48)</td>
<td>51.5 (75)</td>
<td></td>
<td>0.633</td>
</tr>
<tr>
<td>GGT (U/L)</td>
<td>228.0 (597)</td>
<td>149.0 (565)</td>
<td>238.5 (753)</td>
<td>10-70 (m)</td>
<td>0.856</td>
</tr>
<tr>
<td>Bilirubin (µmol/L)</td>
<td>11.0 (10)</td>
<td>11.0 (9)</td>
<td>12.5 (16)</td>
<td>10-45 (f)</td>
<td>0.048</td>
</tr>
<tr>
<td>Thrombocytes (x10^9)</td>
<td>245.0</td>
<td>240.0</td>
<td>240.0</td>
<td>5-25 e</td>
<td>0.779</td>
</tr>
<tr>
<td>Albumin (g/L)</td>
<td>45.0 (5)</td>
<td>46.0 (5)</td>
<td>44.0 (5)</td>
<td>36-48 (&lt;40 y) g</td>
<td>0.122</td>
</tr>
</tbody>
</table>
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.

\[a \geq 1.5 \times \text{ULN}, \quad b \geq 9.8, \quad c \geq 1.28 \text{ m/s}\]

\[d \quad \text{GGT 15-115 U/L for males \(\geq\) 40 years of age, 10-75 U/L for females \(\geq\) 40 years of age.}\]

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

\[g \quad \text{Albumin} 39-50 \text{ g/L for patients <40, 39-48 g/L for patients between 40-69, 36-48 g/L for patients \(\geq\) 70 years of age in the Bergen cohort.}\]

\[h \quad \text{Albumin} 36-45 \text{ g/L for patients 40-69, 34-45 g/L for patients \(\geq\) 70 years of age in the Oslo cohort.}\]
<table>
<thead>
<tr>
<th></th>
<th>Effect size</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELF</td>
<td>Fixed intercept</td>
<td>-0.11</td>
<td>[-0.29, -0.06]</td>
</tr>
<tr>
<td></td>
<td>Fixed slope</td>
<td>0.06</td>
<td>[0.02, 0.09]</td>
</tr>
<tr>
<td></td>
<td>Crude ICC</td>
<td>0.78</td>
<td>[0.72, 0.83]</td>
</tr>
<tr>
<td></td>
<td>Adjusted ICC</td>
<td>0.83</td>
<td>[0.77, 0.87]</td>
</tr>
<tr>
<td>LSM</td>
<td>Fixed intercept</td>
<td>-0.11</td>
<td>[-0.27, 0.06]</td>
</tr>
<tr>
<td></td>
<td>Fixed slope</td>
<td>0.07</td>
<td>[0.02, 0.13]</td>
</tr>
<tr>
<td></td>
<td>Crude ICC</td>
<td>0.56</td>
<td>[0.47, 0.65]</td>
</tr>
<tr>
<td></td>
<td>Adjusted ICC</td>
<td>0.59</td>
<td>[0.48, 0.70]</td>
</tr>
<tr>
<td>ALP</td>
<td>Fixed intercept</td>
<td>-0.03</td>
<td>[-0.21, 0.16]</td>
</tr>
<tr>
<td></td>
<td>Fixed slope</td>
<td>0.04</td>
<td>[0.01, 0.07]</td>
</tr>
<tr>
<td></td>
<td>Crude ICC</td>
<td>0.86</td>
<td>[0.82, 0.89]</td>
</tr>
<tr>
<td></td>
<td>Adjusted ICC</td>
<td>0.89</td>
<td>[0.85, 0.92]</td>
</tr>
<tr>
<td>Bilirubin</td>
<td>Fixed intercept</td>
<td>-0.09</td>
<td>[-0.26, -0.09]</td>
</tr>
<tr>
<td></td>
<td>Fixed slope</td>
<td>0.07</td>
<td>[0.02, 0.12]</td>
</tr>
<tr>
<td></td>
<td>Crude ICC</td>
<td>0.64</td>
<td>[0.55, 0.72]</td>
</tr>
<tr>
<td></td>
<td>Adjusted ICC</td>
<td>0.71</td>
<td>[0.62, 0.78]</td>
</tr>
</tbody>
</table>

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\(^a\).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Outcome</th>
<th>sFE(^b)</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP</td>
<td>ELF</td>
<td>0.47</td>
<td>[0.37, 0.56]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>LSM</td>
<td>0.28</td>
<td>[0.16, 0.39]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Albumin(^c)</td>
<td>ELF</td>
<td>-0.39</td>
<td>[-0.47, -0.32]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>LSM</td>
<td>-0.35</td>
<td>[-0.44, -0.25]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Bilirubin</td>
<td>ELF</td>
<td>0.20</td>
<td>[0.11, 0.29]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>LSM</td>
<td>0.29</td>
<td>[0.18, 0.39]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Mayo risk score(^d)</td>
<td>ELF</td>
<td>0.48</td>
<td>[0.40, 0.56]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>LSM</td>
<td>0.37</td>
<td>[0.26, 0.47]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>FIB-4</td>
<td>ELF</td>
<td>0.56</td>
<td>[0.46, 0.65]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>LSM</td>
<td>0.42</td>
<td>[0.31, 0.53]</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

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 x SD predictor variable) / SD 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.

<table>
<thead>
<tr>
<th></th>
<th>Individual means</th>
<th>Linear change</th>
<th>Fluctuation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Random intercepts)</td>
<td>(Random slopes)</td>
<td>(Residuals)</td>
</tr>
<tr>
<td><strong>ELF as the outcome</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALP</td>
<td>sFE (95% CI)</td>
<td>p</td>
<td>sFE (95% CI)</td>
</tr>
<tr>
<td></td>
<td>0.37 (0.21, 0.52)</td>
<td>&lt;.001**</td>
<td>0.03 (-0.14, 0.19)</td>
</tr>
<tr>
<td>Bilirubin</td>
<td>0.40 (0.26, 0.54)</td>
<td>&lt;.001**</td>
<td>0.16 (-0.01, 0.31)</td>
</tr>
<tr>
<td><strong>LSM as the outcome</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALP</td>
<td>sFE (95% CI)</td>
<td>p</td>
<td>sFE (95% CI)</td>
</tr>
<tr>
<td></td>
<td>0.32 (0.18, 0.46)</td>
<td>&lt;.001**</td>
<td>0.07 (-0.08, 0.21)</td>
</tr>
<tr>
<td>Bilirubin</td>
<td>0.42 (0.30, 0.54)</td>
<td>&lt;.001**</td>
<td>0.23 (0.10, &lt;.001**)</td>
</tr>
</tbody>
</table>

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

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 ≥1.5xULN 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

![Graph showing predicted log (LSM) and predicted log (ELF) over time for different categories of ALP](image-url)
Highlights

- ELF and LSM increased in PSC patients, but only in patients with ALP>1.5*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