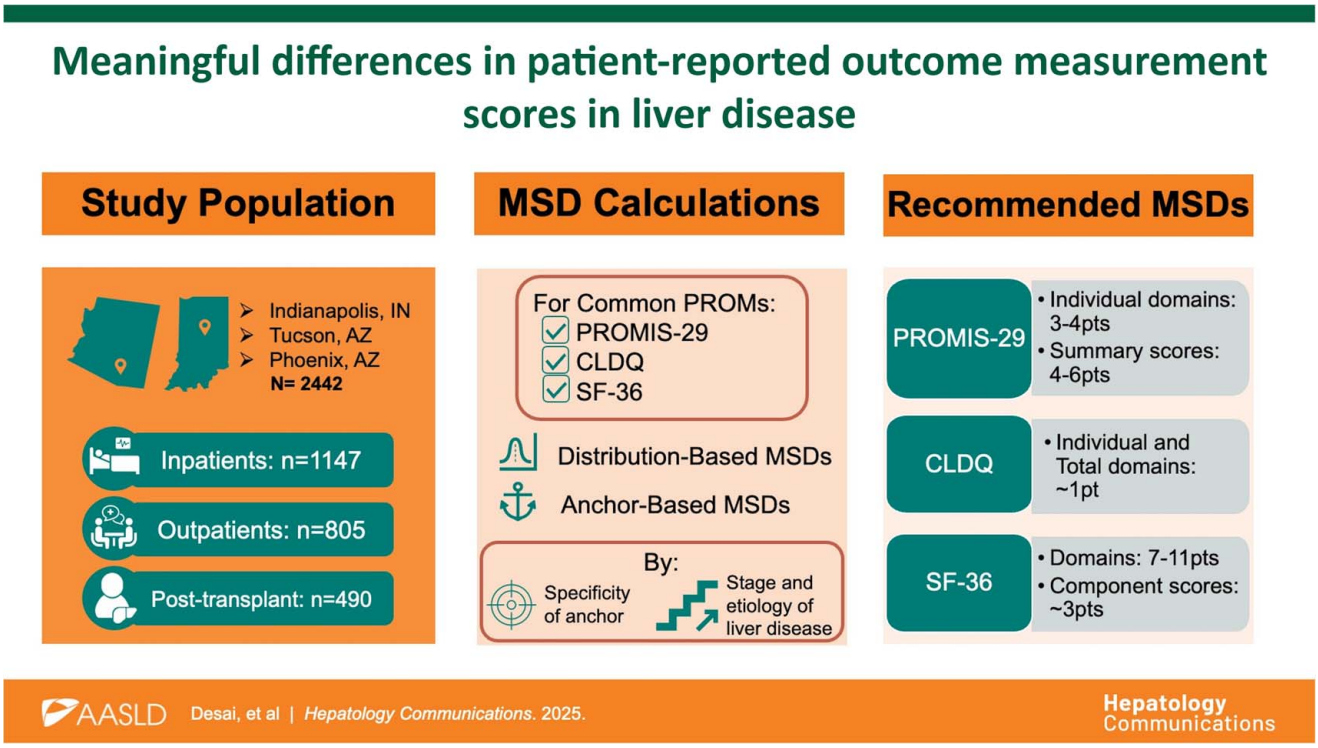
Meaningful differences in patient-reported outcome measurement scores in liver disease

VISUAL ABSTRACT



Received: 2 January 2025 | Accepted: 11 April 2025

DOI: 10.1097/HC9.0000000000000727

ORIGINAL ARTICLE

OPEN

Meaningful differences in patient-reported outcome measurement scores in liver disease

Archita P. Desai[1] | **Eric S. Orman**[1] | **Tarek G. Aridi**[2] | **Timothy Stump**[3] | **Lauren Nephew**[1] | **Marwan S. Ghabril**[1] | **Michael Fallon**[4] | **Naga Chalasani**[1] | **Patrick Monahan**[3]

[1]Division of Gastroenterology and Hepatology, Indiana University School of Medicine, Indianapolis, Indiana, USA

[2]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA

[3]Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, Indiana, USA

[4]Division of Gastroenterology and Hepatology, University of Arizona College of Medicine—Phoenix, Phoenix, Arizona, USA

Correspondence
Archita P. Desai, Division of Gastroenterology and Hepatology, Indiana University School of Medicine, 702 Rotary Building, Suite 225, Indianapolis, IN 46202, USA.
Email: desaiar@iu.edu

Abstract

Background: Patient-reported outcome measures (PROMs) are being used more often in chronic liver disease (CLD) clinical care and research. Their interpretability can be greatly enhanced by establishing the smallest meaningful score difference (MSD). We report scores of commonly used PROMs and their MSDs in patients at different stages of liver disease.

Methods: Patient-Reported Outcomes Measurement Information System (PROMIS)-29 Profile, Chronic Liver Disease Questionnaire (CLDQ), and Short Form-36 (SF-36) v1.0 scores were aggregated from 2442 adults with CLD at 4 different stages: inpatients with decompensated cirrhosis (n = 1146) and outpatients with cirrhosis (n = 677) or CLD (n = 128) or recipients of liver transplant (LT, n = 490) between June 2014 and April 2023 from 3 academic centers. MSDs were estimated using distribution and anchor-based methods.

Results: The study sample's median age was 60.0 (IQR: 51.0–66.0); 55% were male, 17% Hispanic, 84% White, and 49% college educated. The etiology of CLD was alcohol in 36%, metabolic dysfunction–associated steatohepatitis (MASH) in 31%, and viral hepatitis B/C in 26%. Median PROMIS domain scores were generally lowest in inpatients and highest after transplant. For PROMIS, distribution-based and anchor-based MSDs ranged from 3 to 4 for individual domains and 4 to 6 for summary scores. Distribution-based MSDs were 1 for CLDQ and ranged from 7 to 11 for individual SF-36 domains, except role limitations domains, which ranged from 15 to 18, and component scores, which were 3. When compared across stages of liver

Abbreviations: APSRA, ability to participate in social roles and activities v2.0; CLD, chronic liver disease; CLDQ, Chronic Liver Disease Questionnaire; HRQOL, health-related quality of life; LT, liver transplantation; MASH, metabolic dysfunction–associated steatohepatitis; MSD, meaningful score difference; PROMIS, Patient-Reported Outcomes Measurement Information System; PROM, patient-reported outcome measure; PROs, patient-reported outcomes; SF-36, Short Form-36; SPDSA, satisfaction with participation in discretionary social activities v1.0.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.hepcommjournal.com.

disease, PROMIS MSDs were generally similar, although they tended to be 0.5–1.0 points smaller in the decompensated population compared to the stable populations.

Conclusions: This study provides data-driven recommendations for MSDs, enhancing the interpretability of commonly used PROMs in liver disease and facilitating the integration of PROMs in various clinical and research settings.

Keywords: chronic liver disease, health-related quality of life, meaningful score difference, minimal important difference, patient-reported outcome measurement

INTRODUCTION

Measuring and interpreting patient-reported outcomes (PROs) add value to healthcare and align with the aim of practicing patient-centered care established by the Institute of Medicine.[1–3] PROs are often collected using validated measures (patient-reported outcome measures, PROMs), which are poised to improve patient reporting of symptoms and needs, achieve patient-centered care, and improve clinical outcomes.[4–7] Transition to electronic medical records and advances in digital health technology have allowed for the ability of electronic PROM completion by patients as part of their routine clinical care.[3,7,8] Despite these advances in PROM, barriers to using PROs in routine practice remain.[7–10]

An important barrier to PROMs integration into clinical research and practice is that they generate scores that require assessment before facilitating the interpretation and action.[7,8,11–14] In liver disease, for example, health-related quality of life (HRQOL) is a commonly studied PRO and can be measured by PROMs such as NIH's Patient-Reported Outcomes Measurement Information System (PROMIS-29), Chronic Liver Disease Questionnaire (CLDQ), and Short Form-36 (SF-36).[15,16] These PROMs provide several domain and summary scores, which then need to be compared to healthy samples or to prior measurements to have clinical utility.[14,17–19] Establishing differences in scores that are large enough to be meaningful, the meaningful score difference (MSD or minimal clinically important difference), sets a standard for interpretation and is an important step in implementing PROMs in clinical practice and as clinical trial endpoints.[11–14,20,21] MSDs have been established in a variety of disease states for both clinical and research use and facilitated the use of PROs in these settings.[22–28]

There are no established MSDs for HRQOL domains commonly measured during research and clinical practice involving patients with chronic liver disease (CLD) outside of the hepatitis C population.[29,30] To address this need, we aimed to establish MSDs for domain and summary scores for PROMIS-29 Profile (primary aim), SF-36, and CLDQ (secondary aims) using a prospective, diverse cohort of patients at various stages of liver disease.

METHODS

Study design

PROMIS, SF-36, and CLDQ domain scores were aggregated from separate studies that recruited participants aged $\geq$18 years old with CLD with or without cirrhosis. Inpatient participants were recruited from the University of Arizona (Tucson and Phoenix) and Indiana University from June 2014 to April 2023 during admission for a complication related to cirrhosis (infection, ascites, hepatic encephalopathy, GI bleed, fluid overload). Exclusion criteria included significant encephalopathy at the time of recruitment (West Haven Stage $\geq$2), prior liver transplant, death, or hospice enrollment prior to discharge, inability to read the survey on an iPad screen, and use of language other than English and Spanish. Outpatients without a history of transplant were recruited during routine outpatient hepatology visits at the University of Arizona (Tucson and Phoenix) and Indiana University from August 2016 to August 2021. Patients were excluded if they had significant encephalopathy at the time of recruitment (West Haven Stage $\geq$2), inability to read the survey on an iPad screen, or use of language other than English and Spanish. Post-transplant participants were enrolled at Indiana University during routine outpatient transplant visits between June 2019 and February 2023. Patients were excluded if they underwent liver transplant surgery at a different institution, were unable to read the survey on a tablet, used a language other than English or Spanish, or had cognitive impairment limiting their ability to complete the survey.

The diagnosis of cirrhosis and its complications was established through chart review of imaging, clinical

history, and/or liver biopsy. For all participants, demographics, comorbidities as measured by the Charlson Comorbidity Index, and clinical characteristics, including laboratory test results, were all recorded using a secure RedCap database.

HRQOL instruments

Participants electronically completed PROMIS domain short forms. These include: global health v1.1, physical function v1.2, fatigue v1.0, pain behavior v1.0, pain interference v1.1, depression v1.0, anxiety v1.0, satisfaction with participation in discretionary social activities v1.0 (SPDSA), ability to participate in social roles and activities v2.0 (APSRA), and sleep disturbance v1.0. PROMIS domain scores are reported at *T*-scores, and summary scores (physical and mental) were calculated as previously described.[15,31] Higher *T*-score indicates more of the domain measure, and $T > 50$ is higher than the US general adult population, with an SD of 10. Therefore, a score of 55 indicates 0.5 SD more physical function, or conversely, more pain interference than a score of 50. SF-36 v1.0 and CLDQ were also completed electronically and scored and interpreted as previously described.[15]

MSD estimation

We utilized 2 common methodologies for estimating MSDs—one based on distribution of the domain scores and one based on comparison to an anchor as described below.[14,19] Distribution-based methods were used to estimate MSDs for PROMIS, SF-36 v1.0, and the CLDQ domains described above. Additionally, we estimated MSDs for PROMIS domains using the anchor-based method.[19,23,24,32] MSDs were reported using *T*-score (PROMIS domains) or the score linearly transformed to a 0–100 metric (SF-36).

Distribution-based method

The distribution-based method uses the statistical distributions of the data. For this study, we utilized the SEM, which conveys the precision of the outcome measure. The SEM is the smallest difference in a score that may be interpreted as a true difference between scores rather than a measurement error.[23,33,34] For PROMIS measures, each participant has an SE associated with that individual's *T*-score. The sample SEM for each domain was calculated using the square root of the mean of variance (ie, standard error squared) for each individual's *T*-score across persons in the sample. For PROMIS-29 Profile summary scores, as well as for SF-36 and CLDQ, where individual SEs were not available, we calculated SEM by multiplying the SD by the square root of (1 minus Cronbach alpha). We used 1 SEM as a distribution-based MSD, as this corresponds closely with anchor-based MSDs in prior literature.[32,34,35]

Anchor-based method

The anchor-based method maps differences in domains score onto differences in clinically meaningful anchors.[19,24,32] For this study, we based our MSD determination using the PROMIS Global Health questionnaire as a clinically meaningful anchor, as many of its questions measure concepts that align with the domains measured by Profile-29 (Supplemental Table S1, http://links.lww.com/HC9/C9). This tool uses 10 questions, such as "In general, would you say your health is:" most of which are answered with 1 of 5 options: excellent, very good, good, fair, or poor. The MSD for each PROMIS domain was estimated by measuring the mean difference in the domain *T*-scores in individuals answering 2 adjacent categories in each global health anchor question (mean change method). For example, we calculated the mean sleep *T*-score for those who answered "excellent" minus the mean sleep *T*-score for those who answered "very good" on the second global question, "In general, would you say your quality of life is." We then took the average of these differences between all 4 adjacent categories to estimate an overall MSD. This calculation was performed for each global health anchor question.

Selection of global health anchors was based on the strength of association between Profile-29 domains' scores and the global health item as measured by the Spearman rank-order correlation coefficient. A correlation coefficient of ≥ 0.30 was used as the threshold for an acceptable association between the anchor and the PRO.[24,33,36] As all 10 possible anchors met this threshold for each PROMIS domain (Supplemental Table S2, http://links.lww.com/HC9/C9, range of absolute value of coefficients: 0.33–0.75), we also categorized the conceptual similarity between the PROMIS domain of interest and possible anchors as universal (eg, "In general, would you say your health is?"), general (eg, "In general, how would you rate your physical health?") or specific (eg, "How would you rate your fatigue on average?") (Supplemental Table S2, http://links.lww.com/HC9/C9). MSDs for each PROMIS domain were calculated as an overall average of all relevant anchors as well as averages of each category of conceptual similarity to provide multiple possible anchor-based MSDs for triangulation.

Descriptive statistics were presented as percentages for categorical variables and means (SDs) or medians (IQRs) for continuous variables. Baseline characteristics were compared between the 2 groups using the

Pearson chi-squared test or the Fisher exact test for categorical variables, *t* test for continuous variables, and the Wilcoxon Rank-Sum test for continuous skewed data. All analyses were conducted as 2-sided at a 0.05 significance level using SAS v9.4.

All research was conducted in accordance with both the Declarations of Helsinki and Istanbul and was approved by Indiana University Institutional Review Board (IRB). Written consent was given in writing by all subjects.

RESULTS

Questionnaire data from 2442 participants were included in this study. For our primary analysis of PROMIS-29 data, data from 1697 participants were included (n = 745 had only CLDQ data). The demographics and clinical data for these individuals are summarized in (Table 1). The median age was 60 years old (IQR: 51–66), 55.2% were male, 83.7% were White, and 16.9% were Hispanic. Education level of high school diploma or less was reported by 34% of the cohort. Setting at the time of HQROL measurement was an outpatient clinic visit (76.3%) or during hospitalization (23.7%). All inpatients (n = 402, 23.7%) had cirrhosis without a history of transplant. Of the outpatients, 490 (28.9%) were after transplant (median 3.9 y post-transplant, IQR: 1.3, 10.4). Of outpatients prior to transplant, 677 (39.9%) had cirrhosis, with 32% without any complications of cirrhosis. The most common causes of liver disease were viral hepatitis, alcohol, and metabolic-associated steatohepatitis, although there was a broad representation of many liver diseases. Comorbid conditions were common in the cohort, with a median Charlson Comorbidity index of 5 (IQR: 4, 7) even when liver disease was excluded.

HRQOL score distributions

Distribution of PROMIS domain and summary scores by clinical setting, cirrhosis status, and transplant status are displayed in Figure 1 and summarized in Supplemental Table S3, http://links.lww.com/HC9/C9. Compared to the US population mean *T*-score of 50, participants in the overall study cohort had scores generally showing impairment in all domains, with the largest deficits seen in physical function, fatigue, pain behavior, and pain interference (at least 0.5 SD from an expected score of 50). When compared to outpatients with cirrhosis, inpatients had significantly poorer HRQOL in each of the PROMIS domains ($p < 0.001$). In the outpatient subgroup, outpatients with cirrhosis and without cirrhosis had similar domain scores, except outpatients with cirrhosis reported statistically more depressive symptoms (higher score, 52.7 vs. 51.2, $p = 0.012$) compared to outpatients without cirrhosis. Post-transplant patients had significantly improved scores in all PROMIS domains when compared to pre-transplant patients ($p < 0.001$). Post-transplant participants had scores close to the US population mean for many PROMIS-29 domains but continued to report poor physical function (median: 41.8, IQR: 37.6–48.2).

For SF-36 and CLDQ, the distribution of scores is displayed in Supplemental Figure S1, http://links.lww.com/HC9/C9 and S2, http://links.lww.com/HC9/C10, and Supplemental Table S3, http://links.lww.com/HC9/C9. For SF-36 domains (n = 205), no statistically significant differences in domain scores were observed between inpatient and outpatients, nor between outpatients with versus without cirrhosis, except for domains of role limitations in physical function and role limitations in emotional function where outpatients without cirrhosis had significantly better HRQOL scores ($p < 0.05$ and $p < 0.01$, respectively). For CLDQ domains (n = 1321), scores in all domains were lower (poorer HRQOL) for inpatients compared to outpatients ($p < 0.001$) and for outpatients with versus without cirrhosis ($p < 0.01$).

Distribution-based MSD

Distribution-based MSDs for PROMIS domains ranged from 2.6 to 4.2 for individual domains and were 4.3 and 7.9 for mental and physical summary scores, respectively (Table 2 and Figure 2A: open circle). MSDs ranged from 7.3 to 17.8 for SF-36 domains, 3.0–3.2 for SF-36 component scores, 0.5–0.8 for CLDQ domains, and 0.3 for the CLDQ total score (Table 2).

MSDs calculated using various subsamples of our cohort by clinical stage revealed small variations (Figure 2A and Supplemental Table S4, http://links.lww.com/HC9/C9) while MSDs by etiology were generally similar (Supplemental Table S5, http://links.lww.com/HC9/C9). MSDs in the decompensated populations (ie, inpatients, outpatients with decompensated cirrhosis, pre-transplant patients) were marginally smaller than MSDs in stable populations (ie, compensated cirrhosis, post-transplant outpatients). For PROMIS, MSDs calculated in decompensated populations were ~0.5 points smaller than MSDs in stable populations. SF-36 MSDs were generally similar between different subgroups, while CLDQ MSDs were marginally larger in the decompensated subgroups as compared to the stable groups.

Anchor-based MSD

MSDs for each relevant anchor, as well as an assessment of conceptual similarity between the anchor and each PROMIS domain of interest (universal,

TABLE 1 Characteristics of the study population

	Overall (N = 1697), N (%)
Location of HRQOL measurement	
Inpatient	402 (23.7%)
Outpatient with cirrhosis[a]	677 (39.9%)
Outpatient without cirrhosis[a]	128 (7.5%)
Outpatient, post-transplant	490 (28.9%)
Age[b]	60.0 (51.0, 66.0)
Male	936 (55.2%)
Hispanic ethnicity[c]	286 (16.9%)
Race	
White	1420 (83.7%)
American Indian or Alaska Native	96 (5.7%)
Black or African American	88 (5.2%)
Other[d]	65 (3.8%)
Education: High School diploma or less[c]	582 (34.3%)
Insurance[c]	
Medicare	723 (42.6%)
Private/self-pay/none	542 (31.9%)
Medicaid	420 (24.7%)
Income ≤$50,000[c]	1098 (64.7%)
Lives with an adult at home[c]	1153 (67.9%)
Charlson comorbidity score[b,e]	5.0 (4.0, 7.0)
Etiology of liver disease	
Viral hepatitis (hepatitis B, C)	437 (25.8%)
Alcohol	618 (36.4%)
Metabolic dysfunction–associated steatohepatitis	526 (31.0%)
Biliary (PBC, PSC)	142 (8.4%)
Autoimmune hepatitis	96 (5.7%)
Other[f]	426 (25.1%)
Cirrhosis complications[g]	
None	329 (30.5%)
Esophageal varices	508 (47.1%)
Hepatic encephalopathy	502 (46.5%)
Fluid overload	614 (56.9%)
HCC	67 (6.2%)
Time from transplant (y)[b]	3.9 (1.3, 10.4)
Transplant complication	
None	220 (13.0%)
Biliary	70 (4.1%)
Acute rejection	27 (1.6%)
Disease recurrence/graft failure/cirrhosis	60 (3.5%)
Post-transplant malignancy	10 (0.6%)

[a]Without a history of transplant.
[b]Median (IQR).
[c]Missing data: Ethnicity (n = 11, 0.6%), Education (n = 278, 16.4%), Insurance (n = 12, 0.7%), Income (n = 12, 0.7%), and Living with adult data (n = 276, 16.3%).
[d]Other race includes: race not reported (n = 34, 2.0%), Asian (n = 26, 1.5%), and multiracial (n = 25, 1.5%).
[e]Modified to exclude liver disease.
[f]Other etiologies of liver disease include: Wilson disease (n = 3, 0.2%); alpha-1-antitrypsin deficiency (n = 45, 2.7%); hemochromatosis (n = 13, 0.8%); other biliary (n = 13, 0.8%); Drug Induced Liver Injury (n = 8, 0.5%); Budd–Chiari (n = 5, 0.3%); congestive hepatopathy (n = 2, 0.1%); hepatic sarcoidosis (n = 3, 0.2%); polycystic liver disease (n = 4, 0.2%); HIV (n = 1, 0.1%); CMV (n = 1, 0.1%); transplant rejection (n = 1, 0.1%); cryptogenic/unknown (n = 327, 19.3%).
[g]In those with cirrhosis (% reported for n = 1079).
Abbreviations: HRQOL, health-related quality of life; PBC, primary biliary cholangitis; PSC, primary sclerosing cholangitis.

general, or specific), are displayed in Supplemental Table S6, http://links.lww.com/HC9/C9. Using these selected anchors, the overall mean MSDs for each domain ranged from 3.5 to 5.2 while summary score MSDs ranged from 5.0 to 5.5 (Figure 2B and Supplemental Table S7, http://links.lww.com/HC9/C9). As with distribution-based anchors, anchor-based MSDs calculated using various clinical subsamples showed small variations when examined by clinical stage but were similar by liver disease etiology (Figure 2B and Supplemental Tables S7 and S8, http://links.lww.com/HC9/C9). MSDs using the decompensated samples tended to be smaller by ~1 point, except for the inpatient sample, where MSDs for sleep disturbance, pain behavior, and satisfaction with participation in social activities were 2 points smaller.

Recommended ranges for MSDs

Figure 3 displays recommended ranges for PROMIS-based MSDs from the study cohort triangulated using distribution-based and anchor-based methodologies and consideration of the MSDs in various subsamples described above. When applied to an individual, *T*-scores and their corresponding MSDs are rounded to the whole integer. Therefore, for PROMIS, MSDs ranging from 3 to 4 for individual domains and 4 to 6 for summary scores can be recommended.

DISCUSSION

In this study, we prospectively measured HRQOL in a large and diverse sample of individuals with CLD at various stages of liver disease. Using this representative sample, this is the first study to establish MSDs for commonly used PROMs in the liver disease population. The MSDs established by our study are crucial to the usefulness of PROMs use, both in clinical and research settings, as they inform whether PRO-score differences are large enough to matter. The role of MSDs in clinical and research endeavors is unique, yet in both settings, MSDs help bridge the gap between statistical significance and what matters to patients. In research, MSDs allow interpretation that is clinically meaningful and goes beyond using statistically significant differences,

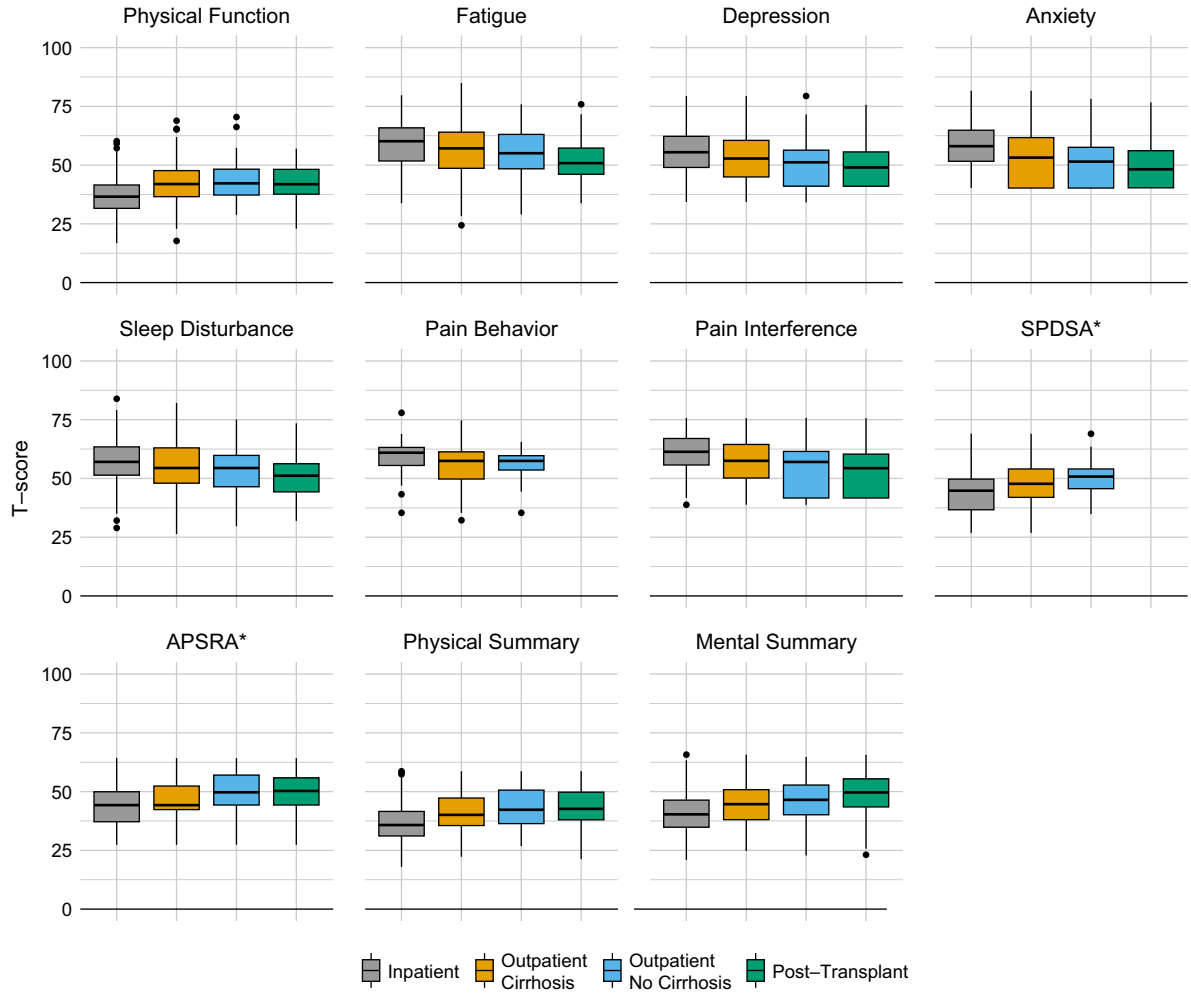


FIGURE 1 PROMIS-29 Profile domains scores by clinical setting and stage of liver disease. HRQOL for each domain was poorer in inpatients versus outpatients with cirrhosis (all *p*-values < 0.001). When comparing those with and without cirrhosis, HQROL was similar except for more depression in those with cirrhosis ($p < 0.05$). Finally, compared to those pre-transplant patients (inpatients and outpatients with cirrhosis), post-transplant participants had improved HRQOL in all domains measured ($p < 0.001$). *APSRA, ability to participate in social roles and activities; SPDSA, satisfaction with participation in discretionary social activities. Abbreviations: HRQOL, health-related quality of life; PROMIS, Patient-Reported Outcomes Measurement Information System.

which are heavily influenced by sample size and distribution of the scores.[14,18,19] Therefore, MSDs can be used throughout the research process (eg, sample size calculations and setting clinically relevant endpoints).[20,21] In clinical practice, MSDs can guide clinician–patient conversations about treatment outcomes and selecting between treatment options.[6,14,17,37]

Our analysis included PROMIS data from 1697 participants, providing a comprehensive dataset for interpreting changes in PROM scores based on both distribution-based and anchor-based methods. In addition to the large sample size, the PROM data included in this analysis represent HRQOL in patients with and without cirrhosis, hospitalized and outpatient settings, as well as before and after liver transplant, allowing us to generate MSDs across different contexts and subpopulations. A notable finding from this study is the similarity of MSDs calculated across different patient subgroups. The consistency of MSDs for PROMIS domains among various settings and disease stages supports the reliability of the MSDs developed in our analysis when applied across diverse clinical contexts and stages experienced by the CLD population. That being said, if the use of MSDs is targeted to a particular subpopulation within CLD (ie, inpatients or post-transplant patients), our study provides detailed and nuanced MSDs estimations for use. It is also important to note that when compared to prior studies in other chronic disease populations, which used similar methods, we find that MSDs in the CLD and cirrhosis populations are likely distinct compared to other chronic disease populations.[23,38–40] For example, in rheumatoid arthritis, MSDs for fatigue were estimated to be 2.6–6.8, similar to anchor-based MSD for fatigue in our population (2.7–7.2) but MSDs for pain interference

TABLE 2 Meaningful score differences for PROMIS, CLDQ, and SF-36 scores by distribution-based methodology

HRQOL domain	n	Standard error of measurement
PROMIS domains (score range 0–100)		
Physical function	1692	3.1
Fatigue	1668	2.7
Depression	1685	3.7
Anxiety	934	4.2
Sleep disturbance	1685	3.3
Pain behavior	649	2.6
Pain interference	1031	3.5
Ability to participate in social roles and activities (APSRA)	932	2.9
Satisfaction with participation in discretionary social activities (SPDSA)	752	2.6
Physical summary score	929	7.9
Mental summary score	929	4.3
SF-36 domains (score range 0–100)		
General health	188	9.7
Physical functioning	190	7.3
Role limitations, physical functioning	189	15.5
Emotional well-being (mental health)	188	7.3
Role limitations, emotional function	189	17.8
Energy/fatigue (vitality)	188	7.6
Social functioning	188	10.8
Bodily pain	188	10.0
Physical component score	188	3.0
Mental component score	188	3.2
CLDQ domains (score range 1–7)		
Abdominal symptoms	1126	0.7
Activity	1126	0.8
Emotional function	1124	0.5
Fatigue	1126	0.5
Systemic symptoms	1124	0.8
Worry	1123	0.5
Total	1124	0.3

Abbreviations: CLDQ, Chronic Liver Disease Questionnaire; HRQOL, health-related quality of life; PROMIS, Patient-Reported Outcomes Measurement Information System; SF-36, Short Form-36.

were 2.7–6.1, which are larger than the range estimated in our study (3.5–4.4).[38] Therefore, our study, which is the first to estimate MSDs in liver disease patients, adds an important data to the PROM field and current repository of MSDs for PROMIS domains.[40]

We also show that the methodology used to develop MSDs impacts the MSD estimate. In our study, using the mostly conceptually similar anchors (termed "specific") to a particular domain generated the largest MSDs, while using more general anchors or the distribution-based methodology generated smaller MSD estimates. Variation in MSDs when comparing distribution versus anchor-based calculations have been reported, fewer studies have examined the effect of conceptual similarity to MSD estimation.[19,33,36,41–43] Prior studies have used cut-offs for correlation coefficients when selecting an anchor, where moderate correlations are deemed to be adequate for identifying a suitable anchor.[24,33,36,44] Despite adhering to these cut-offs (all correlation coefficients for all anchors were >0.30 and often >0.40 in our study), we still noted variability in MSD estimates. For example, for the fatigue domain, when using a universal anchor such as "In general, would you say your health is?" versus a general anchor such as "In general, how would you rate your physical health?" to a more specific anchor such as "How would you rate your fatigue on average?", MSDs varied from 5.2 to 4.6 to 7.2, respectively. Given this variability, future studies that aim to establish MSDs should carefully select anchors not just by statistical correlation between scores but also based on conceptual similarity. Prior work has also shown the added utility of using longitudinal anchor-based methodology to further refine MSDs.[23,43,44] For example, MSDs from samples with worsening disease can be different than those with improving disease.[19,24,32,43] Future studies looking at longitudinal change in PROs partnered with carefully selected anchors are needed to further refine the MSD estimations from our study.

Beyond MSDs, we also establish PROMIS domain scores for CLD, cirrhosis, and patients after liver transplantation (LT) by using a large and diverse study cohort. The observed differences in HRQOL scores among different clinical settings and disease stages offer valuable insights into the dynamic nature of PROM scores in patients with CLD and cirrhosis. For instance, we showed that post-transplant patients had significantly better HRQOL scores compared to pre-transplant patients. Additionally, inpatients with cirrhosis had notably poorer scores than outpatients with cirrhosis. We also observed that outpatients without cirrhosis have poor PROM scores for most domains, scoring as poorly as outpatients with cirrhosis. While the present data are cross-sectional, few studies have reported HRQOL at different stages of liver disease, especially as measured by PROMIS, and those studies were based on older cohorts.[16,45] With the significant shifts in causes of liver disease over the past decade,[46] our study provides a contemporary look at PROMIS-29, CLDQ, and SF-36 scores in the CLD, cirrhosis, and post-LT population. These scores, when partnered with their MSDs to enhance interpretations, can be referenced as benchmarks for anticipated PROM scores and changes in both clinical care and research.

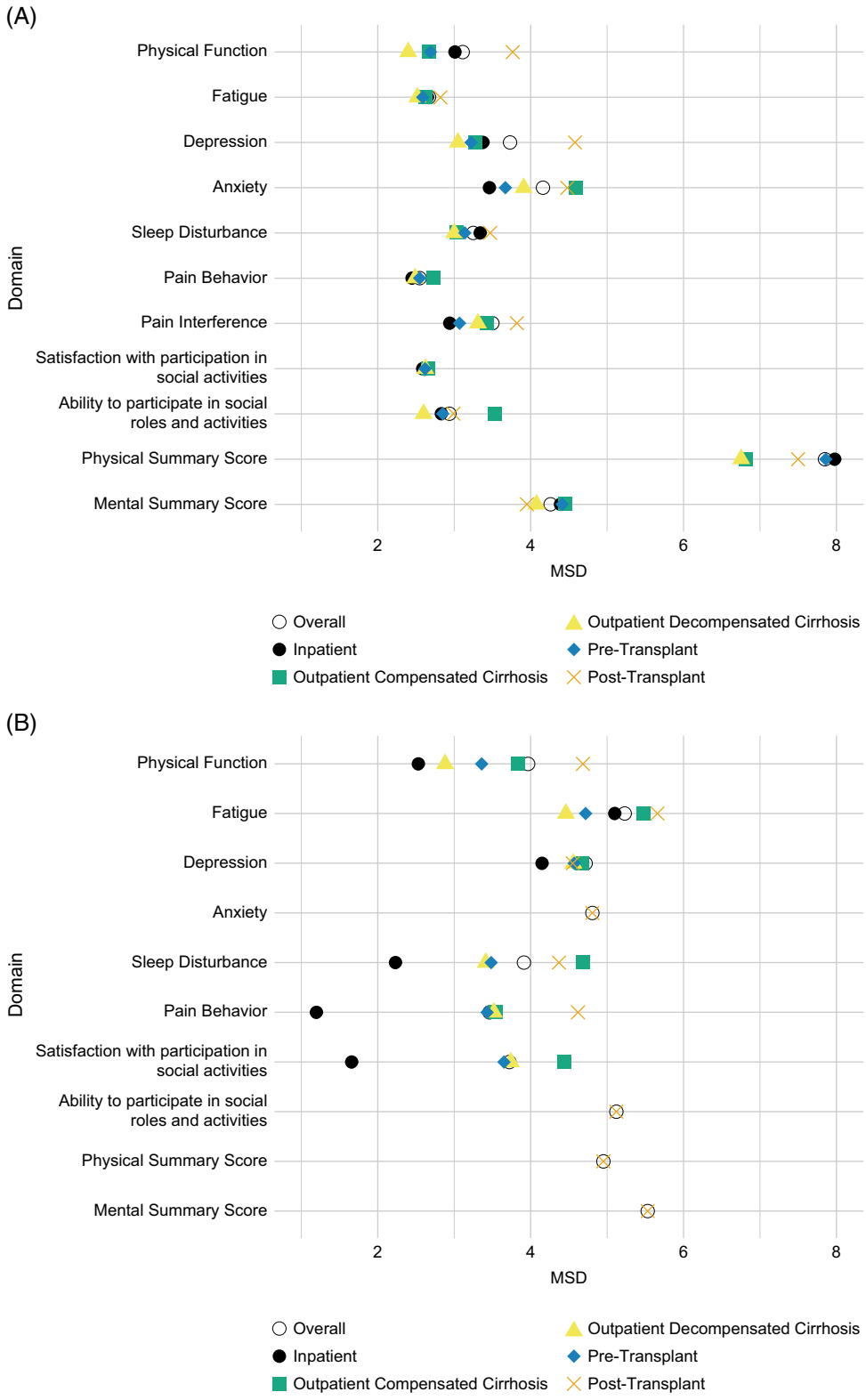


FIGURE 2 PROMIS-29 Profile domain and summary score meaningful score differences by clinical setting and stage of liver disease using (A) distribution-based and (B) anchor-based methodology. Abbreviations: MSD, meaningful score difference; PROMIS, Patient-Reported Outcomes Measurement Information System.

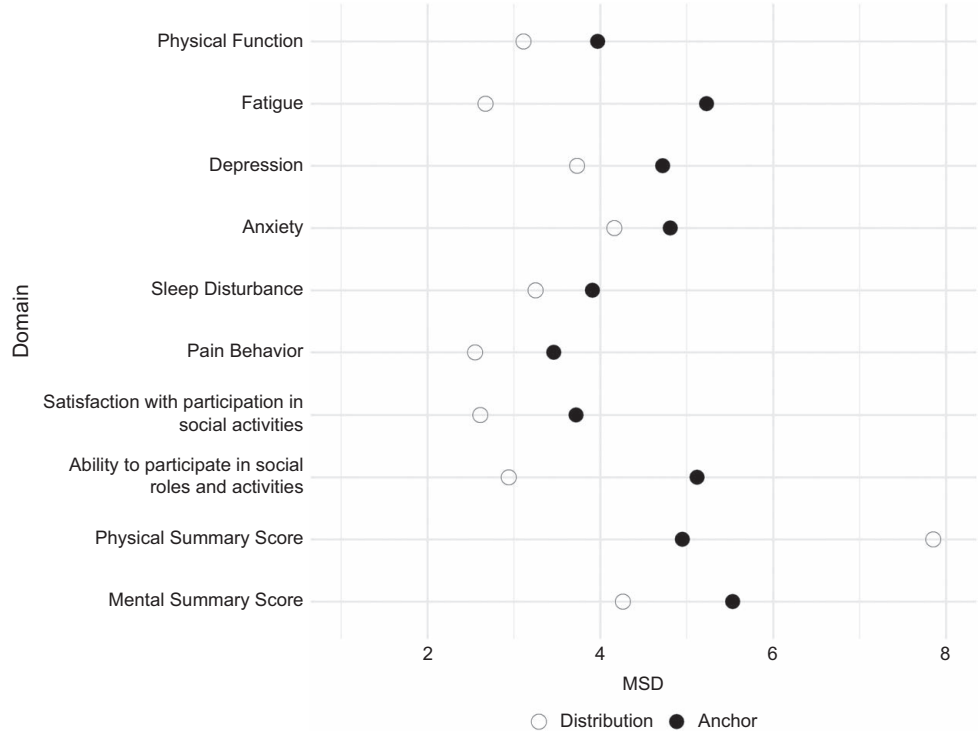


FIGURE 3 Triangulation of PROMIS-29 Profile domain and summary score meaningful score differences. MSDs using the distribution-based method are displayed alongside anchor-based MSDs. Abbreviations: MSD, meaningful score difference; PROMIS, Patient-Reported Outcomes Measurement Information System.

Our study is not without limitations. First, MSDs for SF-36 and CLDQ were derived using distribution-based methods due to the lack of suitable anchor-based data. Similarly, as our data was collected as part of several completed and ongoing studies, anchors were not available for all PROMIS domains of interest within all clinical contexts or stages of liver disease (eg, ability to participate in social roles and activities). Additionally, certain samples of the CLD population, such as those with HCC, were adequate for MSD estimation (generally $n > 10$) but smaller than other subgroups, and therefore, were not specifically reported. Furthermore, while we used individual-level data for MSD estimation, we did not have longitudinal data to derive MSDs based on clinical change at the individual level, which is another important methodology for MSD development.[19,24,32,43] Future work is needed to calculate longitudinal anchor-based MSDs in those experiencing both worsening and improvement of their liver disease and its complications. In addition, MSDs are likely population and context-specific.[19,33,36,43] While we have reported MSDs based on various contexts and subpopulations, future studies could include international populations or focus on particular complications and symptoms of cirrhosis, as these may generate MSD estimates different than ours. Importantly, given the inherent variability of MSD estimates, we provide MSDs for important subgroups of individuals with CLD in addition to carefully triangulated MSDs from the entire cohort. This allows the end user to select within a range of MSDs based on intended clinical or research aim, accounting for the domain of HRQOL, clinical context, and stage of liver disease of the intended population.

In summary, by triangulating MSDs for commonly used PROMs in liver disease, we aimed to facilitate the routine use of PROs in both daily practice and as primary outcomes in health services research. Advances in PROMs support the systematic measurement of self-reported well-being during CLD clinical care and research through tools such as the PROMIS-29 Profile, CLDQ, and the SF-36, which all provide multiple scores across different domains of a patient's health, including physical, emotional, and social well-being. Despite these advances, integration of PROMs into liver disease care and research is in its infancy.[3,37] Lessons from other fields, such as oncology, demonstrate that simply measuring patient well-being is not enough. PROM must be partnered with guidance on the interpretation of PROM scores.[5,12,17,47–50] MSDs play a key role in PROM score interpretation by setting thresholds that can indicate a clinically relevant or meaningful change moving beyond statistically detectable changes. Our study adds a new dimension to the existing knowledge of the role of PROMs in liver disease by offering specific thresholds that can guide treatment evaluations and patient care decisions.

DATA AVAILABILITY STATEMENT

The analytic methods used in this study are detailed in the Methods section. The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

Study concept and design: Archita Desai and Patrick Monahan. Data analysis: Archita Desai, Timothy Stump, and Patrick Monahan. Data analysis: Archita Desai, Timothy Stump, and Patrick Monahan. Manuscript preparation: Archita Desai and Tarek Aridi. Critical manuscript review: all authors.

FUNDING INFORMATION

Archita P. Desai is funded by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award numbers K23DK123408 and R03DK139283. The collection of these data was also supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award numbers K23DK109202 and KL2TR002530 to ESO. The funder was not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the Thomas Boyer Liver Institute and the Division of Gastroenterology and Hepatology at the University of Arizona College of Medicine in Tucson, Arizona, USA, for their support of this work.

CONFLICTS OF INTEREST

Naga Chalasani reports paid consulting agreements with Madrigal, GSK, Zydus, Altimune, BioMea Fusion, Ipsen, Akero, Merck, and Pfizer. He has research grants from Boehringer-Ingelheim and Exact Sciences. He has equity ownership in Avant Sante, a contract research organization, and Heligenics, a drug discovery start-up company. Marwan S. Ghabril reports unpaid consulting agreements with Biocrust, Cymabay, Gilead, and Zydus. He has research support from Bausch/Salix. He also consults for CymaBay, Gilead, Zydus, and BioCryst. Eric Orman consults for BioVie and Sitero. The remaining authors have no conflicts to report.

ORCID

Archita P. Desai https://orcid.org/0000–0001–8731–2370
Eric S. Orman https://orcid.org/0000–0002–8295–9076
Tarek G. Aridi https://orcid.org/0000–0002–3812–7686
Timothy Stump https://orcid.org/0000–0001–9593–3011
Lauren Nephew https://orcid.org/0000–0003–0837–0746
Marwan S. Ghabril https://orcid.org/0000–0002–4784–3246
Naga Chalasani https://orcid.org/0000–0003–4082–3178
Patrick Monahan https://orcid.org/0000–0001–9481–1995

REFERENCES

1. Porter ME. What is value in health care? N Engl J Med. 2010; 363:2477–81.
2. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. National Academies Press; 2001:1192.
3. Verma M, Desai AP. The role of patient-reported outcomes in a patient-centered care model for managing chronic liver diseases. Clin Liver Dis. 2024;23:e0222.
4. Campbell R, Ju A, King MT, Rutherford C. Perceived benefits and limitations of using patient-reported outcome measures in clinical practice with individual patients: A systematic review of qualitative studies. Qual Life Res. 2021;31:1597–620.
5. Gibbons C, Porter I, Goncalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev. 2021;10:CD011589.
6. Aiyegbusi OL, Nair D, Peipert JD, Schick-Makaroff K, Mucsi I. A narrative review of current evidence supporting the implementation of electronic patient-reported outcome measures in the management of chronic diseases. Ther Adv Chronic Dis. 2021; 12:20406223211015958.
7. Liu JB, Kaplan RS, Bates DW, Edelen MO, Sisodia RC, Pusic AL. Mass General Brigham's Patient-Reported Outcomes Measurement System: A decade of learnings. NEJM Catalyst. 2024;5. doi:10.1056/cat.23.0397
8. Groenewegen A, van Muilekom MM, Sieben CHAM, Luijten MAJ, Janssen SL, Venema-Taat N, et al. The Amsterdam PROM Implementation Strategy: Policy and pathway. NEJM Catalyst. 2024;5. doi:10.1056/cat.23.0414
9. Spencer KL, Absolom KL, Allsop MJ, Relton SD, Pearce J, Liao K, et al. Fixing the leaky pipe: How to improve the uptake of patient-reported outcomes-based prognostic and predictive models in cancer clinical practice. JCO Clin Cancer Inform. 2023;7:e2300070.
10. Hancock SL, Ryan OF, Marion V, Kramer S, Kelly P, Breen S, et al. Feedback of patient-reported outcomes to healthcare professionals for comparing health service performance: A scoping review. BMJ Open. 2020;10:e038190.
11. Snyder CF, Smith KC, Bantug ET, Tolbert EE, Blackford AL, Brundage MD; PRO Data Presentation Stakeholder Advisory Board. What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians to improve interpretability. Cancer. 2017;123:1848–59.
12. Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders (Guidance 4); April 2023.
13. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, et al. Guidelines for inclusion of patient-reported outcomes in

clinical trial protocols: The SPIRIT-PRO extension. JAMA. 2018; 319:483–94.
14. Coon CD, Cook KF. Moving from significance to real-world meaning: Methods for interpreting change in clinical outcome assessment scores. Qual Life Res. 2018;27:33–40.
15. Desai AP, Madathanapalli A, Tang Q, Orman ES, Lammert C, Patidar KR, et al. PROMIS Profile-29 is a valid instrument with distinct advantages over legacy instruments for measuring quality of life in chronic liver disease. Hepatology. 2023;78: 1788–99.
16. Younossi ZM, Golabi P, Henry L. A comprehensive review of patient-reported outcomes in patients with chronic liver diseases. J Clin Gastroenterol. 2019;53:331–41.
17. Kroenke K, Miksch TA, Spaulding AC, Mazza GL, DeStephano CC, Niazi SK, et al. Choosing and using patient-reported outcome measures in clinical practice. Arch Phys Med Rehabil. 2022;103:S108–17.
18. Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders (Guidance 3); 2022.
19. Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, et al. Minimal important change (MIC): A conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. 2021;30:2729–54.
20. U.S. Food & Drug Administration. FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient's Voice in Medical Product Development and Regulatory Decision Making. Accessed June 2024. https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical
21. European Medicines Agency: Reflection Paper on the use of patient reported outcome (PRO) measures in oncology studies; 2014.
22. Khanna D, Furst DE, Maranian P, Seibold JR, Impens A, Mayes MD, et al. Minimally important differences of the UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. J Rheumatol. 2011;38:1920–4.
23. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol. 2011;64:507–16.
24. Khanna D, Hays RD, Shreiner AB, Melmed GY, Chang L, Khanna PP, et al. Responsiveness to change and minimally important differences of the Patient-Reported Outcomes Measurement Information System Gastrointestinal Symptoms scales. Dig Dis Sci. 2017;62:1186–92.
25. Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and minimally important differences for 4 Patient-Reported Outcomes Measurement Information System Short Forms: Physical Function, Pain Interference, Depression, and Anxiety in knee osteoarthritis. J Pain. 2017;18:1096–110.
26. Steinhaus ME, Iyer S, Lovecchio F, Khechen B, Stein D, Ross T, et al. Minimal clinically important difference and substantial clinical benefit using PROMIS CAT in cervical spine surgery. Clin Spine Surg. 2019;32:392–7.
27. Kroenke K, Stump TE, Chen CX, Kean J, Bair MJ, Damush TM, et al. Minimally important differences and severity thresholds are estimated for the PROMIS depression scales from three randomized clinical trials. J Affect Disord. 2020;266:100–8.
28. Tack J, Camilleri M, Hale M, Morlion B, Nalamachu S, Webster L, et al. Establishing minimal clinically important differences in quality of life measures in opioid-induced constipation. Clin Gastroenterol Hepatol. 2022;20:855–63.
29. Spiegel BM, Younossi ZM, Hays RD, Revicki D, Robbins S, Kanwal F. Impact of hepatitis C on health related quality of life: A systematic review and quantitative assessment. Hepatology. 2005;41:790–800.
30. Anderson RT, Baran RW, Erickson P, Revicki DA, Dietz B, Gooch K. Psychometric evaluation of the hepatitis C virus patient-reported outcomes (HCV-PRO) instrument: Validity, responsiveness, and identification of the minimally important difference in a phase 2 clinical trial. Qual Life Res. 2014;23: 877–86.
31. Hays RD, Spritzer KL, Schalet BD, Cella D. PROMIS®-29 v2.0 profile physical and mental health summary scores. Qual Life Res. 2018;27:1885–91.
32. Chen CX, Kroenke K, Stump TE, Kean J, Carpenter JS, Krebs EE, et al. Estimating minimally important differences for the PROMIS pain interference scales: Results from 3 randomized clinical trials. Pain. 2018;159:775–82.
33. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61: 102–9.
34. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care. 1999;37:469–78.
35. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol. 1999;52:861–73.
36. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. COPD. 2005;2:63–7.
37. Tapper EB, Lai JC. Time to embrace PROMIS-29 as the standard health-related quality of life instrument for patients with cirrhosis. Hepatology. 2023;78:1688–91.
38. Bartlett SJ, Gutierrez AK, Andersen KM, Bykerk VP, Curtis JR, Haque UJ, et al. Identifying minimal and meaningful change in a patient-reported outcomes measurement information system for rheumatoid arthritis: Use of multiple methods and perspectives. Arthritis Care Res (Hoboken). 2022;74:588–97.
39. Saygin D, DiRenzo D, Raaphorst J, de Groot I, Bingham CO, Lundberg IE, et al. Responsiveness and meaningful thresholds of PROMIS pain interference, fatigue, and physical function forms in adults with idiopathic inflammatory myopathies: Report from the OMERACT Myositis Working Group. Semin Arthritis Rheum. 2024;64:152339.
40. Carrasco-Labra A, Devji T, Qasim A, Phillips MR, Wang Y, Johnston BC, et al. Minimal important difference estimates for patient-reported outcomes: A systematic survey. J Clin Epidemiol. 2021;133:61–71.
41. Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the patient-reported outcomes measurement information system (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. Ann Rheum Dis. 2015;74:104–7.
42. Beaumont JL, Davis ES, Curtis JR, Cella D, Yun H. Meaningful change thresholds for Patient-Reported Outcomes Measurement Information System (PROMIS) fatigue and pain interference scores in patients with rheumatoid arthritis. J Rheumatol. 2021; 48:1239–42.
43. Wang Y, Devji T, Qasim A, Hao Q, Wong V, Bhatt M, et al. A systematic survey identified methodological issues in studies estimating anchor-based minimal important differences in patient-reported outcomes. J Clin Epidemiol. 2022;142:144–51.
44. Zigler CK, Li Z, Hernandez A, Randell RL, Mann CM, Weitzman E, et al. Evaluating anchor variables and variation in meaningful score differences for PROMIS(R) Pediatric measures in children and adolescents living with a rheumatic disease. Qual Life Res. 2024;33:3449–57.

45. Tapper EB, Kanwal F, Asrani SK, Ho C, Ovchinsky N, Poterucha J, et al. Patient-reported outcomes in cirrhosis: A scoping review of the literature. Hepatology. 2018;67:2375–83.
46. Devarbhavi H, Asrani SK, Arab JP, Nartey YA, Pose E, Kamath PS. Global burden of liver disease: 2023 update. J Hepatol. 2023;79:516–37.
47. Skovlund PC, Ravn S, Seibaek L, Thaysen HV, Lomborg K, Nielsen BK. The development of PROmunication: A training-tool for clinicians using patient-reported outcomes to promote patient-centred communication in clinical cancer settings. J Patient Rep Outcomes. 2020;4:10.
48. Snyder C, Smith K, Holzner B, Rivera YM, Bantug E, Brundage M. Panel PRODPD. Making a picture worth a thousand numbers: Recommendations for graphically displaying patient-reported outcomes data. Qual Life Res. 2019;28:345–56.
49. Mierzynska J, Piccinin C, Pe M, Martinelli F, Gotay C, Coens C, et al. Prognostic value of patient-reported outcomes from international randomised clinical trials on cancer: A systematic review. Lancet Oncol. 2019;20:e685–98.
50. European Medicines Agency. EMA Regulatory Science to 2025—Strategic reflection; 2020.

How to cite this article: Desai AP, Orman ES, Aridi TG, Stump T, Nephew L, Ghabril MS, et al. Meaningful differences in patient reported outcome measurement scores in liver disease. Hepatol Commun. 2025;9:e0727. https://doi.org/10.1097/HC9.0000000000000727