

Prospective validation of the short form liver disease quality of life instrument

F. KANWAL*, †, B. M. R. SPIEGEL†, ‡, §, R. D. HAYS‡, ¶, F. DURAZO‡, S. B. HAN‡, §, S. SAAB‡, R. BOLUS†, S. J. KIM† & I. M. GRALNEK**

*John Cochran VA Medical Center, Saint Louis, and Division of Gastroenterology, Saint Louis University, St Louis, MO, USA; †Center for Outcomes Research and Education at UCLA, ‡Department of Internal Medicine, David Geffen School of Medicine at UCLA and SVA Greater Los Angeles Healthcare System, Los Angeles, CA, USA; ¶RAND, Santa Monica, CA, USA; **Department of Gastroenterology, Rappaport Faculty of Medicine, Technion Israel Institute of Technology, Rambam Medical Center, Haifa, Israel

Correspondence to:
Prof. I. M. Gralnek, Department of Gastroenterology, Rappaport Faculty of Medicine, Technion Israel Institute of Technology, Rambam Health Care Campus, Haifa, Israel 31096.
E-mail: i_gralnek@rambam.health.gov.il

Publication data

Submitted 22 February 2008
First decision 13 March 2008
Resubmitted 06 July 2008
Accepted 25 July 2008
Epub Accepted Article 30 July 2008

SUMMARY

Background

Despite the realization that health-related quality of life (HRQOL) is an important outcome in patients with liver disease, there is scarcity of disease-targeted HRQOL measures that have undergone prospective evaluation.

Aim

To validate prospectively the short form of liver disease quality of life instrument (the SF-LDQOL) in patients with advanced liver disease.

Methods

The SF-LDQOL includes 36 disease-targeted items representing nine domains: symptoms of liver disease, effects of liver disease, memory/concentration, sleep, hopelessness, distress, loneliness, stigma of liver disease and sexual problems. We administered the SF-LDQOL to 156 advanced liver disease patients at baseline and at 6-month follow-up. We estimated internal consistency reliability for multi-item scales, item discrimination across scale and evaluated construct validity by estimating the associations of SF-LDQOL scores with SF-36 scores, symptom severity and disability days. To evaluate the SF-LDQOL's responsiveness, we compared HRQOL changes for patients who received with those who did not receive liver transplantation (LT).

Results

The internal consistency reliability coefficients were ≥ 0.70 for seven of nine scales in baseline and for all scales in follow-up administration. The SF-LDQOL correlated highly with SF-36 scores, symptom severity, disability days and global health. Patients undergoing LT reported improved HRQOL compared with patients without LT and the responsiveness indices were excellent.

Conclusions

This study provides support for the reliability and validity of the SF-LDQOL in patients with advanced chronic liver disease. This instrument may be useful in everyday clinical practice and in future clinical trials.

Aliment Pharmacol Ther 28, 1088–1101

INTRODUCTION

Cirrhosis is a prevalent and expensive condition affecting 5.5 million Americans at a cost of over \$1.5 billion annually.¹⁻³ This economic burden is multiplied by the dramatic impact of cirrhosis on health-related quality of life (HRQOL) resulting from complications of advanced liver disease, including encephalopathy, ascites and hepatocellular cancer.⁴⁻¹⁰ Despite the realization that HRQOL is an important outcome in patients with cirrhosis, clinicians rarely assess HRQOL in patients with advanced liver disease.

One explanation underlying this apparent disconnect may be the perceived respondent burden (length) of available HRQOL outcome measures.¹¹ Although several multidimensional scales have been developed to measure HRQOL outcomes in medicine,¹²⁻¹⁶ these instruments are rarely used in everyday practice because of time and resource constraints. Shorter HRQOL instruments reduce respondent burden and thus increase the likelihood of their use in clinical practice. However, although this approach is conceptually appealing, it remains untested whether the shorter instruments can adequately capture HRQOL in patients with advanced liver disease.¹¹ Another reason for limited use of HRQOL assessments in patients with advanced liver disease may be the scarcity of disease-targeted HRQOL measures that can detect clinically meaningful change in patients' health over time. Outcome measures are expected to detect change in health over time to assess effectiveness of health care.^{17, 18} This property, known as responsiveness, is key in the evaluation of an HRQOL measure and must be demonstrated prior to its use in routine clinical care and prospective clinical trials.¹⁷⁻²⁰ In the absence of a responsive disease-targeted HRQOL measure in chronic liver disease, clinicians and investigators must rely on either biological outcomes (e.g. MELD scores) or generic HRQOL measurements (e.g. SF-36) to assess their patients' health status over time. Unfortunately, biological outcomes may fail to measure the patients' perception of their disease and generic HRQOL measurements may fail to capture the most important disease-targeted components of HRQOL in chronic liver disease.^{4, 19-21}

We sought to develop and prospectively evaluate a short, disease-targeted, multidimensional HRQOL instrument in patients with cirrhosis and advanced liver disease, the short form liver disease quality of life (SF-LDQOL). We hypothesized that in patient with

advanced liver disease: (i) the SF-LDQOL would capture the decrement in HRQOL engendered by cirrhosis; (ii) HRQOL would decline with worsening overall health (measured by patient self-report) and liver disease severity (measured by biological disease severity indices); (iii) HRQOL would significantly improve after undergoing liver transplantation (LT) and (iv) these changes in HRQOL would be accurately measured over time using the SF-LDQOL.

METHODS

Development of the SF-LDQOL

The Liver Disease Quality of Life version 1.0 (LDQOL 1.0) is a disease-targeted, self-completed, multidimensional HRQOL instrument for patients with advanced chronic liver disease.²² The LDQOL 1.0 was previously developed and tested in a cross-sectional, multicentre study including 221 patients with advanced liver disease.²² The content validity of the LDQOL 1.0 was supported by a systematic literature review, a focus group with expert hepatologists, cognitive interviews of patients with advanced liver disease and confirmation of *a priori* hypotheses regarding relevant HRQOL domains. The resulting questionnaire was field tested in a cohort of geographically diverse patients with advanced liver disease. The results of this multicentre field test provide support for the reliability and construct validity of the LDQOL 1.0 and are detailed elsewhere.²²

The LDQOL 1.0 uses the Short Form-36 version 2.0 (Quality-metric, Lincoln, RI, USA) as the generic core, supplemented by 75 disease-targeted items grouped in 12 scales. The SF-36 is a widely used generic HRQOL instrument.²³ It measures eight aspects of HRQOL: physical functioning, role limitation – physical, bodily pain, general health, role limitation – emotional, vitality, emotional well-being and social functioning and these in turn are compiled into two summary scores: the physical component summary (PCS) score and the mental component summary (MCS) score.²³ The disease-targeted scales of LDQOL 1.0 include liver disease-related symptoms, liver disease-related effects on activities of daily living, concentration, memory, quality of social interaction, health distress, sleep, loneliness, hopelessness, self-perceived stigma of liver disease, sexual functioning and sexual problems.²²

Using HRQOL data from the original psychometric evaluation of the LDQOL 1.0, we selected a subset of

disease-targeted items to be included in the SF-LDQOL. We based the disease-targeted item selection on maximum R^2 regression approach and internal consistency reliability coefficients (Cronbach α).^{24–26} Refer to Supporting information Table S1 for the descriptions of our disease-targeted item selection process and maximum R^2 regression technique. Briefly, using maximum R^2 regression method, we selected the ‘best’ subset of disease-targeted items that accounted for $\geq 90\%$ of the variance in the total score for each disease-targeted scale.²⁶ We retained the selected subset if the Cronbach α statistic of the resulting scale exceeded the prespecified threshold ≥ 0.70 .^{14, 25} Conversely, we added back original disease-targeted items in the scales, if the Cronbach α value did not meet the threshold specified above. The selected items comprise the ‘Short Form’ of the LDQOL (See Table S2 for the disease-targeted items included in the SF-LDQOL).

We based our *a priori* scale configurations on the original scales included in the LDQOL 1.0, interscale correlations and content coverage. We subsequently tested our *a priori* scales using multitrait scaling and factor analysis of patient response data, as described below. We retained the generic SF-36 core of LDQOL 1.0 to supplement the disease-targeted SF-LDQOL scales, thereby maintaining the original hybrid structure of the instrument.¹⁹ The hybrid structure combines the generic and disease-targeted components, provides more comprehensive HRQOL information and is preferable to generic or disease-targeted measures alone.¹⁹

Prospective testing of the SF-LDQOL

We prospectively administered the resulting SF-LDQOL instrument to a cohort of 156 patients with advanced liver disease. Subjects were eligible to participate if they were >18 years old, had advanced chronic liver disease and were awaiting LT at the UCLA–Dumont Liver Transplant Program, Los Angeles, CA. Patients were excluded if they previously received LT, had grade ≥ 2 hepatic encephalopathy despite optimal medical management as determined by the primary hepatologist or if they participated in the initial cross-sectional field test of the LDQOL 1.0. We followed all participants longitudinally and invited them to complete the SF-LDQOL, 6 months after the initial administration, for prospective validation. Participants completed the SF-LDQOL at the time of routine office visits. We mailed the SF-LDQOL (with return postage

paid) to those participants who did not have a patient care encounter scheduled within a month of the expected follow-up date. We maximized participant response rates for mailed questionnaires by using a reminder postcard and telephone calls. All patients completed the generic SF-36 and rated self-reported symptom severity and self-reported disability days during both baseline and follow-up administrations of the SF-LDQOL. We also computed MELD scores for all patients according to published algorithms.^{27, 28} This study was approved by the Institutional Review Board at the University of California, Los Angeles.

Psychometric analyses

Multitrait scaling and factor analyses. To test our hypotheses regarding *a priori* HRQOL scales, we conducted a multi-item, multitrait scaling analysis to examine the extent to which items in the SF-LDQOL correlate with their hypothesized scales (item correlation, corrected for item overlap with the scale of $r \geq 0.4$) and to determine whether that correlation was significantly higher than the correlations of the item with other scales (i.e. ≥ 2 standard error).^{19, 28–30} We assessed the intercorrelation of the subscales to determine whether a single, higher-order ‘overall’ disease-targeted SF-LDQOL scale was empirically supported.³⁰ Items within scales were averaged together and scale scores were linearly transformed to a 0–100 possible range, where higher scores denote better HRQOL.^{26, 30}

Reliability testing. Reliability and validity are the two fundamental psychometric properties of an HRQOL instrument.^{19, 25, 30} Internal consistency reliability refers to the extent to which different items in an instrument are measuring the same underlying construct of interest.^{14, 25} We estimated the internal consistency reliability of the SF-LDQOL by calculating a Cronbach’s α statistic for each scale. As stated above, we considered a Cronbach’s $\alpha > 0.70$ adequate evidence for internal consistency reliability.^{14, 25}

Construct validation. Construct validity is the degree to which a measure is associated with other variables in hypothesized ways.^{19, 25, 30} To establish construct validity of an HRQOL instrument, the HRQOL scores can be compared with concurrently measured disease-specific biological, clinical or patient-reported

measures (a.k.a. anchors). We compared the disease-targeted SF-LDQOL scores with the following concurrently measured anchors:

SF-36 health survey. The individual SF-36 scales and summary scores are scored from 0 to 100 (100 = best HRQOL). The SF-36 summary scores are scored on a *T*-score metric [mean of 50, standard deviation (s.d.) of 10 in US general population].²³ On the basis of content coverage of the SF-LDQOL and SF-36 scales as well as previously published empiric data,²² we hypothesized that there would be moderate positive correlation ($r \geq 0.4$) between the overall SF-LDQOL and SF-36 PCS and MCS. For example, the SF-36 PCS measures bodily pain. The SF-LDQOL 'symptoms of liver disease' scale, in addition to measuring pain, also quantifies symptoms that are more specific to cirrhosis such as swelling, shortness of breath and gum bleeding. We further hypothesized a positive correlation ($r \geq 0.4$) between the SF-36 PCS and symptoms of liver disease and effects of liver disease scales; and between the SF-36 MCS and hopelessness, distress, loneliness and stigma of liver disease scales.

Patient self-reported symptom severity and disability days. In addition to completing the SF-36, we asked patients to report their self-rated severity of liver disease symptoms on a five-point categorical response scale (1 = no symptoms; 5 = extremely severe symptoms). Patients also reported the number of disability days attributable to their liver disease in the preceding month as 0, 1–10, 11–20 and ≥ 21 days. On the basis of previous data, we anticipated a moderate negative correlation ($r \geq -0.4$) between SF-LDQOL scores and patients' self-reported symptom severity and disability days.²²

Patient liver disease severity – MELD score. The MELD score is an objective liver disease severity scoring index based on three laboratory values – serum bilirubin, creatinine and international normalized ratio.^{27, 28} MELD score is a reliable and valid predictor of short-term mortality in patients with cirrhosis.²⁸ On the basis of our previous work, we hypothesized that there would be weak negative correlation ($r \geq -2.0$ and $r < -0.4$) between the SF-LDQOL and MELD scores.⁴

Responsiveness testing. Responsiveness is an important aspect of validity and is the degree to which an instrument is able to detect clinically meaningful change in an individual's health over time.^{17, 31} *A priori*, we hypothesized that patients receiving LT

between their baseline and follow-up SF-LDQOL administrations would experience a clinically meaningful change.^{7, 32, 33} To measure this and establish the responsiveness of the SF-LDQOL, we first computed between group *t*-tests to examine differences in changes in HRQOL between transplanted patients (i.e. changed group) and nontransplanted patients (i.e. stable group). Second, we conducted multivariate analyses to determine factors predictive of change in the overall SF-LDQOL score and in the scores for the individual disease-targeted scales. In these analyses, we examined the socio-demographic variables (e.g. age, gender, race, socioeconomic status, etc.), baseline severity of disease, baseline HRQOL scores and interval between baseline and follow-up HRQOL administrations. Third, we assessed the SF-LDQOL's responsiveness to LT using three standardized measures: the effect size (ES), standardized response means (SRMs) and responsiveness statistics (RS).²⁵ For all three responsiveness measures, the numerator is the mean change in HRQOL for the 'changed group' and the denominators are the s.d. at baseline (ES), the s.d. of change for the entire sample (SRM) and the s.d. of change for the 'stable group' (RS). Results from all three measures are generally comparable for a given criterion of change.²⁵ The ES also provides a measure of the magnitude of change with an ES ≥ 0.2 representing a small change, ≥ 0.5 a medium change and ≥ 0.8 a large change.³⁴ Finally, we determined associations between changes in SF-LDQOL score and changes in SF-36 PCS, MCS, patients' rating of their global health, and MELD scores between the baseline and follow-up SF-LDQOL administrations.

Our sample had 80% power (two-tailed 5% significance, assuming 0.60 pre-post correlations) to detect a 0.5 s.d. (10 point) difference in HRQOL scale scores.

RESULTS

Development of the SF-LDQOL

We included 36 disease-targeted items in the SF-LDQOL on the basis of results of the maximum R^2 regression analysis and internal consistency reliability coefficients. Please refer to the Supporting information Table S1 for detailed results of our item selection process. There was a significant overlap of content in the concentration and memory scales (interscale correlation = 0.88) and in sexual functioning and sexual

problems scales (interscale correlation = 0.96) of the LDQOL 1.0. As these scales provide essentially no unique information, these items were combined into one scale each – memory/concentration and sexual functioning/problem scales respectively. As a result, 36 items were grouped into nine disease-targeted scales (Supporting information Table S1). When administered together with the SF-36 generic core, the SF-LDQOL took a mean of 18 (s.d. = 9) min to complete. This compares to 38 (s.d. = 20) min for the LDQOL 1.0 as documented in the original field-test.²²

Subject characteristics

Table 1 lists other descriptive results for the sample characteristics. The mean age of the study participants was 54 (s.d. = 11) years and 55% were male. Sixty-six per cent of our cohort was Caucasian and 36% described themselves as Hispanic. Two-thirds of the subjects were married, one-third had attended college and approximately half of the cohort had either HMO/PPO or private health insurance coverage. Chronic hepatitis C was the most common diagnosis (32%) causing advanced liver disease followed by alcohol-related liver disease (12%) and cryptogenic cirrhosis (12%). The mean MELD score of our cohort at the baseline administration of SF-LDQOL was 18.5 (s.d. = 4). Of the 156 participants, 86 (55%) completed the 6-month follow-up SF-LDQOL survey. Twenty-seven patients passed away before completing the follow-up survey. The remaining 43 patients did not complete the follow-up survey as they were lost to follow-up ($n = 24$) or were too sick to complete study questionnaire ($n = 19$). With the exception of race distribution (Caucasian patients were more likely to complete the follow-up survey than non-Caucasians, 58% vs. 42%, $P = 0.04$), there were no significant differences in the baseline characteristics between patients who completed and those who did not complete the follow-up survey.

Multitrait scaling and factor analyses

With the exception of items included in the sexual functioning/problems scales, data were missing for <5% of the items. As the sexual functioning/problem scale applies only to the subset of population with recent sexual activity (i.e. within the previous 4 weeks), these items were completed by only 49% of patients in our cohort. The frequency distribution of

individual items showed that all of the response choices were used. Results of the multitrait multi-item correlation matrix demonstrated that item–scale correlations were higher (by two standard errors) for the hypothesized scale, corrected for overlap, than for the other scales.^{29, 30} (Please refer to Supporting information Table S2 for the results of the multitrait multi-item correlation matrix). These analyses supported our hypothesized item grouping and showed that distinct constructs were being measured by different disease-targeted scales.

Intercorrelation of the disease-targeted scales supported the validity of a single, higher order, disease-targeted ‘overall SF-LDQOL’ score. Thus, when administered as a hybrid instrument (with both generic and disease-targeted scales), the SF-LDQOL can be summarized into three summary measures, the PCS and the MCS from the SF-36 core and an overall SF-LDQOL score from the disease-targeted core. Whereas the SF-36 summary measures represent the general physical and mental health, the overall SF-LDQOL score summarizes the disease-targeted HRQOL in patients with advanced liver disease.

Reliability

Table 2 displays the number of disease-targeted items per scale, the mean scale scores, and Cronbach’s α from the baseline measurement in our cohort. The Cronbach’s α exceeded our threshold of 0.70 for the overall instrument and for seven of nine disease-targeted scales at the baseline administration and for all scales at the follow-up administration indicating that these scales performed well together as a composite measure. The Cronbach’s α values for the loneliness and hopelessness scales were 0.63 and 0.62 respectively. However, during follow-up administration of the SF-LDQOL, these increased to 0.80 and 0.70 respectively, thus providing support for the reliability of these two scales. As shown in Table 2, the Cronbach’s α for all SF-36 scales exceeded ≥ 0.70 at both the baseline and follow-up administrations.

Construct validity

Table 3 presents the supporting evidence for the construct validity of the SF-LDQOL. The direction and magnitude of the correlation coefficients for all comparisons were consistent with our *a priori* hypotheses.

Table 1. Baseline sample characteristics

Variable	<i>n</i>	Results
Age in year, mean (s.d.)	156	53.9 (11)
Gender, male (%)	85	54.8
Race (%)		
White	104	66.2
African American	9	5.7
Asian/Pacific Islander	7	4.5
Native American	2	1.3
Other	34	22.3
Hispanic ethnicity (%)	55	35.7
Marital status (%)		
Single	24	15.6
Married	102	66.2
Separated/divorced	23	14.9
Widowed	5	3.3
Education level (%)		
Eight grade or less	18	11.8
Some high school	19	12.5
High school diploma	26	17.1
Some college	48	31.5
Professional or graduate degree	26	17.1
Employment status (%)		
Working full time	20	12.7
Working part time	3	1.9
Unemployed	2	1.3
Retired	33	21.0
Disabled	76	48.4
Homemaker	11	7.0
Health insurance coverage (%)		
Medicare only	7	4.5
Medicare and supplemental	19	12.1
Medicaid only	28	17.8
Veterans administration	1	0.6
Private	8	5.1
HMO, PPO, IPA*	69	43.9
Other	15	9.5
None/not sure	6	3.8
Total household income (%)		
<\$5000	9	6.3
\$5000–\$10 000	17	11.9
\$10 001–\$25 000	30	20.9
\$25 001–\$50 000	25	17.5
\$50 001–\$75 001	24	16.8
>\$75 000	26	18.2
Not sure	12	8.4
Primary aetiology of liver disease (%)		
HCV	49	31.8
Alcohol	15	12.3
Alcohol and HCV	13	8.4
PBC/PSC	17	11.0
HBV	11	7.1
Cryptogenic	18	11.6
Other	21	13.6
MELD score, mean (s.d.)	155	18.5 (3.9)

HMO, health maintenance organization; PPO, preferred provider organization; IPA, independent practice association; HCV, chronic hepatitis C; PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis; HBV, chronic hepatitis B.

* These represent different models used by managed care organizations to deliver healthcare in the US.

Table 2. Number of items, mean scores and internal consistency reliability of the SF-LDQOL scales

Scales	Items	Cronbach's α	Mean	s.d.
<i>Disease-targeted scales</i>				
Overall score	36	0.91	55.3	18.8
Symptoms of liver disease	6	0.74	54.1	24.5
Effects of liver disease	3	0.73	49.8	28.9
Concentration/memory	4	0.90	57.3	24.4
Health distress	2	0.88	72.6	28.6
Sleep	5	0.73	41.5	20.6
Loneliness	5	0.63	73.7	20.0
Hopelessness	3	0.62	68.6	20.85
Stigma of liver disease	4	0.80	66.1	29.1
Sexual functioning/ problem	4	0.83	49.4	32.9
<i>Generic (SF-36) scales</i>				
PCS	35	0.93	30.8	10.8
MCS	35	0.90	41.0	10.8
Physical functioning	10	0.93	39.7	30
Role – physical	4	0.93	30.5	29.3
Bodily pain	2	0.90	45.2	28.5
General health	5	0.78	27.9	22.3
Vitality	4	0.78	34.6	21.4
Social functioning	2	0.74	44.4	28.5
Role – emotional	3	0.94	45.0	33.3
Mental health	5	0.79	57.5	20.7

Cronbach's α values for the overall score and for all individual scales were ≥ 0.70 during the follow-up administration of the SF-LDQOL. Loneliness and hopelessness scales during the second SF-LDQOL administration were 0.80 and 0.70 respectively.

PCS, physical component summary score of the SF-36; MCS, mental component summary score of the SF-36.

The SF-LDQOL overall score was statistically significantly associated with the SF-36 PCS ($r = 0.52$; $P < 0.0001$) and MCS ($r = 0.62$; $P < 0.0001$), patient-reported symptom severity ($r = -0.62$; $P < 0.0001$) and disability days ($r = -0.42$; $P < 0.0001$). The nine individual disease-targeted scale scores also demonstrated moderate correlations with these construct validity anchors. Using the data from the follow-up SF-LDQOL administration, we repeated the bivariate associations between the SF-LDQOL scores and previously described anchors and all relationships were consistent with the results from the baseline administration (Table 3). As hypothesized, MELD score demonstrated weak negative correlations with the HRQOL scales during both baseline and follow-up administrations (Table 3).

Table 3. Cross-sectional construct validation of the SF-1DQOL [correlation (*P*-value)]

	SF 36-PCS		SF 36-MCS		Symptom severity		Disability days		MELD score	
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
SF-1DQOL	0.52 (<0.0001)	0.53 (<0.0001)	0.68 (<0.0001)	0.80 (<0.0001)	-0.62 (<0.0001)	-0.60 (<0.0001)	-0.42 (<0.0001)	-0.41 (<0.0001)	-0.11 (0.17)	-0.35 (0.001)
Symptoms of liver disease	0.49 (<0.0001)	0.66 (<0.0001)	0.46 (<0.0001)	0.63 (<0.0001)	-0.49 (<0.0001)	-0.61 (<0.0001)	-0.38 (<0.0001)	-0.37 (0.0008)	-0.15 (0.14)	-0.31 (0.005)
Effects of liver disease	0.57 (<0.0001)	0.53 (<0.0001)	0.53 (<0.0001)	0.55 (<0.0001)	-0.55 (<0.0001)	-0.55 (<0.0001)	-0.41 (<0.0001)	-0.39 (0.0003)	-0.21 (0.009)	-0.12 (0.29)
Memory/concentration	0.33 (<0.0001)	0.28 (0.01)	0.49 (<0.0001)	0.53 (<0.0001)	-0.48 (<0.0001)	-0.40 (0.0002)	-0.33 (<0.0001)	-0.37 (0.0005)	0.003 (0.96)	-0.25 (0.02)
Sleep	0.44 (<0.0001)	0.52 (<0.0001)	0.47 (<0.0001)	0.63 (<0.0001)	-0.52 (<0.0001)	-0.43 (<0.0001)	-0.41 (<0.0001)	-0.30 (0.005)	-0.15 (0.05)	-0.30 (0.007)
Hopelessness	0.19 (0.02)	0.17 (0.04)	0.38 (<0.0001)	0.53 (<0.0001)	-0.14 (0.07)	-0.24 (0.02)	-0.25 (0.003)	-0.02 (0.83)	0.02 (0.80)	-0.23 (0.04)
Distress	0.40 (<0.0001)	0.47 (<0.0001)	0.67 (<0.0001)	0.75 (<0.0001)	-0.59 (<0.0001)	-0.43 (<0.0001)	0.37 (<0.0001)	-0.31 (0.0005)	-0.14 (0.09)	-0.26 (0.01)
Loneliness	0.14 (0.10)	0.25 (0.02)	0.31 (<0.0001)	0.58 (<0.0001)	-0.08 (0.2)	-0.25 (0.02)	-0.01 (0.8)	-0.20 (0.06)	-0.04 (0.5)	-0.16 (0.13)
Stigma of liver disease	0.30 (0.0004)	0.43 (0.0001)	0.43 (<0.0001)	0.58 (<0.0001)	-0.45 (<0.0001)	-0.48 (<0.0001)	-0.27 (0.001)	-0.31 (0.006)	0.02 (0.8)	-0.30 (0.007)
Sexual functioning/ problems	0.398 (0.002)	0.05 (0.7)	0.37 (0.002)	0.24 (0.16)	-0.38 (0.0009)	-0.26 (0.13)	-0.16 (0.2)	-0.3 (0.08)	-0.18 (0.12)	-0.27 (0.13)

PCS, physical component summary score of the SF-36; MCS, mental component summary score of the SF-36; MELD, model for end stage liver disease score.

Responsiveness

Twenty nine patients received an LT during study follow-up. Figure 1a displays the mean HRQOL score changes between baseline and follow-up administrations in patients with and without LT. Transplantation was associated with improvement in both overall and individual scale scores. In contrast, patients not receiving LT reported similar or worse HRQOL compared with the baseline HRQOL assessment (all $P > 0.05$).

Between-group comparisons showed significant differences in the mean HRQOL score changes between the transplanted patients and nontransplanted patients across the overall and eight of nine disease-targeted scales. Figure 1b displays the mean changes in generic SF-36 scale scores between the two HRQOL assessments in patients with and without LT.

The results of the multivariable regression analyses (Table 4) showed that for all scales, the corresponding baseline score was predictive of the magnitude of

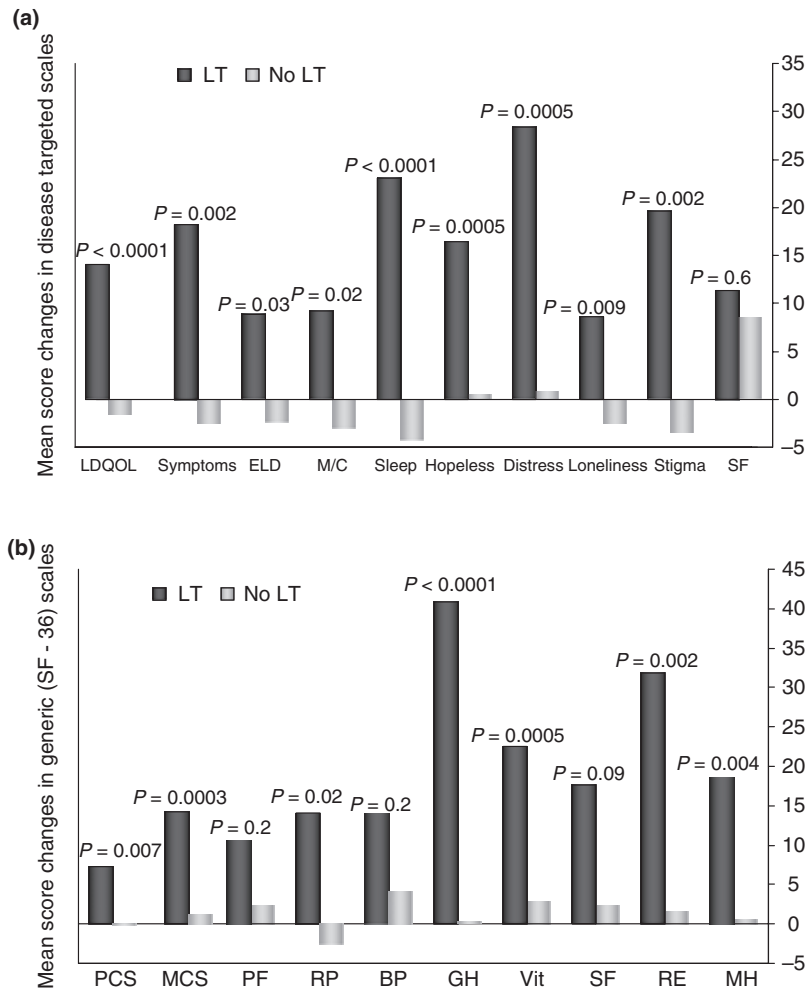


Figure 1. Comparison of mean score changes of SF-LDQOL between baseline and follow-up Administrations in patients who received vs. those who did not receive liver transplantation (LT). (a) and (b) The mean score changes in the disease-targeted and generic (SF-36) scale scores. Patients with LT reported improved HRQOL over time compared to patients without LT. Most of these differences were highly statistically significant. The differences in HRQOL persisted after adjusting for patients' age, gender, marital status, household income and baseline liver disease severity (i.e. MELD score) in multivariable models. LDQOL, overall score; Symptoms, symptoms of liver disease; ELD, effects of liver disease; M/C, memory and concentration; Hopeless, hopelessness; Stigma, stigma of liver disease; SF, sexual functioning and problems; PCS, physical component score; MCS, mental component score; PF, physical functioning; RP, role – physical; BP, bodily pain; GH, general health; Vit, vitality; SF, social functioning; RE, role – emotional; MH, mental health.

change in the HRQOL scores, with a lower baseline score predictive of a larger magnitude of improvement. Similarly, receipt of LT was highly associated with improvement in HRQOL across the overall SF-LDQOL and five of the nine disease-targeted scales. Between-group (LT vs. no LT) differences observed in the univariate analyses of HRQOL change scores were no longer significant for the symptoms and effects of liver disease scales after adjusting for the prespecified variables. Patient's socio-demographic characteristics and baseline liver disease severity did not predict HRQOL change over time.

Table 5 presents the responsiveness indices in patients who received LT. The results were comparable

Table 4. Multiple regression analyses for prediction of improvement in health-related quality of life

Disease-targeted scale	Variable(s)	Coefficient	P-value
Overall SF-LDQOL	Baseline score	-0.38	<0.0001
	LT	12.8	0.01
Symptoms of liver disease	Baseline score	-0.51	<0.0001
	LT	11.29	0.07
Effects of liver disease	Baseline score	-0.44	0.001
	LT	5.3	0.4
Memory/concentration	Baseline score	-0.37	0.001
	LT	13.9	0.05
Sleep	Baseline score	-0.41	0.002
	LT	26.1	<0.0001
Hopelessness	Baseline score	-0.51	<0.0001
	LT	20.0	0.007
Distress	Baseline score	-0.61	<0.0001
	LT	25.1	0.01
Loneliness	Baseline score	-0.46	<0.0001
	LT	17.7	0.004
Stigma of liver disease	Baseline score	-0.52	<0.0001
	LT	16.7	0.02
Sexual functioning/problems	Baseline score	-0.56	0.09
	LT	9.4	0.6

These data indicate that for all scales, the corresponding baseline score was predictive of the magnitude of change in the HRQOL scores, with a lower baseline score predictive of a larger magnitude of improvement. For example, for each point increase in the baseline SF-LDQOL, the HRQOL deteriorated by approximately 0.4 point between initial and follow-up administrations. Receipt of liver transplant (LT) was highly associated with improvement in HRQOL across the overall SF-LDQOL and five of the nine disease-targeted scales. For example, patients receiving LT during the study duration reported a 13-point increment in the SF-LDQOL score between the baseline and follow-up surveys.

across all three measures. The overall SF-LDQOL score, symptoms of liver disease, sleep, hopelessness, distress, stigma of liver disease, PCS, MCS, general health, vitality, role - emotional, mental health and social functioning scale scores showed medium ($ES \geq 0.5$) to large changes ($ES \geq 0.8$) indicating that these scales were highly responsive to change in patients' health status over time. The responsiveness statistic values for the effects of liver disease, memory/concentration, loneliness and sexual functioning/problems were 0.35, 0.43, 0.46 and 0.32 respectively.

Table 6 displays the correlation coefficients between the change in the SF-LDQOL score and the change in the SF-36 PCS, MCS and MELD scores. The change in the overall SF-LDQOL score showed highly statistically significant correlations with changes in SF-36

Table 5. Responsiveness of the SF-LDQOL to liver transplantation

Scale	Change score	Effect size	SRM*	RS
<i>Disease-targeted scales</i>				
LDQOL overall score	14.1	0.78	0.78	0.97
Symptoms of liver disease	18.2	0.45	0.42	0.51
Effects of liver disease	8.8	0.39	0.30	0.35
Memory/concentration	9.2	0.37	0.38	0.43
Sleep	23	1.12	0.93	1.17
Hopelessness	16.4	0.66	0.62	0.67
Distress	28.4	0.86	0.74	0.79
Loneliness	8.6	0.40	0.39	0.46
Stigma of liver disease	19.6	0.61	0.71	0.80
Sexual functioning/problems	11.3	0.31	0.33	0.32
<i>Generic (SF-36) scales</i>				
PCS	5.16	0.50	0.56	0.63
MCS	11.6	1.06	0.97	1.2
Physical functioning	10.4	0.27	0.42	0.43
Role - physical	14.0	0.45	0.55	0.57
Bodily pain	13.9	0.48	0.45	0.49
General health	40.8	1.8	1.6	2.3
Vitality	22.5	1.0	1.0	1.14
Social functioning	17.6	0.61	0.54	0.59
Role - emotional	31.8	0.95	0.87	0.94
Mental health	18.5	0.89	0.78	0.81

SRM, standardized responsiveness mean; RS, responsiveness statistic; PCS, physical component summary score of the SF-36; MCS, mental component summary score of the SF-36. * We used s.d. of change score for nontransplanted patients in calculating the SRM.

Table 6. Correlation of the SF-LDQOL change scores with change scores of key anchors

Scale	MELD	SF-PCS	SF-MCS
LDQOL summary score	-0.30 (0.007)	0.65 (<0.0001)	0.41 (0.0007)
Symptoms of liver disease	-0.48 (<0.0001)	0.45 (0.0002)	0.52 (<0.0001)
Effects of liver disease	-0.09 (0.4)	0.38 (0.002)	0.35 (0.005)
Memory/concentration	-0.16 (0.14)	0.08 (0.5)	0.42 (0.0005)
Sleep	-0.37 (0.001)	0.37 (0.01)	0.61 (<0.0001)
Hopelessness	-0.006 (0.95)	0.27 (0.02)	0.30 (0.01)
Distress	-0.26 (0.02)	0.31 (0.01)	0.66 (<0.0001)
Loneliness	-0.14 (0.22)	0.35 (0.005)	0.44 (0.0004)
Stigma of liver disease	-0.17 (0.14)	0.22 (0.07)	0.40 (0.001)
Sexual functioning/problems	-0.32 (0.13)	0.30 (0.21)	0.10 (0.60)

MELD, model for end stage liver disease score; PCS, physical component summary score of the SF-36; MCS, mental component summary score of the SF-36.

PCS ($r = 0.65$, $P < 0.0001$) and SF-36 MCS ($r = 0.41$, $P = 0.0007$). In addition, there was a significant correlation between changes in the overall SF-LDQOL score and MELD score over time ($r = -0.30$, $P = 0.007$) indicating that patients' HRQOL declined with worsening liver disease severity. Changes in the individual scale scores also demonstrated moderate correlations with patient self-reported anchor score changes and most of the relationships were highly statistically significant. Symptoms of liver disease showed moderate correlation, whereas sleep and health distress change scores showed weak correlations with changes in patients' MELD scores over time.

As statistical significance does not always correlate with clinical relevance, it is important to determine whether a given HRQOL difference between groups is clinically important to patients themselves.³⁵ In an exploratory analysis, we used the patient-reported change in overall health status between the baseline and follow-up SF-LDQOL administrations to estimate the minimally clinically important difference (MCID) for the overall SF-LDQOL score. We defined MCID as the mean change in HRQOL for the subset of patients who reported a minimal yet perceptible change in health (i.e. reported hardly, little, or somewhat changed health between the baseline and follow-up HRQOL administration). Only 16 patients reported minimal yet perceptible change in overall health between baseline and follow-up administrations. Given the small sample size, we could not reliably estimate the MCID for the subset of patients with improved health ($n = 6$). The mean change in SF-LDQOL score, i.e. MCID, for

patients who reported worse health ($n = 10$) was 5.1 points (ES = 0.28).

DISCUSSION

We found that disease-targeted HRQOL can be accurately measured in advanced liver disease using a short-form measure (the SF-LDQOL) both as a single concept (i.e. using the 'overall score') and as nine individual domains. The SF-LDQOL disease-targeted scales explained greater than 90% of the variation in the corresponding LDQOL 1.0. scale scores. The internal consistency reliability coefficients of the SF-LDQOL disease-targeted scales were excellent with seven of nine scales at baseline and all scales at follow-up administration exceeding the prespecified threshold of ≥ 0.70 for the Cronbach's α values. Multitrait scaling analyses revealed that the items' correlation was higher for the hypothesized scales than for other scales, thus providing support of the convergent and discriminant validity of the included items. The direction and magnitude of the correlation coefficients between HRQOL and concurrently measured clinical anchors were consistent with our *a priori* hypotheses. Our study also found that the SF-LDQOL is able to detect clinically meaningful changes in patients with advanced liver disease over time. The SF-LDQOL scores declined significantly with the worsening of patients' self-reported health (r with PCS = 0.65, $P < 0.0001$; r with MCS = 0.40, $P = 0.0007$) and with worsening MELD score over time ($r = -0.30$, $P = 0.007$). Although consistent with our *a priori* hypothesis, the

magnitude of the correlation between HRQOL and MELD score changes was small. Moreover, change in MELD scores explained only 9% of the variation in HRQOL changes over time (data not shown). These data reaffirm the knowledge that biological disease severity indices fail to capture patients' perceptions fully and thus highlight the need to complement objective measures of disease severity with HRQOL assessments to assess health status accurately and comprehensively.^{4, 20-22} As hypothesized, we found marked improvements in HRQOL in patients undergoing LT vs. those not receiving transplantation across a wide range of HRQOL domains and these data are consistent with previously published reports.^{7, 22, 32, 33} Indeed, in our multivariable models, with the exception of baseline HRQOL, receipt of LT was the only other factor that consistently predicted an improvement in HRQOL in our patient cohort. Combined, these results provide strong support for the reliability and validity (including responsiveness) of the SF-LDQOL.

Previous studies evaluating changes in HRQOL before vs. after LT suggest that generic HRQOL measures may indeed be able to detect some HRQOL differences after LT.³⁶ Despite these data, we believe that our study provides incrementally useful information for the following reasons. First, contrary to the generic HRQOL measures, our outcome measure is directly relevant to patients with advanced chronic liver disease. The liver disease-targeted items and scales in the SF-LDQOL were derived from disease-specific perceptions and concerns elicited during patient focus groups and cognitive interviews.²² The SF-LDQOL, therefore, captures the spectrum of HRQOL in advanced liver disease (e.g. stigma of liver disease, symptoms and effects of liver disease, difficulty with sleep, etc.) not otherwise measurable by generic instruments alone. Second, the SF-LDQOL is a hybrid instrument. This hybrid structure further maximizes the measurement spectrum of the SF-LDQOL,²¹ both in terms of the breadth (with SF-36 component) and depth (with disease-targeted component) of HRQOL information compared with other available HRQOL measures.³⁷ Prospective validation of the SF-LDQOL therefore significantly contributes to the available armamentarium of relevant and clinically useful HRQOL measures for patients with advanced liver disease. Third, because the SF-36 serves as a generic core of the SF-LDQOL, our study provides the most comprehensive data to date demonstrating responsiveness of the SF-36 to change in the health status pre- vs. post-LT. Fourth, the SF-LDQOL is rela-

tively short and when administered without the SF-36 takes an average of 12 min to complete. These data are important because although several multidimensional scales have been developed to measure HRQOL in medicine, their clinical utility is limited by the respondent's burden associated with their routine clinical use. In summary, the prospective validation of the SF-LDQOL provides incremental data beyond the available literature. Augmentation of the generic with disease-targeted scales in the SF-LDQOL comprehensively assesses HRQOL without compromising the clinical usefulness and feasibility of these assessments in patients with advanced chronic liver disease.

Given its psychometric properties, the SF-LDQOL may provide researchers with a novel outcome measure for prospective clinical trials in advanced liver disease. With the increasing availability of new treatment alternatives for cirrhosis and related clinical complications (e.g. rifaximin vs. lactulose for management of hepatic encephalopathy; transjugular intrahepatic portosystemic shunts vs. large volume paracentesis for management of ascites; and anti-viral treatments vs. conservative management for viral hepatitis in cirrhosis),³⁸⁻⁴⁰ it is important to establish the most effective therapeutic alternative for patients with advanced liver disease. For reasons of the uncertainty regarding how best to use the available agents in this growing patient population, data from studies using HRQOL outcomes may assist in clinical decision making. The SF-LDQOL can now fulfil the need for a reliable and responsive HRQOL outcome measure in clinical trials among patients with cirrhosis. In addition to providing researchers with an outcome measure for clinical trials, the SF-LDQOL may allow physicians to monitor better, patient outcomes in clinical practice. The HRQOL measures should meet the minimum standard for reliability (≥ 0.90) before they can be used for individual patient level assessment.²⁵ As the Cronbach's α for the overall disease-targeted component (the SF-LDQOL overall score) and for several SF-36 scales exceeded this minimum standard of 0.90, the SF-LDQOL may now be used to assess changes in individual patients' HRQOL in daily clinical practice.²⁵

Our study provides a preliminary insight into the clinical significance (in contrast to the statistical significance) of the SF-LDQOL score differences. Specifically, we found that a difference of 5.1 points may be perceived as minimally clinically important by the patients. Given the limited sample size of patient reporting minimal yet perceptible change in health, our MCID

estimates may not be robust. Despite this limitation, our estimate (ES = 0.28) is consistent with published reports of minimally important differences in HRQOL in other areas of medicine.^{41, 42} The consistency of MCID estimates across various studies provides convergent validity to our finding and suggests that our preliminary estimate of the MCID in the overall SF-LDQOL score may be useful in everyday clinical practice and in clinical trials to evaluate the effectiveness of care.

In addition to the prospective evaluation of HRQOL, our study has several strengths. First, we evaluated the instrument in a large sample size with sufficient number of 'changed' vs. 'stable' patients, thereby minimizing the probability of type-II error. Second, we measured construct validity across a wide range of clinically relevant anchors, including the SF-36, self-reported disability days, self-reported symptom severity and MELD scores. Third, in addition to measuring instrument validity and responsiveness, we ensured that the SF-LDQOL is reliable in terms of its structure as measured by Cronbach's α . Fourth, with both generic and disease-targeted components, the hybrid structure of the SF-LDQOL maximizes the spectrum of HRQOL captured with the SF-LDQOL both in terms of the breadth (with SF-36 component) and depth (with disease-targeted component) of HRQOL information.¹⁹ The disease-targeted SF-LDQOL scales do not overlap the domains covered by the SF-36 (e.g. physical functioning, social functioning, general health perceptions, etc.), but instead expand more deeply into the domains hypothesized to be important to patient with advanced liver disease (e.g. symptoms of liver disease, effects of liver disease, liver-related health distress, etc). Thus, this approach allows for measurement of different aspect of health in patients with advanced liver disease. Future use of the SF-LDQOL as a hybrid instrument vs. as a combination of various components/scales will depend on the goal of the evaluation. For example, in a clinical trial evaluating a new treatment for hepatic encephalopathy, the investigators may decide to use concentration/memory, sleep and MCS scales as their primary outcomes. Similarly, in a trial of anti-viral treatment in patients with advanced liver disease, the overall disease-targeted SF-LDQOL score with or without general health and vitality scales of SF-36 may be the relevant outcomes. In clinical practice, the hybrid instrument may be useful in screening individual patients comparing the relative burden of liver disease with that of other chronic diseases, and in differentiating the health benefits

produced by a wide range of disparate treatments in patients with advanced liver disease.

Our study has several limitations. First, the sample population was limited to English-speaking patients and therefore cannot be generalized to non-English speaking patients. Second, the sample consisted of ambulatory patients with stable health insurance coverage who were referred for LT, limiting the generalizability of our findings to non-insured patients with poor access to care and to patients not referred for LT. Similarly, the SF-LDQOL was designed to measure the HRQOL of patients with advanced chronic liver disease. As a result, only patients with documented advanced liver disease (i.e. cirrhosis) were included in this study. It is plausible that some of the disease-targeted domains (such as effects of liver disease, concentration/memory, health distress, etc.) may be relevant to patients with compensated noncirrhotic liver disease. However, in the absence of properly designed studies including patients with different stages of liver disease, the validity and reliability of the SF-LDQOL in assessing HRQOL of patients with less severe liver disease cannot be supported. Third, because of the small sample size of patients who completed the sexual functioning/problems items, our responsiveness estimate for this scale may not be sufficiently robust. The responsiveness values for sexual functioning/problems were low (RS = 0.32). However, this may represent lack of significant change in sexual functioning/problems scale scores with LT. Specifically, although we observed wide between-group differences in eight of nine scale scores, changes in scores in the sexual functioning/problem scale were not significantly different between transplanted patients and nontransplanted patients. These data are consistent with published reports suggesting that transplantation fails to improve sexual dysfunction in patients with advanced liver disease.⁴³ In addition to its limited responsiveness, the sexual functioning/problem scale applies only to the subset of population with recent sexual activity, i.e. within the previous 4-week and this may explain the proportion of missing data on these items. Given these considerations, future research will need to address whether this scale should remain in the SF-LDQOL. Fourth, our MCID estimate will also need to be confirmed using a larger sample size and longer duration of patient follow-up.

In conclusion, our data provide evidence that the SF-LDQOL is a reliable and valid measure of HRQOL in patients with advanced liver disease. When administered without the SF-36 core, it takes on average

12 min to complete and provides rich information across a wide range of physical, psychological and social health domains targeted to liver disease patients. The SF-LDQOL is sensitive to change in health status over time and may prove useful for measuring HRQOL in every day clinical practice and in clinical trials.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

Table S1. Results of maximum *R*-squared stepwise regression analyses¹ and reliability testing

Table S2. Multitrait multi-item correlation matrix

Table S3. Disease-targeted items in the short form of liver disease quality of life instrument (the SF-LDQOL)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting material supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article

ACKNOWLEDGEMENTS

Declaration of personal interests: Dr Kanwal has received research funding from Gilead Pharmaceuticals. Dr Spiegel has served as a consultant for AstraZeneca, TAP Pharmaceuticals, Phynova and Takeda, and has received research funding from Amgen, AstraZeneca, TAP Pharmaceuticals, Takeda, Salix. Dr Bolus has served as a consultant for the Rome Foundation and Pfizer, and owns stocks in Pfizer. Dr Saab has served as a consultant and speaker for Schering-Plough, Roche, Three-Rivers, Gilead, BMS and Vertex, owns stocks in Schering-Plough and Vertex, and has received grant support from Schering-Plough, Roche, Three-Rivers, Gilead and BMS. Dr Durazo has served on the advisory board for McNeil Laboratories. Dr Gralnek, Dr Han and Mr Kim have no disclosures to declare. *Declaration of funding interests:* This study was funded by the American College of Gastroenterology Clinical Research Award 2003–2004.

REFERENCES

- Dufour MC. Chronic liver disease and cirrhosis. In: Everhart JE, ed. *Digestive Diseases in the United States: Epidemiology and Impact*. Washington, DC: US Government Printing Office, 1994: 613–46.
- Kim WR, Brown RS Jr, Terrault NA, El-Serag H. Burden of liver disease in the United States: summary of a workshop. *Hepatology* 2002; **36**: 227–42.
- Anonymous. *The Burden of Gastrointestinal Diseases*. Bethesda, MD: The American Gastroenterological Association, 2001: 41–60.
- Kanwal F, Hays RD, Kilbourne AM, Dulai GS, Gralnek IM. Are physician-derived disease severity indices associated with health-related quality of life in patients with end-stage liver disease? *Am J Gastroenterol* 2004; **99**: 1726–32.
- Marchesini G, Bianchi G, Amodio P, *et al.* Factors associated with poor health-related quality of life of patients with cirrhosis. *Gastroenterology* 2001; **120**: 170–8.
- Steel JL, Chopra K, Olek MC, Carr BI. Health-related quality of life: hepatocellular carcinoma, chronic liver disease, and the general population. *Qual Life Res* 2007; **16**: 203–15.
- Younossi ZM, McCormick M, Price LL, *et al.* Impact of liver transplantation on health-related quality of life. *Liver Transpl* 2000; **6**: 779–83.
- Younossi ZM, Kiwi ML, Boparai N, Price LL, Guyatt G. Cholestatic liver diseases and health-related quality of life. *Am J Gastroenterol* 2000; **95**: 497–502.
- Younossi ZM, Boparai N, Price LL, Kiwi ML, McCormick M, Guyatt G. Health-related quality of life in chronic liver disease: the impact of type and severity of disease. *Am J Gastroenterol* 2001; **96**: 2199–205.
- Younossi ZM. Chronic liver disease and health-related quality of life. *Gastroenterology* 2001; **120**: 305–7.
- Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; **118**: 622–9.
- Devinsky O, Vickrey BG, Cramer J, *et al.* Development of the quality of life in epilepsy inventory. *Epilepsia* 1995; **36**: 1089–104.
- Hofer S, Lim L, Guyatt G, Oldridge N. The MacNew Heart Disease health-related quality of life instrument: a summary. *Health Qual Life Outcomes* 2004; **2**: 3.
- Litwin MS, Hays RD, Fink A, Ganz PA, Leake B, Brook RH. The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure. *Med Care* 1998; **36**: 1002–12.
- Puhan MA, Behnke M, Frey M, *et al.* Self-administration and interviewer-administration of the German Chronic Respiratory Questionnaire: instrument development and assessment of validity and reliability in two randomised studies. *Health Qual Life Outcomes* 2004; **2**: 1.
- Wu AW, Huang IC, Gifford AL, Spritzer KL, Bozzette SA, Hays RD. Creating a crosswalk to estimate AIDS Clinical Trials Group quality of life scores in a nationally representative sample of persons in care for HIV in the United States. *HIV Clin Trials* 2005; **6**: 147–57.
- Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987; **40**: 171–8.
- Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985; **38**: 27–36.
- Atherly A. Condition-specific measures. In: Kane R, ed. *Understanding Health Care Outcomes Research*. Sudbury, MA, USA:

- Jones and Bartlett Publishers 2004: 165–184.
- 20 Kessler RC, Mroczek DK. Measuring the effects of medical interventions. *Med Care* 1995; 4(Suppl.): AS109–19.
 - 21 Saab S, Ibrahim AB, Shpaner A, *et al.* MELD fails to measure quality of life in liver transplant candidates. *Liver Transpl* 2005; 11: 218–23.
 - 22 Gralnek IM, Hays RD, Kilbourne A, *et al.* Development and evaluation of the Liver Disease Quality of Life instrument in persons with advanced, chronic liver disease – the LDQOL 1.0. *Am J Gastroenterol* 2000; 95: 3552–65.
 - 23 Ware JE, Snow KK, Kosinski MK, Gandek B. Scoring the SF-36. In: Ware JE, ed. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston, MA: Nimrod Press, 1992: 6.
 - 24 Tubbs JD. *Maximum R2 Improvement (MAXR)*. Available at: http://www3.baylor.edu/~Jack_Tubbs/Courses/Regression/Notes/regression.pdf. Accessed on 15 June 2008.
 - 25 Hays RD, Anderson RT, Revicki D. Assessing reliability and validity of measurement in clinical trials. In: Staquet MJ, Hays RD, Fayers PM, eds. *Quality of Life Assessment in Clinical Trials: Methods and Practice*. Oxford: Oxford University Press, 1998: 169–82.
 - 26 Litwin MS. Scaling and scoring. In: Litwin MS, ed. *How to Measure Survey Reliability and Validity*. Thousand Oaks, CA, USA: SAGE Publications Inc., 1995: 47–52.
 - 27 Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL, *et al.* A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001; 33: 464–70.
 - 28 Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, *et al.* Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology* 2003; 124: 91–6.
 - 29 Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959; 56: 81–105.
 - 30 Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. *International Quality of Life Assessment. J Clin Epidemiol* 1998; 51: 945–52.
 - 31 Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989; 42: 403–8.
 - 32 Belle SH, Porayko MK, Hoofnagle JH, Lake JR, Zetterman RK. Changes in quality of life after liver transplantation among adults. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Liver Transplantation Database (LTD). *Liver Transpl Surg* 1997; 3: 93–104.
 - 33 Gross CR, Malinchoc M, Kim WR, Evans RW, Wiesner RH, Petz JL, *et al.* Quality of life before and after liver transplantation for cholestatic liver disease. *Hepatology* 1999; 29: 356–64.
 - 34 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. London: Academic Press, 1969.
 - 35 Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it?. *Pharmacoeconomics* 2000; 18: 419–23.
 - 36 Tome S, Wells JT, Said A, Lucey MR. Quality of life after liver transplantation. A systematic review. *J Hepatol* 2008; 48: 567–77.
 - 37 Younossi ZM, Guyatt G, Kiwi M, Boparai N, King D. Development of a disease specific questionnaire to measure health related quality of life in patients with chronic liver disease. *Gut* 1999; 45: 295–300.
 - 38 Mas A, Rodes J, Sunyer L, *et al.* Comparison of rifaxamin and lactitol in the treatment of acute hepatic encephalopathy: results of a randomized, double-blind, double-dummy, controlled trial. *J Hepatol* 2003; 38: 51–8.
 - 39 Salerno F, Cammà C, Enea M, *et al.* Transjugular intrahepatic portosystemic shunt for refractory ascites: a meta-analysis of individual patient data. *Gastroenterology* 2007; 133: 825–34.
 - 40 Di Marco V, Almasio PL, Ferraro D, *et al.* Peg-interferon alone or combined with ribavirin in HCV cirrhosis with portal hypertension: a randomized controlled trial. *J Hepatol* 2007; 47: 484–91.
 - 41 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.
 - 42 Steel JL, Eton DT, Cella D, Olek MC, Carr BI. Clinically meaningful changes in health-related quality of life in patients diagnosed with hepatobiliary carcinoma. *Ann Oncol* 2006; 17: 304–12.
 - 43 Sorrell JH, Brown JR. Sexual functioning in patients with end-stage liver disease before and after transplantation. *Liver Transpl* 2006; 12: 1473–7.