

# Psychometric validation of the SF-36 Health Survey in light chain amyloidosis: results from community-based and clinic-based samples

Michelle K White<sup>1</sup>  
 Kristen L McCausland<sup>1</sup>  
 Vaishali Sanchorawala<sup>2</sup>  
 Spencer D Guthrie<sup>3</sup>  
 Martha S Bayliss<sup>1</sup>

<sup>1</sup>Optum, Johnston, RI, <sup>2</sup>Department of Medicine, Section of Hematology/Oncology, Amyloidosis Center, Boston University School of Medicine, Boston, MA, <sup>3</sup>Medical Affairs, Prothena Biosciences Inc, South San Francisco, CA, USA

**Background:** Light chain (AL) amyloidosis, a rare and life-threatening protein misfolding disorder, causes organ damage and severely impacts health-related quality of life (HRQoL). No patient-reported outcome (PRO) HRQoL measure has been validated for use in an AL amyloidosis patient population, leaving a gap for researchers conducting observational studies and clinical trials for drug development. The SF-36 Health Survey (SF-36) has been the most frequently used PRO in AL amyloidosis studies to date, and early qualitative validation studies support its use in this population. The aim of this study was to assess the psychometric properties of the SF-36 among patients with AL amyloidosis.

**Methods:** Data from community-based (n=341) and clinic-based (n=1,438) observational studies were used to document the psychometric properties of the SF-36 in this disease population. Reliability was estimated using internal consistency (Cronbach's alpha) and test-retest reliability (intraclass correlation). Convergent validity, known-groups validity, and the ability to detect change were assessed with available criterion variables.

**Results:** Scale reliability (Cronbach's alpha  $\geq 0.780$  for all scores) and test-retest reliability (intraclass correlation coefficients  $\geq 0.731$  for all) were acceptable. Scale convergent validity was supported by strong correlations with conceptually related measures. Mean SF-36 scores varied by response to treatment ( $P < 0.05$  for all scores) and a self-reported measure of disease severity ( $P < 0.001$  for all scores). Data indicate that the SF-36 is sensitive to changes in other measures over time.

**Conclusion:** This study provided clear and consistent evidence of the psychometric properties of the SF-36 in both community-based and clinic-based samples of patients with AL amyloidosis.

**Keywords:** patient-reported outcomes, quality of life, psychometric properties

## Background

Light chain (AL) amyloidosis is a rare and life-threatening protein misfolding disorder which causes organ damage and severely impacts health-related quality of life (HRQoL).<sup>1-4</sup> AL amyloidosis can impact any organ in the body but most commonly affects the heart and kidneys.<sup>5</sup> The clinical features of the disease are highly dependent on the type and number of organs involved in a particular patient.

To date, no patient-reported outcome (PRO) HRQoL measure has been validated for use in an AL amyloidosis patient population. This leaves a gap for researchers conducting observational studies and clinical trials for drug development, as well as for physicians who seek to understand the HRQoL of their patients over time. Given the heterogeneity of the disease and the nonspecific symptoms that are experienced by patients, broad generic health status measures are best suited to capture the impact of the condition on how patients feel and function. The SF-36 Health Survey (SF-36)

Correspondence: Kristen L McCausland  
 Optum, 1301 Atwood Avenue, Johnston, RI 02919, USA  
 Tel +1 401 642 9220  
 Email kmccausland@qualitymetric.com

is currently the most used generic PRO in AL amyloidosis patient studies,<sup>3,4</sup> and early qualitative validation studies support its use in this population.<sup>6,7</sup>

Until now, there has been no published evidence to support the psychometric validity of the SF-36 specifically in the AL amyloidosis patient population, although it has been shown to be a reliable and valid measure in numerous disease areas.<sup>8,9</sup> The aim of this study is to describe the psychometric properties of the SF-36 in AL amyloidosis. We provide evidence from two large samples of patients with AL amyloidosis: one community-based sample and one clinic-based sample. This approach ensures that our findings are generalizable to patients with a broad range of demographic and disease characteristics. Although the clinic-based study administered an earlier version of the SF-36 measure (ie, SF-36v1), the instrument is very similar; therefore, the results are relevant and supportive of psychometric properties based on the SF-36v2<sup>®</sup>. To simplify the exposition of results, we will refer to the measure as the “SF-36” regardless of the version used.

## Methods

### Sample/study procedures

#### Community-based sample (Study #1; n=341)

The community-based sample of patients with AL amyloidosis (Study #1) was obtained from the AL Amyloidosis Patient Health-Related Quality of Life Study (ClinicalTrials.gov; NCT02574676). This longitudinal, noninterventional, observational study was specifically designed to examine HRQoL within a diverse sample of patients with AL amyloidosis in the US with varied demographics, and disease characteristics, including organ involvement, duration of disease, disease severity, and treatment history. Online recruitment occurred between October and December 2015 through two amyloidosis patient advocacy groups: the Amyloidosis Support Groups and the Amyloidosis Foundation. The recruitment strategy included IRB-approved announcements on websites, social media sites, and in membership emails. The announcements provided a hyperlink for an electronic informed consent form and an online study screener. Individuals were eligible to participate if they 1) were at least aged 18 years; 2) had a self-reported diagnosis (made by a physician) of AL amyloidosis; and 3) were interested and able to complete four online surveys administered in English over a 12-month period. Eligible individuals were immediately directed to the online baseline survey.

Online surveys were administered at baseline, as well as at 1, 6, and 12 months following the initial baseline data collection. A variety of measures, including HRQoL (as measured by the SF-36 and other PROs) and sociodemographic,

disease, diagnosis, and treatment history characteristics, were collected at each time point. This study was reviewed and approved by the New England Institutional Review Board (IRB# 15-355).

#### Clinic-based sample (Study #2; n=1,438)

Data from the clinic-based sample (Study #2) were extracted from a cohort of patients with AL amyloidosis seen at the Amyloidosis Center of Boston University School of Medicine at Boston Medical Center (BMC) between 1994 and 2014 (n=1,822). Information regarding patient sociodemographic characteristics, disease characteristics, and HRQoL was extracted from medical records. Patients completed the SF-36 during clinic visits, but there was no standard protocol for the frequency of administration of the measure. Of the 1,822 patients, 79% of patients (n=1,438) completed the SF-36 at their initial evaluation. The BMC IRB determined that this retrospective review of medical records did not meet the definition of “human subject research” and, therefore, did not require consent. All procedures performed in these studies involving human participants were in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

## Measures

### Validation measure: SF-36

The SF-36 is a 36-item, generic measure of HRQoL. Thirty-five of the 36 items are used to calculate scores representing eight dimensions of functioning and well-being: physical functioning (PF), role physical (RP; role limitations due to physical problems), bodily pain (BP), general health (GH) perceptions, vitality (VT), social functioning (SF), role emotional (RE; role limitations due to emotional problems), and mental health (MH), and two summary scores, such as physical component summary (PCS) and mental component summary (MCS). All eight scale scores and the two summary scores are derived using a norm-based scoring strategy that yields standardized distributions with a mean of 50 and a standard deviation of 10 for a nationally representative sample of US adults. SF-36 data were scored using the developer’s software using the norm-based scoring method.<sup>9</sup> Missing items were given person-specific estimates using the developer’s missing data estimation algorithm.<sup>9</sup> Higher SF-36 scores represent better health for all scores.

### Criterion measures

Several other HRQoL PROs were administered concurrently with the SF-36 and were used as criterion measures

to evaluate the validity of the SF-36. The following PROs served as criterion measures in Study #1: a numeric rating of pain; a global assessment of functioning; the Patient Global Impression – Severity (PGI-S) scale;<sup>10</sup> the Patient Global Assessment of Change (PGAC); the Work Productivity and Activity Impairment Questionnaire: Specific Health Problem (WPAI) scale;<sup>11</sup> and the Kansas City Cardiomyopathy Questionnaire (KCCQ-12) scale.<sup>12</sup> Furthermore, patients in Study #1 were also asked to describe their current response to treatment status from their most recent physician visit. In Study #2, the Southwest Oncology Group (SWOG) performance status measure was assessed at each clinic visit. Table 1 provides further description of all PROs used as criterion measures in the psychometric analyses, in terms of recall period, scoring, and content.

## Statistical analyses

Cross-sectional and longitudinal analyses were conducted to test specific validation hypotheses in each sample. In Study #1, cross-sectional data from the baseline survey (n=341) were used to examine internal consistency, convergent validity, and known-groups validity, whereas data from the 1- and 6-month follow-up surveys were analyzed to examine test-retest reliability and ability to detect change, respectively. For Study #2, all patients who completed the SF-36 at their initial evaluation were included in the psychometric analysis of internal consistency reliability (n=1,438). Additional complete case analytic samples were created for each type of validation test based on the availability of data from the SF-36 and appropriate criterion measures from the same time point. A list of psychometric tests by sample and time point is provided in Table 2.

**Table 1** Summary of criterion measures

PRO name	Description of items/domains	Recall period	Scoring	Data source
Numeric pain rating	Single item rating of pain	Past week	<ul style="list-style-type: none"> <li>• Range: 0–10</li> <li>• Direction: higher scores represent worse pain</li> </ul>	Study #1
Global assessment of functioning	Single item of functioning	Nonspecific	<ul style="list-style-type: none"> <li>• Range: 0–100</li> <li>• Direction: higher scores represent better functioning</li> </ul>	Study #1
WPAI questionnaire: specific health problem scales	<p>Only administered to a subsample of patients who were currently employed</p> <p>Three subscales:</p> <ul style="list-style-type: none"> <li>• Absenteeism: % work missed due to disease</li> <li>• Presenteeism: % time spent at work that is impacted by disease</li> <li>• Overall work productivity loss: % work productivity loss (accounts for both absenteeism and presenteeism)</li> </ul>	Past 7 days	<ul style="list-style-type: none"> <li>• Range: 0–100 (presented as percentages)</li> <li>• Direction: higher scores represent worse outcomes</li> </ul>	Study #1
KCCQ-12	<p>Twelve-item disease-specific measure of HRQoL</p> <p>Four subscales: physical limitation, symptoms, quality of life, and social limitation</p> <p>Only administered to a subsample of patients with cardiac involvement</p>	Past 2 weeks	<ul style="list-style-type: none"> <li>• Range: 0–100 (presented as percentages)</li> <li>• Direction: higher scores represent better functioning</li> </ul>	Study #1
PGI-S	Single item to measure patient's perceived severity of disease	Past month	<ul style="list-style-type: none"> <li>• Ordinal scale: not severe at all, mild, moderate, severe, very severe</li> </ul>	Study #1
PGAC	Single item to gauge and measure patient's perceived change in severity	Since last survey	<ul style="list-style-type: none"> <li>• Ordinal scale: very much improved, somewhat improved, no change, somewhat worse, very much worse</li> </ul>	Study #1
Recent hematologic response	Single item to gauge recent response to treatment	Current status	<ul style="list-style-type: none"> <li>• Ordinal scale: no response, partial response, complete remission, or hematologic response</li> </ul>	Study #1
SWOG performance status	Single item gauging the percent of day spent in bed	Nonspecific	<ul style="list-style-type: none"> <li>• Ordinal scale: 0%, 25%, 50%, 75%, and 100%</li> </ul>	Study #2

**Notes:** Study #1, community-based sample; Study #2, clinic-based sample.

**Abbreviations:** HRQoL, health-related quality of life; KCCQ-12, Kansas City Cardiomyopathy Questionnaire; PGAC, Patient Global Assessment of Change; PGI-S, Patient Global Impression – Severity; PRO, patient-reported outcome; SWOG, Southwest Oncology Group; WPAI, Work Productivity and Activity Impairment.

**Table 2** Summary of the psychometric validation approaches for the SF-36 in an AL amyloidosis population

Psychometric property	Analysis/indicator	Data source	Criterion measure
<b>Reliability</b>			
Test–retest reliability	ICCs $\geq 0.70$	Study #1: patients with stable disease between baseline and 1-month follow-up (n=179) <sup>a</sup>	N/A
Internal consistency reliability	Cronbach's alpha $\geq 0.70$	Study #1: baseline sample Study #2: baseline sample	N/A
<b>Validity</b>			
Scale convergent validity	Pearson's moment-product correlation $\geq 0.40$	Study #1: baseline sample	Numeric pain rating; global assessment of functioning; WPAI; KCCQ-12
Discriminant validity (known-groups approach)	ANOVA tests	Study #1: baseline sample Study #2: baseline sample	Response to treatment; PGI-S SWOG – performance status
<b>Responsiveness</b>			
Ability to detect change	Paired t-tests	Study #1: 6-month follow-up sample (change from baseline) Study #2: follow-up sample (change from baseline)	PGAC SWOG – performance status

**Notes:** <sup>a</sup>A 1-month interval was chosen for the assessment of test–retest reliability because: 1) it was long enough so that patients were unlikely to remember their previous responses and/or be affected by survey response fatigue and 2) it was short enough that substantial changes in health status were unlikely. To help ensure that no substantial change in health status occurred during that timeframe, we restricted this analysis to patients who reported their disease had been stable on the PGAC survey item. Study #1, community-based sample; Study #2, clinic-based sample.

**Abbreviations:** AL, light chain; ANOVA, analysis of variance; KCCQ-12, Kansas City Cardiomyopathy Questionnaire; ICC, intraclass correlation coefficient; N/A, not applicable; PGAC, Patient Global Assessment of Change; PGI-S, Patient Global Impression – Severity; SF-36, SF-36 Health Survey; SWOG, Southwest Oncology Group; WPAI, Work Productivity and Activity Impairment.

## Data quality

Data quality was evaluated by examining out-of-range values and missing data. The SF-36 Response Consistency Index (RCI) was calculated for each patient to gauge how he/she responded to 15 paired items with specific hypothesized relationships indicating a logically consistent pattern of responses.<sup>9</sup> Inconsistent responses within an item pair are scored as 1 per pair, therefore the range of RCI scores is 0 (no inconsistent pairs) to 15, all pairs answered inconsistently.

## Reliability

The reliability of the SF-36 scales was tested using internal consistency and test–retest reliability methods. Data from both Study #1 and Study #2 were used to examine internal consistency reliability. Cronbach's alphas were calculated for each scale to evaluate the extent to which each item in a scale measures the same underlying construct. A threshold of 0.70 was also used to identify scales with good internal consistency.<sup>13</sup>

Data from Study #1 were used to assess test–retest reliability based on intraclass correlation coefficients (ICCs).<sup>14</sup> ICCs are an indicator of how consistently an instrument measures a construct over time when no change would otherwise be expected. In this study, SF-36 scores were obtained from patients at two consecutive time points ~1 month apart. ICCs were calculated among a subset of patients who indicated that there had not been any changes in their disease severity based

on their responses to the PGAC at the 1 month follow-up. A generally accepted cutoff of 0.70 was used to identify scales with good test–retest reliability.<sup>13</sup>

## Convergent validity

To assess convergent validity, Pearson correlation coefficients were calculated between SF-36 scores and other PROs measuring similar concepts based on baseline data from Study #1. We hypothesized that the following SF-36 scores and similar PROs should correlate at least moderately ( $r > 0.30$ ): 1) BP and the numeric pain rating (negative correlation); 2) RP and all four WPAI scales (negative correlations among a subset of patients who were currently employed); 3) GH and global functioning rating (positive correlation); 4) PF and KCCQ-12 Physical Limitation (positive correlation); 5) PCS and KCCQ-12 Physical Limitation (positive correlation); 6) SF and KCCQ-12 social limitation (positive correlation); 7) MCS and KCCQ-12 social limitation (positive correlation); 8) PCS and KCCQ-12 quality of life (positive correlation); and 9) MCS and KCCQ-12 quality of life (positive correlation). Correlations between SF-36 scores and KCCQ-12 scores were examined in the subset of patients with self-reported cardiac involvement.

## Discriminant validity (known-groups approach)

Analysis of variance was conducted to test for significant differences in mean scores across groups known to vary in

disease severity. For Study #1, differences were examined by patients' most recent self-reported response to treatment and by their responses to the PGI-S. For Study #2, differences were examined by the SWOG performance status measure. Given the high number of tests conducted for discriminant validity, we present both raw *P*-values as well as Hommel-adjusted *P*-values to adjust for multiple comparisons.

### Ability to detect change

The ability to detect change, or responsiveness, of the SF-36 was tested in both Study #1 and Study #2 samples by examining whether SF-36 scores were sensitive to changes in other PRO scores.

First, we examined whether SF-36 scores from Study #1 were sensitive to self-reported changes in patients' condition. Based on responses to the PGAC during the 6-month follow-up survey, patients were classified as "improved" (*n*=70) if they described their condition as "very much improved" or "somewhat improved", and they were classified as "worsened" (*n*=47) if they described their condition as "somewhat worse" or "very much worse".

Based on data from Study #2, we examined whether SF-36 scores were responsive to changes in SWOG performance status over time. Among patients with baseline values for both the SF-36 and performance status, 201 patients also provided responses to both measures again at a follow-up visit of at least 3 months, but <1 year after their baseline observation. Patients were classified as "improved" (*n*=43) if they reported spending less time in bed as compared to baseline, and they were classified as "worsened" if they reported an increase in the proportion of time spent in bed (*n*=36).

For both studies, paired *t*-tests were used to compare the baseline and follow-up values of HRQoL within groups of patients classified as improved or worsened for each criterion variable. Hommel-adjusted *P*-values were used to control for multiple comparisons.

## Results

Patient characteristics for the two study samples are reported in Table 3 (Study #1: *n*=341; Study #2: *n*=1,438). The mean age of patients in each study was comparable (60.6 years [Study #1] and 61.6 years [Study #2]). The majority of patients in both studies were white, had at least a college degree, and were married. More patients in Study #1 were females compared with Study #2 (52.9 vs 35.9) and in terms of highest level of educational attainment, a smaller proportion of patients in Study #1 had earned at most a high school diploma (8.1% vs 26.6% completed ≤ a high school

diploma) as compared to patients in Study #2. With regard to clinical characteristics, the kidneys and the heart were the most commonly affected organs in both studies. There was more heterogeneity in terms of the duration of disease among patients in Study #1 as compared to Study #2. The time since diagnosis ranged from 1 month to 28 years among patients in Study #1 (median =3.5 years) vs 0 days to 15.5 years among patients in Study #2 (median =2.4 months).

### Data quality

Overall, the data in both studies were of good quality, with few patients providing inconsistent responses. Based on the RCI, 94.1% of patients in Study #1 and 99.9% of patients in Study #2 had perfect RCI scores, indicating consistent responses across all 15-item pairs. The online data collection method used in Study #1 did not allow for missing item responses; therefore, no additional considerations regarding missing data were warranted within this sample. In Study #2, 90.2% of the sample had complete SF-36 data. After the missing data estimation algorithm was implemented, scores were obtained for 99.0% of the sample.

### Reliability

Internal-consistency reliability estimates for each scale exceeded 0.70, as measured by Cronbach's alpha in both study samples (Table 4). Cronbach's alphas in Study #1 ranged from 0.78 for the GH scale to 0.97 for the RP scale. In Study #2, the Cronbach's alphas ranged from 0.76 for GH to 0.95 for PF.

In Study #1, ~71% of patients who completed the 1 month follow-up survey met the criteria of stable disease and were included in assessments of test-retest reliability. Based on data from these disease-stable patients (*n*=179), all SF-36 scales demonstrated good test-retest reliability, with ICCs exceeding the acceptable threshold of 0.70 on all scales, and ranging from 0.73 for the SF scale to 0.86 for the GH scale (Table 4).

### Convergent validity

As shown in Table 5, the correlation coefficients between the SF-36 scores and those obtained from other PROs in Study #1 were all significantly different from zero and in the hypothesized directions. Overall, correlations exceeded the conventional threshold of 0.30 for all of the hypothesized relationships.

As shown in Table 5, the SF-36 BP scale was strongly correlated with the numeric pain rating scale ( $r=-0.82$ ). The correlations between the SF-36 RP scale and all four

**Table 3** Baseline sociodemographic and disease characteristics by study sample

Characteristic	Study #1, community-based sample (n=341)		Study #2, clinic-based sample (n=1,438)	
	n	%	n	%
Age, mean years (SD)	60.6 (10.2)		61.6 (10.8)	
Age range (median)	23–85 (61)		27–101 (62)	
Gender				
Male	160	47.1	870	60.5
Female	180	52.9	568	39.5
Race/ethnicity				
White	304	89.1	1,275	88.7
Others	37	10.9	162	11.3
Education				
≤ High school diploma or GED	26	8.1	319	26.6
Some college, associate degree, or technical certificate	99	30.7	212	17.7
Bachelor's degree (BA, BSc)	109	33.9	306	25.6
Graduate degree	88	27.3	360	30.1
Marital status				
Married	271	82.1	1,043	79.0
Others	59	17.9	274	21.0
Employment status				
Currently employed for pay	115	38.3		N/A
Time since diagnosis, mean years (SD)	4.5 (4.0)		0.6 (1.3)	
Time since diagnosis range (median)	1 month to 28 years (3.5 years)		0 days to 15.5 years (2.4 months)	
Types of organs/systems impacted <sup>a</sup>				
Heart (cardiac)	178	52.2	777	54.0
Kidney	214	62.8	1,038	72.2
Liver	49	14.4	389	27.1
Nervous system	126	37.0	432	30.0
Gastrointestinal	148	43.4	466	32.4
Number of organs involved				
One	95	27.9	47	3.3
Two	89	26.1	346	24.3
Three or more	157	46.0	1,029	71.6
Most recent hematologic response status				N/A
No response to treatment	23	6.7		
Partial hematologic response or partial remission	126	37.0		
Complete hematologic response or complete remission	141	41.3		
I do not know	51	15.0		

**Note:** <sup>a</sup>Multiple responses allowed.

**Abbreviations:** BA, Bachelors of Art; BSc, Bachelors of Science; GED, general educational development; N/A, not applicable; SD, standard deviation.

**Table 4** Reliability of the SF-36 scales

Reliability	SF-36 scale							
	PF	RP	BP	GH	VT	SF	RE	MH
Test–retest reliability								
ICC <sup>a</sup>	0.85	0.77	0.78	0.86	0.77	0.73	0.76	0.78
Internal consistency								
Study #1 <sup>b</sup> – Cronbach's alpha	0.93	0.97	0.91	0.78	0.87	0.90	0.95	0.88
Study #2 <sup>c</sup> – Cronbach's alpha	0.95	0.91	0.90	0.76	0.88	0.87	0.89	0.82

**Notes:** <sup>a</sup>Calculated among the subsample of patients in the community-based sample (Study #1) who reported “no change” on the Patient Global Assessment of Change item on the 1-month follow-up survey (n=179). <sup>b</sup>Community-based sample of AL amyloidosis patients (n=341). <sup>c</sup>Clinic-based sample of AL amyloidosis patients (n=1,438).

**Abbreviations:** AL, light chain; BP, bodily pain; GH, general health; ICC, intraclass correlation coefficient; MH, mental health; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; SF-36, SF-36 Health Survey; VT, vitality.

WPAI scales were moderate to strong (all were  $<-0.50$ ) and in the expected direction. The data supported a strong relationship between the SF-36 PF and SF scales with similar

KCCQ-12 scales (physical limitation and social limitation, respectively;  $r=0.73$  for both). Finally, both SF-36 summary scores demonstrated significant correlations with KCCQ-12

scales (PCS with physical limitations [ $r=0.65$ ] and quality of life [ $r=0.61$ ]; MCS with social limitations [ $r=0.58$ ] and quality of life [ $r=0.51$ ]).

## Discriminant validity

Based on data from Study #1, evidence of the SF-36's ability to discriminate among groups of patients with AL amyloidosis based on their most recent self-reported hematologic response and their PGI-S rating is reported in Tables 6 and 7,

**Table 5** Convergent validity in Study #1: Pearson correlation coefficients between SF-36 scores and criterion measures

SF-36 scale/ component summary	Criterion measure	N	r
BP	Numeric pain rating	341	-0.82
RP	WPAI <sup>a</sup> : absenteeism	108	-0.45
RP	WPAI <sup>a</sup> : presenteeism	94	-0.67
RP	WPAI <sup>a</sup> : work productivity	92	-0.68
RP	WPAI: activity impairment	341	-0.72
GH	Global functioning rating	341	0.60
PF	KCCQ-12 <sup>b</sup> : physical limitation	153	0.73
PCS	KCCQ-12 <sup>b</sup> : physical limitation	153	0.65
SF	KCCQ-12 <sup>b</sup> : social limitation	153	0.73
MCS	KCCQ-12 <sup>b</sup> : social limitation	153	0.58
PCS	KCCQ-12 <sup>b</sup> : quality of life	153	0.61
MCS	KCCQ-12 <sup>b</sup> : quality of life	153	0.51

**Notes:** All correlations are significant ( $P<0.001$ ). <sup>a</sup>Administered to a subsample of study participants who are currently employed. <sup>b</sup>Administered to a subsample of study participants with self-reported cardiac involvement. *r*, Pearson correlation.

**Abbreviations:** BP, bodily pain; GH, general health; KCCQ-12, Kansas City Cardiomyopathy Questionnaire; MCS, mental component summary; PCS, physical component summary; PF, physical functioning; RP, role physical; SF, social functioning; SF-36, SF-36 Health Survey; WPAI, Work Productivity and Activity Impairment.

respectively. As shown in Table 6, mean SF-36 scores across hematologic response groups were significantly different (adjusted  $P$ -values  $<0.01$  for all). As anticipated, the lowest scores were observed among patients who had not responded to treatment and more favorable scores were observed among patients who reported a complete hematologic response or remission. Similar patterns were also observed for the PGI-S responses as shown in Table 7. Based on these data, the lowest scores were observed among patients who described their condition as "severe" or "very severe" and greater among those who reported their condition as "not severe at all".

Among the 1,438 patients from Study #2 who completed the SF-36 at their initial evaluation, ~61% ( $n=884$ ) also had data for SWOG performance status from the same time point. SF-36 mean scores were examined by level of SWOG performance status to determine if the mean SF-36 scores varied by the proportion of time spent in bed. As expected, mean SF-36 scores were inversely associated with the amount of time spent in bed (Table 8).

## Ability to detect change

Among the 225 patients who completed the 6-month follow-up survey in Study #1, ~52% ( $n=117$ ) reported a change in their condition since baseline data collection and were included in the assessment of the SF-36's ability to detect change. Of those, 60% reported an improvement in their condition and 40% reported their condition as worsened. SF-36 scores changed positively among patients who described their condition as improved and declined among those whose condition reportedly worsened, as illustrated

**Table 6** Known-groups validity test: SF-36 scores by recent hematologic response (Study #1)

SF-36 scale/ score	Most recent hematologic response			Raw $P$ -value <sup>a</sup>	Adjusted $P$ -value <sup>b</sup>	Do not know ( $n=51$ )
	No response to treatment ( $n=23$ )	Partial hematologic response ( $n=126$ )	Complete hematologic response/ remission ( $n=141$ )			
SF-36 norm-based scales						
PF	35.0	41.1	45.6	<0.001	<0.001	40.9
RP	33.3	37.5	44.3	<0.001	<0.001	40.6
BP	38.6	44.5	49.2	<0.001	<0.001	47.9
GH	33.2	35.1	43.3	<0.001	<0.001	41.4
VT	40.7	41.7	47.9	<0.001	<0.001	43.6
SF	36.2	42.2	46.8	<0.001	<0.001	41.6
RE	40.3	43.8	47.6	0.005	0.009	45.0
MH	44.7	47.7	51.4	0.002	0.005	48.2
SF-36 summary scores						
PCS	32.8	37.9	44.3	<0.001	<0.001	41.3
MCS	44.2	46.5	50.2	0.009	0.009	46.8

**Notes:** <sup>a</sup> $P$ -values for the ANOVA tests were based on the following groups: no response to treatment; partial hematologic response; and complete hematologic response/remission. The ANOVA tests did not include the "do not know" category. <sup>b</sup>Adjusted for multiple comparisons using Hommel's method.

**Abbreviations:** BP, bodily pain; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; SF-36, SF-36 Health Survey; VT, vitality.

**Table 7** Known-groups validity test: SF-36 scores by PGI-S scale (Study #1)

SF-36 scale/score	PGI-S scale				Raw, P-value	Adjusted, P-value <sup>a</sup>
	Not severe at all (n=133)	Mild (n=92)	Moderate (n=81)	Severe/very severe (n=35)		
SF-36 norm-based scales						
PF	48.9	41.7	38.4	30.1	<0.001	<0.001
RP	47.3	40.7	34.5	27.8	<0.001	<0.001
BP	51.8	47.0	43.5	32.8	<0.001	<0.001
GH	45.1	39.2	34.6	27.9	<0.001	<0.001
VT	49.7	44.1	40.9	33.8	<0.001	<0.001
SF	49.8	44.5	37.8	31.1	<0.001	<0.001
RE	50.7	46.7	40.1	33.4	<0.001	<0.001
MH	53.5	50.3	45.7	37.0	<0.001	<0.001
SF-36 summary scores						
PCS	47.0	40.0	36.4	28.5	<0.001	<0.001
MCS	52.4	49.3	43.7	36.9	<0.001	<0.001

**Note:** <sup>a</sup>Adjusted for multiple comparisons using Hommel's method.

**Abbreviations:** BP, bodily pain; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; PGI-S, Patient Global Impression – Severity; RE, role emotional; RP, role physical; SF, social functioning; SF-36, SF-36 Health Survey; VT, vitality.

**Table 8** Known-groups validity test: SF-36 scores by SWOG performance status (Study #2)

SF-36 scale/score	Performance status (% of time spend in bed)				Raw P-value	Adjusted P-value <sup>a</sup>
	0% (n=70)	25% (n=471)	50% (n=213)	≥75% (n=131)		
SF-36 norm-based scales						
PF	48.4	37.4	26.1	21.6	<0.001	<0.001
RP	48.2	36.6	30.8	29.0	<0.001	<0.001
BP	53.4	48.7	43.3	41.2	<0.001	<0.001
GH	46.6	40.7	35.7	33.0	<0.001	<0.001
VT	49.5	41.4	34.4	33.2	<0.001	<0.001
SF	47.8	40.4	32.3	27.9	<0.001	<0.001
RE	49.4	43.3	37.4	34.8	<0.001	<0.001
MH	49.4	46.9	43.5	43.3	<0.001	<0.001
SF-36 summary scores						
PCS	49.1	39.0	31.0	27.6	<0.001	<0.001
MCS	48.9	45.8	41.7	40.5	<0.001	<0.001

**Note:** <sup>a</sup>Adjusted for multiple comparisons using Hommel's method.

**Abbreviations:** BP, bodily pain; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; SF-36, SF-36 Health Survey; SWOG, Southwest Oncology Group; VT, vitality.

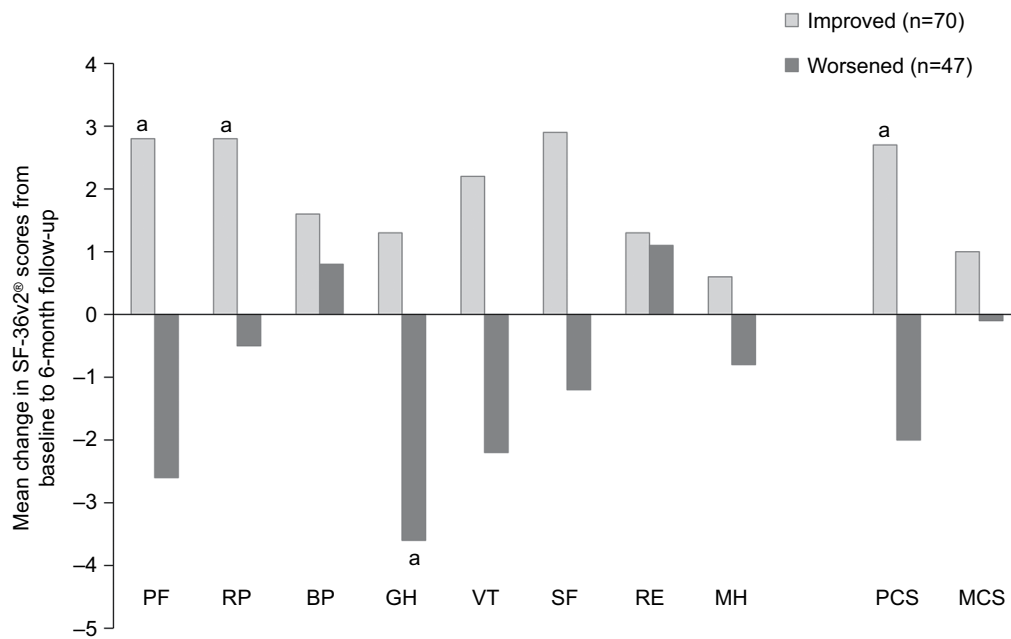
in Figure 1. Among patients whose condition improved, the positive changes in PF, RP, and PCS were significant. Among patients who described their condition as worse, there was a significant decline in GH.

As shown in Figure 2, SF-36 scores were also related to changes in patients' SWOG performance status based on data from Study #2. Positive mean changes were observed among the subgroup of patients who were classified as improved based on their changes in performance status. These positive changes were significant for PF, RP, VT, SF, and RE, as well as for the PCS and the MCS ( $P \leq 0.01$  for all, adjusted for multiple comparisons). Significant declines in PF were observed among patients who were classified as worsened ( $P < 0.05$ , adjusted for multiple comparisons).

## Discussion

This study provides clear and consistent evidence of the psychometric properties of the SF-36 in patients with AL amyloidosis. All SF-36 scores met the minimum standards of internal consistency reliability and test-retest reliability. Correlations between SF-36 scores and criterion measures also met minimum values to support convergent validity. The strongest correlations were observed between SF-36 scores and PROs designed to measure the same construct (ie, the SF-36 BP scale and a numeric pain rating); whereas, correlations between SF-36 scales and PROs that measured similar, but slightly different constructs also were sufficient. Significant differences in SF-36 scores by levels of disease severity helped to illustrate discriminant validity using a

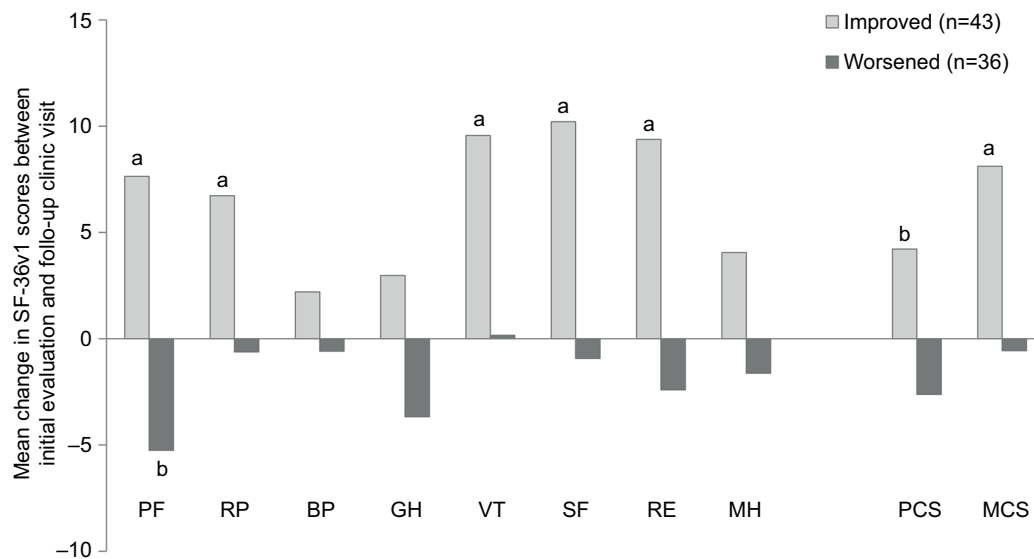




**Figure 1** Change in SF-36v2 scores by self-reported patient global assessment of change in AL amyloidosis patients.

**Notes:** Data represent responses from patients in the community-based sample (Study #1; n=117). A total of 107 patients endorsed "no change" on the PGAC, data not shown. <sup>a</sup>P<0.05, adjusted for multiple comparisons.

**Abbreviations:** AL, light chain; BP, bodily pain; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; PGAC, Patient Global Assessment of Change; RE, role emotional; RP, role physical; SF, social functioning; VT, vitality.



**Figure 2** Change in SF-36 scores by SWOG performance status in AL amyloidosis patients.

**Notes:** Data represent responses from patients in the clinic-based sample (Study #2; n=201). A total of 122 patients had "no change" in their SWOG performance status, data not shown. <sup>a</sup>P<0.01, adjusted for multiple comparisons. <sup>b</sup>P<0.05, adjusted for multiple comparisons.

**Abbreviations:** AL, light chain; BP, bodily pain; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; SF-36, SF-36 Health Survey; SWOG, Southwest Oncology Group; VT, vitality.

known-groups approach. It is important to note that self-reported criterion measures of disease severity were used to evaluate discriminant validity. Future research should confirm these results using biomarker and/or clinician-reported outcomes, such as N-terminal pro B-type natriuretic peptide and the New York Heart Association Functional Classification,

both of which are used to help stage cardiac involvement in patients with AL amyloidosis.

AL amyloidosis is a rare condition with an estimated incidence of 8–12 cases per million person-years.<sup>15,16</sup> Obtaining sufficient representative data to adequately demonstrate the psychometric properties of an HRQoL instrument in such

a rare disease can be challenging. We addressed this in two ways: 1) the use of web-based purposive sampling in partnership with patient advocacy groups for the community-based sample and 2) working with a Center of Excellence to analyze 20 years of patient records. The availability of HRQoL data from patients with AL amyloidosis seen at the clinic during a 20-year timeframe translates into a large sample for such a rare disease and is a true strength.

The two-sample approach allowed us to study a relatively large number of patients and ensure diversity in our sample demographics and clinical characteristics. Despite some educational differences, the demographic characteristics were fairly similar in both groups, which may indicate that these two samples collectively represent a broader spectrum of patients ranging from the newly diagnosed to longer term survivors and representing a larger continuum of disease severity. To that end, the inclusion of both samples in this study strengthened the generalizability and robustness of the study findings.

Despite these strengths, our use of SF-36v1 and SF-36v2 data is not ideal; however, given AL amyloidosis is a rare condition, the advantages of this approach outweighed the limited weaknesses. Specifically, the SF-36v2 data from the community-based study were used to assess psychometric properties in patients with AL amyloidosis. The SF-36v1 data from the clinic-based sample were supplementary and were used to confirm known-groups validity and responsiveness in a sample of patients with clinician-confirmed diagnoses and more recently diagnosed patients – areas where the community-based sample had limitations. The clinic-based study offered >20 years of data, which were seen as a true asset, particularly for a rare disease. It is important to note that the SF-36v2 was designed as an improvement over the SF-36v1; however, it maintains the purpose and content of the SF-36v1, making it comparable and relevant to this validation study.

There are many advantages to using a generic measure of HRQoL, such as the SF-36, for AL amyloidosis. First, generic measures allow comparisons to general population and disease norms that can provide context for lesser known diseases.<sup>17,18</sup> Second, given its wide use, many stakeholders are familiar with the measure and its interpretations. Third, the heterogeneity of organ involvement and symptomology of this disease can result in a wide variety of disease experiences and impacts on patients' quality of life. Based on these results, a variety of stakeholders, ranging from health care providers, patients, payers, developers of new drugs, and regulatory agencies, may be able to use this measure to elicit and better understand HRQoL among patients receiving treatment for AL amyloidosis.

This is the first study to report the quantitative psychometric properties of a HRQoL measure specifically intended for use in an AL amyloidosis disease population. These findings extend those from qualitative studies that describe the content validity of the SF-36 in AL amyloidosis.<sup>6–7</sup> Collectively, these qualitative and quantitative results demonstrate the appropriateness of using the SF-36 to assess HRQoL among patients with AL amyloidosis.

## Acknowledgments

We gratefully acknowledge the numerous colleagues in the Amyloidosis Center, Cancer Clinical Trials Office, and Center for Cancer and Blood Disorders at BMC who assisted with the multidisciplinary evaluation and treatment of the patients with AL amyloidosis in the clinic-based study. We thank the patients for their participation in this study and the Amyloidosis Support Group and the Amyloidosis Foundation for their help in recruitment of the community-based sample. We also thank Colin Oliver and Tiffany Quock for their review of the analyses and manuscript drafts. This study was funded by Prothena Biosciences Inc. Portions of this work were presented in poster format at the 21st Congress of the European Hematology Association, June 9–12, 2016, Copenhagen, Denmark, and at the International Society of Amyloidosis 2016 Symposium, July 3–7, 2016, Uppsala, Sweden.

## Author contributions

All authors contributed toward data analysis, drafting and critically revising the paper and agree to be accountable for all aspects of the work.

## Disclosure

At the time of the original submission, MKW, MSB, and KLM were full-time employees of Optum, Inc, which publishes the SF-36v2, and received research funding from Prothena Biosciences Inc. to conduct the study. SDG was a full-time employee of the study sponsor, Prothena Biosciences Inc. at the time of the original submission. The authors report no other conflicts of interest in this work.

## References

1. Gertz MA. Immunoglobulin light chain amyloidosis: 2013 update on diagnosis, prognosis, and treatment. *Am J Hematol*. 2013;88(5):416–425.
2. Smorti M, Cappelli F, Bergesio F, Perfetto F. Anxiety and depression among AL amyloidosis patients: the role of cardiac symptoms. *Amyloid*. 2012;19(3):123–128.
3. Seldin DC, Anderson JJ, Sanchowala V, et al. Improvement in quality of life of patients with AL amyloidosis treated with high-dose melphalan and autologous stem cell transplantation. *Blood*. 2004;104(6):1888–1893.

4. Caccialanza R, Palladini G, Klersy C, et al. Nutritional status independently affects quality of life of patients with systemic immunoglobulin light-chain (AL) amyloidosis. *Ann Hematol.* 2012;91(3):399–406.
5. Merlini G, Palladini G. Light chain amyloidosis: the heart of the problem. *Haematologica.* 2013;98(10):1492–1495.
6. Lin HM, Seldin D, Hui A, Berg D, Dietrich CN, Flood E. The patient's perspective on the symptom and everyday life impact of AL amyloidosis. *Amyloid.* 2015;22(4):244–251.
7. White MK, Bayliss M, Guthrie SD, Raymond KP, Rizio AA, McCausland KL. Content validation of the SF-36v2® Health Survey with AL amyloidosis patients. *Journal of Patient Reported Outcomes.* In press 2017.
8. Ware JE Jr. SF-36 health survey update. *Spine.* 2000;25(24):3130–3139.
9. Maruish ME [webpage on the Internet]. *User's Manual for the SF-36v2 Health Survey.* 3rd ed. Lincoln, RI: QualityMetric, Inc; 2011. Available from: <https://campaign.optum.com/optum-outcomes.html>. Accessed June 15, 2016.
10. Guy W; National Institute of Mental Health (U.S.) [webpage on the Internet]. *Early Clinical Drug Evaluation Program. ECDEU Assessment Manual for Psychopharmacology.* Rev. Rockville, MD: U.S. Dept. of Health Education and Welfare Public Health Service Alcohol Drug Abuse and Mental Health Administration National Institute of Mental Health Psychopharmacology Research Branch; 1976. Available from: <http://worldcatlibraries.org/wcpa/oclc/2344751>. Accessed October 27, 2017.
11. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics.* 1993;4(5):353–365.
12. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. *Circ Cardiovasc Qual Outcomes.* 2015;8(5):469–476.
13. Nunnally JC, Bernstein IH. *Psychometric Theory.* 3rd ed. New York: McGraw-Hill; 1994.
14. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420–428.
15. Kyle RA, Linos A, Beard CM, et al. Incidence and natural history of primary systemic amyloidosis in Olmsted County, Minnesota, 1950 through 1989. *Blood.* 1992;79(7):1817–1822.
16. Pinney JH, Smith CJ, Taube JB, et al. Systemic amyloidosis in England: an epidemiological study. *Br J Haematol.* 2013;161(4):525–532.
17. Bosch AM, Burlina A, Cunningham A, et al. Assessment of the impact of phenylketonuria and its treatment on quality of life of patients and parents from seven European countries. *Orphanet J Rare Dis.* 2015; 10:80.
18. Hendriksz CJ, Berger KI, Lampe C, et al. Health-related quality of life in mucopolysaccharidosis: looking beyond biomedical issues. *Orphanet J Rare Dis.* 2016;11(1):119.

## Patient Related Outcome Measures

### Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

Dovepress

The journal is included in PubMed. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.