Psychometric Validation of the SF-36v2® Health Survey in an AL Amyloidosis Population

Michelle K. White,[1] Martha Bayliss,[1] Kristen L. McCausland,[1] Spencer D. Guthrie[2]

[1]Optum, Lincoln, Rhode Island, United States; [2]Prothena Biosciences Inc, South San Francisco, California, United States

E1441

BACKGROUND

- Amyloid light chain (AL) amyloidosis, a rare protein-misfolding disease, leads to deficits in health-related quality of life (HRQoL).[1] Patients with AL amyloidosis have a wide variety of nonspecific symptoms, organ involvement, and functional impairment[2]
- There is no disease-specific measure of HRQoL for AL amyloidosis. Given the heterogeneity in disease characteristics and symptom burden, a general measure of HRQoL in AL amyloidosis may be an efficient strategy to estimate HRQoL
- The SF-36v2® Health Survey (SF-36v2; Optum), a general HRQoL measure, is the most frequently used patient-reported outcome (PRO) endpoint in clinical trials[3]
- The SF-36v2 has been used to quantify the impact of AL amyloidosis on HRQoL,[1] but to date there is no evidence of its psychometric validity for use in patients with AL amyloidosis

OBJECTIVE

- To document the psychometric properties of the SF-36v2—including tests of data quality, scaling success, reliability, and validity—among patients with AL amyloidosis

METHODS

Patient Sample and Study Design

- Adults (≥18 years of age) with self-reported AL amyloidosis completed online surveys to assess HRQoL and clinical and sociodemographic characteristics
- Online recruitment strategies were implemented to promote the study opportunity through AL amyloidosis patient advocacy group websites, social media sites, and email lists
- Data from the initial (N = 341) and 1-month follow-up (n = 251) surveys, collected in 2015-2016, were used in these analyses

Measure: SF-36v2

- The SF-36v2 is a 36-item, self-report survey that measures patient HRQoL across 8 different dimensions of functioning and well-being[4]
  - Physical functioning (PF)
  - Role physical (RP)
  - Bodily pain (BP)
  - General health (GH)
  - Vitality (VT)
  - Social functioning (SF)
  - Role emotional (RE)
  - Mental health (MH)
- Scores on all domains are used to calculate summary measures for overall physical and mental health
  - Physical Component Summary (PCS)
  - Mental Component Summary (MCS)
- The standard (4-week) recall form in US English was used in the present study
- All scales and summary scores were normalized to 50 ± 10 (mean ± SD) and then adjusted to the distribution of scores observed in the US population
- Based on norm-based scores, a mean of 50 is equal to the average score observed in the general population. Higher scores represent better health and functioning

Criterion Measures for HRQoL Domains

- Numeric rating of pain in the past week: 0-10 scale, where 10 represents worst pain
- Global assessment of functioning rating: 0-100 scale, where 100 represents best possible functioning
- Work Productivity and Activity Impairment Questionnaire (WPAI): Specific Health Problem[5] scales
  - Absenteeism, presenteeism, overall work productivity loss, and activity impairment
  - Recall period: past 7 days
  - Score range and direction: 0% to 100% (higher score represents worse outcome)
- Kansas City Cardiomyopathy Questionnaire (KCCQ-12)[6]
  - Physical limitation, symptoms, quality of life, and social limitation
  - Recall period: past 2 weeks
  - Score range and direction: 0 to 100 (higher score represents better functioning)

Psychometric Analysis

Data Quality

- Item and scale distributions and summary measures were evaluated against assumptions of summated rating scales (ie, approximately comparable means and standard deviations)
- A response consistency index (RCI) was calculated for each patient to gauge how he or she responded to 15 paired items one would expect a patient to answer similarly
  - Consistent pair score = 0; inconsistent pair score = 1
  - The final score is the sum across all 15 pairs of items
- The online system did not allow out-of-range values or missing data

Reliability

- Internal consistency reliability: Cronbach's α was calculated for each domain to measure the extent to which each item in a domain measured the same underlying construct
- Test–retest reliability: Intraclass correlation coefficients (ICCs) between initial and 1-month follow-up scores were calculated among patients with stable disease (those who reported "no change" on the Patient Global Assessment of Change survey item at the 1-month follow-up [n = 179])

Convergent Validity

- Pearson correlation coefficients were calculated between scores from the SF-36v2 and other PROs measuring similar concepts. We hypothesized that there would be significant correlations between
  - BP and numeric pain rating
  - RP and all 4 WPAI scales
  - GH and global functioning rating
  - PF, PCS, and KCCQ-12 Physical Limitation
  - SF, MCS, and KCCQ-12 Social Limitation
  - PCS, MCS, and KCCQ-12 Quality of Life

Discriminant Validity (known-groups approach)

- Analysis of variance (ANOVA) was conducted to test for significant differences in mean scores across groups known to vary in disease severity. Differences were examined by
  - Most recent hematologic response status
  - Responses to the Patient Global Impression–Severity (PGI-S) scale[7]
- Hommel-adjusted *P* values were used to control for multiple comparisons

RESULTS

Data Quality

- Assessment of the data showed excellent response distribution; summated rating scale assumptions were satisfactory
- More than 94% of patients had RCIs of zero, indicating that they responded to similar items as expected (exceeded conventional threshold of 90%)

Reliability

- Indicators were well above the conventional thresholds to support adequate internal consistency. Cronbach's α for each of the 8 domains was >0.70 (**Table 1**)
- SF-36v2 scores were stable over time among a subsample of patients with stable disease. All ICCs were >0.70 for all domains and scales (**Table 1**)

Table 1. Scale Reliability

SF-36v2 Scale	No. of Items	Cronbach's α	ICC
Physical functioning	10	0.93	0.85
Role physical	4	0.97	0.77
Bodily pain	2	0.91	0.78
General health	5	0.78	0.86
Vitality	4	0.87	0.77
Social functioning	2	0.90	0.73
Role emotional	3	0.95	0.76
Mental health	5	0.88	0.78
Physical Component Summary	—	—	0.84
Mental Component Summary	—	—	0.81

ICC, intraclass correlations for test–retest reliability. Cronbach's α was used to assess internal consistency of scales. ICCs were calculated using data from baseline and 1-month follow-up visits among a subset of patients who completed each survey and reported "no change" on the Patient Global Assessment of Change item at 1-month.

Convergent Validity

- SF-36v2 scores correlated well with conceptually related measures; all hypothesized relationships had correlations of ≥0.40 (**Table 2**)

Table 2. Convergent Validity

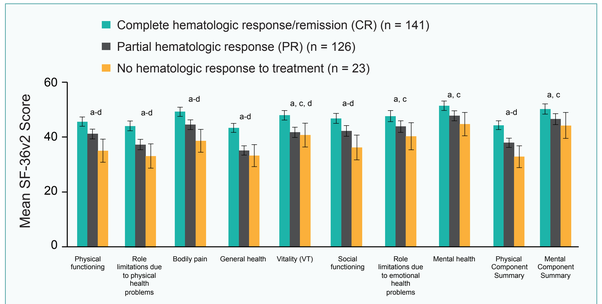
SF-36v2 Scale Comparison	Pearson Correlation Coefficient
BP with numeric pain rating	–0.82
RP with WPAI[a]: absenteeism	–0.45
RP with WPAI[a]: presenteeism	–0.67
RP with WPAI[a]: work productivity	–0.68
RP with WPAI: activity impairment	–0.72
GH with global functioning rating	0.60
PF with KCCQ-12[b]: physical limitation	0.73
PCS with KCCQ-12[b]: physical limitation	0.65
SF with KCCQ-12[b]: social limitation	0.73
MCS with KCCQ-12[b]: social limitation	0.58
PCS with KCCQ-12[b]: quality of life	0.61
MCS with KCCQ-12[b]: quality of life	0.51

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; WPAI, Work Productivity and Activity Impairment.
[a]Administered to a subsample of study participants currently employed.
[b]Administered to a subsample of study participants with cardiac involvement.

Known-Groups Validity

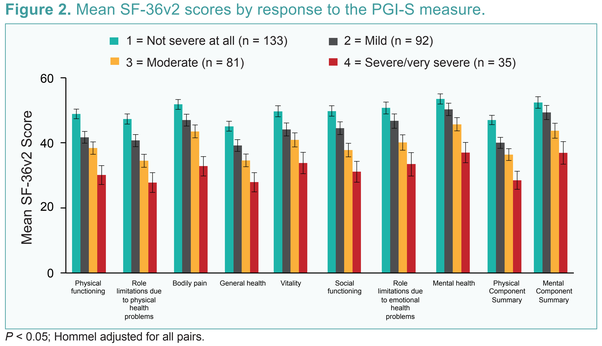
- Tests for known-groups validity indicated that patient groups with greater disease severity had deficits in physical and mental functioning
- Scores for patients with self-reported complete hematologic response or remission were significantly greater than scores for patients with no response to treatment (unadjusted $P < 0.05$ for all scores). After adjusting for multiple comparisons, the differences in means remained significant for most scales and summary scores; the exceptions were RE, MH, and MCS (**Figure 1**)
- No significant differences were observed between partial hematologic response and no response after adjusting for multiple comparisons
- All SF-36v2 scores were also significantly associated with responses to the PGI-S based on an overall ANOVA ($P < 0.001$ for all scores). All pairwise comparisons were significant before and after adjusting for multiple comparisons ($P < 0.05$ for all) (**Figure 2**)

Figure 1. Mean SF-36v2 scores by most recent hematologic response status.*



*Status unknown for 51 respondents.
[a]CR vs PR; unadjusted $P < 0.05$.
[b]CR vs PR; Hommel adjusted $P < 0.05$.
[c]CR vs no response; unadjusted $P < 0.05$.
[d]CR vs no response; Hommel adjusted $P < 0.05$.

Figure 2. Mean SF-36v2 scores by response to the PGI-S measure.



$P < 0.05$; Hommel adjusted for all pairs.

CONCLUSIONS

- This study provided robust evidence of the psychometric properties of the SF-36v2 in a diverse sample of patients with AL amyloidosis
- This research extends previous qualitative studies that support the SF-36v2 as a valid measure of HRQoL in patients with AL amyloidosis[8,9]
- Planned future analyses will assess responsiveness and confirm psychometric properties of the SF-36v2 in clinic-based samples of patients with AL amyloidosis

REFERENCES

1. Seldin DC et al. *Blood*. 2004;104:1888-1893.
2. Falk RH et al. *N Engl J Med*. 1997;337:898-909.
3. Scoggins JF, Patrick DL. *Contemp Clin Trials*. 2009;30:289-292.
4. Maruish ME. *User's Manual for the SF-36v2 Health Survey*. 3rd ed. Lincoln, RI: QualityMetric, Inc; 2011. https://campaign.optum.com/optum-outcomes.html. Accessed May 23, 2016.
5. Reilly MC et al. *PharmacoEconomics*. 1993;4:353-365.
6. Spertus JA, Jones PG. *Circ Cardiovasc Qual Outcomes*. 2015;8:460-462.
7. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. http://worldcatlibraries.org/wcpa/oclc/2344751. Accessed May 12, 2016.
8. White MK et al. Cognitive debriefing of the SF-36v2 with immunoglobulin light chain (AL) amyloidosis patients. Presented at: 21st Annual ISPOR; May 21-25, 2016; Washington, DC.
9. White MK et al. Health-related quality of life in patients with AL amyloidosis: qualitative interviews with physicians and patients. Presented at: 57th ASH Annual Meeting and Exposition; December 5-8, 2016; Orlando, FL.

SF Health Surveys

ACKNOWLEDGMENT

This study was funded by Prothena Biosciences Inc.