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

Michelle K. White,1 Martha Bayliss,1 Kristen L. McCaustland,1 Spencer D. Guthrie2

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

BACKGROUND

• Amyloid light chain (AL) amyloidosis, a rare protein-misfolding disease, leads to deficits in health-related quality of life (HRQoL). Patients with AL amyloidosis have a wide variety of nonspecific symptoms, organ involvement, and functional impairment.

• 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 explore 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.

• The SF-36v2 has been used to quantify the impact of AL amyloidosis on HRQoL, 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 groups, 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 dimensions of functioning and well-being including:

  - Physical functioning (PF) — Vitality (VT)
  - Role physical (RP) — Social functioning (SF)
  - Bodily pain (BP) — Role emotional (RE)
  - General health (GH) — 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 normed 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 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 outcomes)

• Kansas City Cardiomyopathy Questionnaire (KCCQ-12)

  - 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 (e.g., approximately comparable means and standard deviations)

• A response consistency index (RCI) was calculated for each patient to gauge how far 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 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 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 rating pain

  - RP with WPAI: absenteeism

  - RP with WPAI: presenteeism

  - RP with WPAI: work productivity

  - RP with WPAI: activity impairment

  - GH with global functioning rating

  - PF with KCCQ-12: physical limitation

  - PCS with KCCQ-12: physical limitation

  - SF with KCCQ-12: social limitation

  - MCS with KCCQ-12: social limitation

  - PF, PCS, 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

  - Hemoglobin-adjusted P-values were used to control for multiple comparisons

RESULTS

Data Quality

• Assessment of the data showed excellent response distribution; summated rating scales were satisfactory

• More than 94% of patients had RCI’s 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 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

<table>
<thead>
<tr>
<th>SF-36v2 Scale</th>
<th>No. of Items</th>
<th>Cronbach’s α</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical functioning</td>
<td>10</td>
<td>0.93</td>
<td>0.86</td>
</tr>
<tr>
<td>Role physical</td>
<td>1</td>
<td>0.91</td>
<td>0.77</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>2</td>
<td>0.91</td>
<td>0.78</td>
</tr>
<tr>
<td>General health</td>
<td>5</td>
<td>0.78</td>
<td>0.86</td>
</tr>
<tr>
<td>Physical limitation</td>
<td></td>
<td>0.87</td>
<td>0.77</td>
</tr>
<tr>
<td>Social functioning</td>
<td>2</td>
<td>0.90</td>
<td>0.73</td>
</tr>
<tr>
<td>Role emotional</td>
<td>3</td>
<td>0.95</td>
<td>0.76</td>
</tr>
<tr>
<td>Mental health</td>
<td>5</td>
<td>0.88</td>
<td>0.78</td>
</tr>
<tr>
<td>Physical Component Summary</td>
<td>1</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td>Mental Component Summary</td>
<td></td>
<td>0.78</td>
<td>0.61</td>
</tr>
</tbody>
</table>

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

Table 2. Convergent Validity

<table>
<thead>
<tr>
<th>SF-36v2 Scale</th>
<th>Comparison</th>
<th>Pearson Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP with numeric rating pain</td>
<td>-0.82</td>
<td></td>
</tr>
<tr>
<td>RP with WPAI: absenteeism</td>
<td>-0.45</td>
<td></td>
</tr>
<tr>
<td>RP with WPAI: presenteeism</td>
<td>-0.67</td>
<td></td>
</tr>
<tr>
<td>RP with WPAI: work productivity</td>
<td>-0.68</td>
<td></td>
</tr>
<tr>
<td>RP with WPAI: activity impairment</td>
<td>-0.72</td>
<td></td>
</tr>
<tr>
<td>GH with global functioning rating</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>PF with KCCQ-12: physical limitation</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>PCS with KCCQ-12: physical limitation</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>SF with KCCQ-12: social limitation</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>MCS with KCCQ-12: social limitation</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>PF with KCCQ-12: quality of life</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>PCS with KCCQ-12: quality of life</td>
<td>0.51</td>
<td></td>
</tr>
</tbody>
</table>

Convergent Validity

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

• 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.

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

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.

• Planned future analyses will assess responsiveness and confirm psychometric properties of the SF-36v2 in clinic-based samples of patients with AL amyloidosis.

REFERENCES

[3] Prothena Biosciences Inc. South San Francisco, California, United States

ACKNOWLEDGMENT

This study was funded by Prothera Biosciences Inc.