UC Davis

UC Davis Previously Published Works

Title

Do the unlabeled response categories of the Minnesota Living with Heart Failure Questionnaire satisfy the monotonicity assumption of simple-summated scoring?

Permalink

<https://escholarship.org/uc/item/2zz956q2>

Journal

Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation, 29(5)

ISSN

0962-9343

Authors

Uy, Visith
Hays, Ron D
Xu, Jay J
et al.

Publication Date

2020-05-01

DOI

10.1007/s11136-020-02422-8

Peer reviewed

Do the unlabeled response categories of the Minnesota Living with Heart Failure Questionnaire satisfy the monotonicity assumption of simple-summed scoring?

Visith Uy¹ · Ron D. Hays^{1,2,3} · Jay J. Xu⁴ · Peter M. Fayers^{5,6} · Andrew D. Auerbach⁷ · Jeanne T. Black⁸ · Lorraine S. Evangelista⁹ · Theodore G. Ganiats¹⁰ · Patrick S. Romano^{11,12} · Michael K. Ong^{1,2,13}

Accepted: 8 January 2020
© Springer Nature Switzerland AG 2020

Abstract

Purpose Half of the 21-item Minnesota Living with Heart Failure Questionnaire (MLHFQ) response categories are labeled (0 = No, 1 = Very little, 5 = Very much) and half are not (2, 3, and 4). We hypothesized that the unlabeled response options would not be more likely to be chosen at some place along the scale continuum than other response options and, therefore, not satisfy the monotonicity assumption of simple-summed scoring.

Methods We performed exploratory and confirmatory factor analyses of the MLHFQ items in a sample of 1437 adults in the Better Effectiveness After Transition—Heart Failure study. We evaluated the unlabeled response options using item characteristic curves from item response theory—graded response models for MLHFQ physical and emotional health scales. Then, we examined the impact of collapsing response options on correlations of scale scores with other variables.

Results The sample was 46% female; 71% aged 65 or older; 11% Hispanic, 22% Black, 54% White, and 12% other. The unlabeled response options were rarely chosen. The standard approach to scoring and scores obtained by collapsing adjacent response categories yielded similar associations with other variables, indicating that the existing response options are problematic.

Conclusions The unlabeled MLHFQ response options do not meet the assumptions of simple-summed scoring. Further assessment of the performance of the unlabeled response options and evaluation of alternative scoring approaches is recommended. Adding labels for response options in future administrations of the MLHFQ should be considered.

Keywords Health-related quality of life · Minnesota Living with Heart Failure Questionnaire · Item characteristic curves · Survey response options · Heart failure

Introduction

The Minnesota Living with Heart Failure Questionnaire (MLHFQ) is one of the most widely used health-related quality of life (HRQOL) measures targeted at individuals with heart failure [1, 2]. Garin et al. recommended the MLHFQ based on a systematic review of disease-specific HRQOL measures for heart failure [3]. In addition, the

MLHFQ is one of only two Medical Device Development Tools approved by the Food and Drug Administration (FDA) [4]. The 21-item instrument assesses mobility, physical symptoms (e.g., fatigue, shortness of breath, swelling in ankles or legs, and treatment side effects), emotional distress (e.g., loss of self-control, worry, and difficulty concentrating, depressed mood), sleep, social/role function, sexual activity, hospitalization, and medical costs [5]. Half of the MLHFQ response categories have verbal labels (0 = *No*, 1 = *Very little*, 5 = *Very Much*) and half are unlabeled (2, 3 and 4). The MLHFQ is an “end-labeled” response scale whereby responses at the extremes (low or high) are more likely to be selected than the unlabeled choices [6].

The 21-item overall MLHFQ score ranges from 0 to 105, with a higher score signifying greater impairment (worse HRQOL) [5]. Also recommended for scoring is an

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11136-020-02422-8>) contains supplementary material, which is available to authorized users.

✉ Ron D. Hays
drhays@ucla.edu

Extended author information available on the last page of the article

8-item *physical health scale* (items 2–7 and 12–13) and a 5-item *emotional health scale* (items 17–21). Bilbao confirmed the physical and emotional health scales and also suggested scoring a 6-item social health scale (items 8–10, 14–16) originally proposed by Munyombwe et al. [7, 8]. All MLHFQ scales are created using simple-summed scoring. That is, the distance between response categories is assumed to be equal. For example, an item response of 2 is treated as indicating more impairment by one unit than a response of 1.

Each polytomous response category in a survey item should have the highest likelihood of being selected somewhere along the underlying distribution of the measured concept. Unless a nominal response model is employed, the location on the continuum where response categories are most likely to be chosen should indicate a monotonic relationship between scale scores and the probability of selecting item response options [9]. Item characteristic curves (ICCs) can be used to evaluate whether this requirement is met [10, 11]. Figure 1 illustrates an ICC that provides support for the performance of the *poor* to *excellent* response scale for a general health rating item [12]. This curve shows that people with the most positive physical health scores have the highest probability of selecting *excellent* while those with the most negative scores have the highest probability of selecting *poor*. The *fair*, *good*, and *very good* response options are monotonically ordered between the two extreme options.

We analyzed data from responses to the MLHFQ in a sample of patients with heart failure to assess the performance of the unlabeled response categories and determine

if simple-summed scoring of the 0–5 response scale is warranted.

Methods/design

Sample

This randomized control trial assigned patients with heart failure at an individual level to either a telemedicine intervention or usual care arm at six academic medical centers in California between October 12, 2011, and September 30, 2013. The sample was 1437 individuals participating in the Better Effectiveness After Transition—Heart Failure (BEAT-HF) Study [13, 14]. The baseline sample was 29% 50–64, 35% 65–79, and 36% 80 years of age or older; 46% were female; 11% were Hispanic, 22% (non-Hispanic) Black, 54% (non-Hispanic) White; and 12% other.

The study was approved by the University of California, Los Angeles (UCLA) institutional review board (IRB#14-000579) and all other study institutions were subject to this review. A data and safety monitoring board was convened for the study and reviewed data during the study enrollment period. The study was registered at clinicaltrials.gov (NCT01360203).

Analysis plan

The MLHFQ was administered at baseline and at 7, 30, and 180 days post-baseline. Because of the variability of

Fig. 1 Item characteristic curve for how would you rate your health item

the MLHFQ factor structure in prior studies, we conduct an exploratory factor analysis of the MLHFQ items using baseline data. Then, we compute item–scale correlations (corrected for item overlap with scale scores) for physical, emotional, and social health scales, and we estimate internal consistency reliability coefficients for the physical and emotional health scales [15, 16]. Next, we estimate categorical confirmatory factor analytic models and evaluate model fit using the comparative fit index (CFI). We then estimate item response theory (IRT)—graded response models to obtain ICCs [17]. Based on the ICCs, we examine different scoring approaches for the MLHFQ by collapsing adjacent response categories justified by the ICCs. We compare IRT model fit for the original 6 categories versus a collapsed 3-category scoring of the MHLFQ scales using the Akaike information criterion (AIC). We also compare fit using the widely applicable information criterion or Watanabe–Akaike information criterion, WAIC [18]. We fit bidimensional Bayesian graded response models by specifying identical centered normal priors with shared standard deviation for the set of person parameters, the set of item easiness parameters, and the set of item discrimination parameters. Specifically [19], we specified $N(0, \sigma_\theta^2)$ priors for the person parameters and a Half-Normal(0, 3) prior for the standard deviation, σ_θ ; $N(0, \sigma_\xi^2)$ priors for the item easiness parameters and a Half-Normal(0, 3) prior for the standard deviation, σ_ξ ; and $N(0, \sigma_\alpha^2)$ priors for the item discrimination parameters on the log scale and a Half-Normal(0, 1) prior for the standard deviation, σ_α . A consequence of the set of person, item easiness, and item discrimination parameters each sharing the same standard deviation is that their posterior mean estimates “borrow strength” from the entire sample and shrink towards a common mean, providing more stable estimates, a phenomenon called partial pooling [20]. This modeling approach is recommended for Bayesian IRT models [19]. We generated a single chain of 7500 Markov chain Monte Carlo samples from the posterior distribution, discarding the first 2500 samples as burn-in.

We also compare the original scoring of the MHLFQ items with different options for collapsing the intermediate response options by examining associations of the different scoring schemes with other variables. We hypothesize that the MLHFQ scales (higher score is worse) will be negatively associated with self-rated general health and positively associated with New York Heart Association functional classification [21], the 15-item Geriatric Depression Scale [22], a count of all-cause hospital readmission within 30 days, and comorbidities [23, 24]. We examine the AHRQ Elixhauser Comorbidity Index (ECI) [25, 26] using two set of weights: the first based on 30-day readmission and the second on in-hospital mortality.

Each set of weights is summed to generate an index score where higher values represent increased likelihood of each respective outcome: a readmission-based ECI and mortality-based ECI. Finally, we assess whether the different MLHFQ scoring schemes change conclusions about the 30-day and 180-day evaluation of the BEAT-HF study intervention on the MLHFQ [5].

Factor analyses were conducted with Mplus version 6.12. We used the brms version 2.10.0 R package [27] to fit Bayesian graded response models for the original 6-category and the collapsed 3-category models. The brms package performs Bayesian estimation using the Stan programming language [28]. All other analyses were completed using STATA version 14.2 [29].

All statistical tests were two-tailed and $p < 0.05$ was the threshold to be considered statically significant. Because the focus is on whether results are similar or not for the original scoring and alternative scoring options for the MLHFQ, we do not adjust for multiple significance tests and significant results should be interpreted cautiously.

Results

The number of factor criteria suggested the possibility of three underlying factors (e.g., 3 principal component eigenvalues exceeded 1.0). A three-factor promax-rotated solution (Table 1) indicated that two of the three factors largely reflected the physical and emotional health factors described in the MLHFQ user manual [5]. An oblique rather than orthogonal rotation was performed to enable estimation of the correlations among factors rather than assume they were uncorrelated. The third factor was a mix of diverse content but was closest to the social health factor suggested by prior investigators [8]. Item 1 (swelling in ankles and legs) and item 11 (eat less food than you like), which were not included in the original proposed scales, were most strongly associated with the physical and social health factors, respectively, in our study.

We estimated an item–scale correlation matrix supported by the exploratory factor analysis for the physical, emotional, and social health scales (Table 2). Eight of the ten items in our physical health scale are items in the standard eight-item MLHFQ physical health scale. Four of the five items in our emotional health scale are items in the standard 5-item MLHFQ scale. Item 9 (recreational activities difficult) fit best with physical health and item 16 (side effects from treatment) with emotional health. Only three items were consistent with the previously proposed social health scale: item 8 (working to make a living), item 10 (sexual activities difficult), and item 15 (medical cost). Items 14 (hospitalization) and 20 (making it difficult for you to

Table 1 Promax-rotated three-factor pattern matrix (standardized regression coefficients) for the Minnesota Living with Heart Failure Questionnaire items

Items	Description	Physical	Emotional	Social
1	Swelling in ankles or legs	0.466	0.121	0.044
2	Sit or lie down to rest during the day	0.753	0.052	− 0.013
3	Walking or climbing stairs difficult	0.955	− 0.113	− 0.076
4	Working around house or yard difficult	0.874	− 0.016	− 0.038
5	Going places away from home difficult	0.752	0.117	− 0.025
6	Sleeping difficult	0.495	0.103	0.224
7	Relating to/doing things with friends/family difficult	0.508	0.284	0.139
9	Recreational activities difficult	0.492	0.143	0.260
12	Short of breath	0.606	0.093	− 0.007
13	Tired, fatigue, or low on energy	0.630	0.093	0.111
14	Hospitalization	0.319	0.350	0.047
16	Side effects from treatment	0.114	0.571	− 0.033
17	Feeling burden to family or friends	0.007	0.682	0.118
18	Feeling a loss of self-control	0.069	0.810	− 0.068
19	Being worried	0.096	0.704	0.012
21	Feel depressed	0.035	0.668	0.055
8	Working to earn a living difficult	− 0.022	− 0.068	0.794
10	Sexual activities difficult	0.056	− 0.058	0.757
11	Eat less of the foods you like	0.116	0.058	0.522
15	Medical cost	− 0.102	0.251	0.495
20	Difficult to concentrate or remember things	0.089	0.244	0.340

Pietri et al. [5] recommended scoring an 8-item *physical health scale* (items 2–7 and 12–13) and a 5-item *emotional health scale* (items 17–21). Bilbao et al. [7] suggested scoring a 6-item social health scale (items 8–10, 14–16) originally proposed by Munyombwe et al. [8]

concentrate or remember things) correlated similarly with all the scales.

Because our data strongly agreed with what has been reported for the physical and emotional health scales but indicated inconsistencies with previous findings for the social health scale, we focused our evaluation on the former two and excluded the social health scale.

Physical health scale

Coefficient alpha was 0.91 and ordinal alpha was 0.95 for the 10 physical health items. A categorical one-factor model fit the data reasonably well (CFI=0.94). ICCs for the items show that the unlabeled response categories (2, 3, and 4) were rarely chosen and did not satisfy the monotonicity assumption (left column of Fig. 2). The most extreme is item 1, where response categories 1–4 are never the most likely to be selected. The third response category, the mid-point between labeled categories, performed adequately for the nine other items. Overall, the ICCs indicated sub-optimal performance for most of the intermediate response categories.

Therefore, we trichotomized the response options. A response of 0 (equal to “no effect”) or 1 (“very little”) was unchanged, responses of 1–4 were recoded to 1 (equal to

“very little”), and a response of 5 was recoded to 2 (equal to “very much”). We then recalculated the psychometric properties and estimated ICCs for the recoded items. Coefficient alpha (0.90), ordinal alpha (0.95), and a one-factor model were largely unchanged (CFI=0.94). The ICCs drastically improved and suggested the MLHFQ response categories should be collapsed. The graded response model for the three collapsed categories fit the data better than it did for the original six categories (AIC difference=11,983.34). The WAIC for the original 6-category model was 30,998.25 with a standard error of 240.27, and the WAIC for the collapsed 3-category model was 19,461.57 with a standard error of 173.44. Clearly, the collapsed 3-category model provides a superior fit to the data.

Emotional health scale

Results for the emotional health scale were like what was found for the physical health scale. Coefficient alpha was 0.84 and ordinal alpha was 0.89 for the 5 emotional health items. A categorical one-factor model fit the data well (CFI=0.99). ICCs for the 5 emotional health items show that unlabeled response categories (2, 3, and 4) were again rarely chosen (left column of Fig. 3).

Table 2 Item–scale correlations for Minnesota Living with Heart Failure scales

Item	Description	Physical health	Emotional health	Social health
1	Swelling in ankles or legs	0.56 ^a	0.44	0.34
2	Sit or lie down to rest during the day	0.75 ^a	0.53	0.39
3	Walking or climbing stairs difficult	0.76 ^a	0.47	0.37
4	Working around house or yard difficult	0.78 ^a	0.54	0.41
5	Going places away from home difficult	0.79 ^a	0.57	0.41
6	Sleeping difficult	0.69 ^a	0.53	0.49
7	Relating to/doing things with friends/family difficult	0.77 ^a	0.66	0.51
9	Recreational activities difficult	0.73 ^a	0.59	0.54
12	Short of breath	0.62 ^a	0.48	0.35
13	Tired, fatigue, or low on energy	0.72 ^a	0.54	0.46
16	Side effects from treatment	0.49	0.55 ^a	0.34
17	Feeling burden to family or friends	0.55	0.70 ^a	0.49
18	Feeling a loss of self-control	0.59	0.76 ^a	0.41
19	Being worried	0.59	0.71 ^a	0.44
21	Feel depressed	0.53	0.65 ^a	0.41
8	Working to earn a living difficult	0.41	0.36	0.66 ^a
10	Sexual activities difficult	0.47	0.41	0.64 ^a
11	Eat less of the foods you like	0.47	0.42	0.52 ^a
15	Medical cost	0.37	0.45	0.51 ^a
14	Hospitalization	0.57	0.56	0.41
20	Difficult to concentrate or remember things	0.46	0.48	0.48

Correlations were 0.64 between physical and emotional, 0.54 between physical and social, and 0.57 between emotional and social

^aCorrelation is corrected for overlap with the scale score

We followed the same procedures mentioned above and trichotomized the response choices. Coefficient alpha decreased slightly to 0.82, ordinal alpha remained the same (0.89), and a one-factor categorical model fit the data well (CFI = 0.99). The resulting ICCs performed much better and strengthened the argument to collapse the MLHFQ response choices. The graded response model for the three collapsed categories fit the data better than it did for the original six categories (AIC difference = 6924.41). The WAIC for the original 6-category model was 17,754.07 with a standard error of 160.25, and the WAIC for the collapsed 3-category model was 10,876.81 with a standard error of 118.17. Again, the collapsed 3-category model provides a superior fit to the data.

Associations of MLHFQ scales with other variables

As noted above, collapsing MLHFQ response categories produced superior ICCs and suggested examination of other recoding schemes to identify the optimal one. A second option was to retain 4 response categories by scoring responses of 1 and 2 together as 1, scoring 3 as 2, and scoring 4 and 5 as 3. A third option preserved 5 categories

by scoring 4 and 5 as 4. We compared associations of the original and these three new simple-summed physical and emotional health scale scores with other variables.

Relationships with the self-rated general health item were similar across the scoring schemes, but product moment correlations were slightly larger for the 6-category scoring (Online Appendix Table 1). Correlations ranged from 0.24 to 0.26 for physical health and from 0.18 to 0.21 for emotional health. Associations between the NYHA functional classification and physical health scale score estimates were similar but a little larger for the 5-category scoring (one-way ANOVA F-statistics ranged from 10.26 to 14.22). For hospital readmissions within 30 days, only the 5- and 6-category physical health scales produced statistically significant product moment coefficients with correlations of 0.058 and 0.061, respectively. A slightly larger correlation of the emotional health scales with the Geriatric Depression Scale was found for the 6-category scoring ($r = 0.37$ compared to 0.37 for 3 categories, 0.35 for four categories, and 0.34 for five categories).

The MLHFQ summary score had statistically significant product moment coefficients with readmission-based ECI, but not mortality-based ECI. The summary score for the 4-category scoring produced the largest coefficient with readmission-based ECI. Estimates from the 3-category

scoring produced the smallest, with 5- and 6-category scoring in between. However, absolute values for all coefficients were small (r 's ranged from 0.08 to 0.09).

The significant positive effect of the BEAT-HF intervention on the 180-day MLHFQ summary score was robust to different scoring approaches (original

6-category versus 3-, 4-, and 5-category scoring): z statistics for the intervention ranged from -2.2 to -2.4 and p -values from 0.018 to 0.026 across the four different scoring options (Online Appendix Table 2).

Fig. 2 Item characteristic curves for ten physical health items using original six response categories and recoded three response categories

Fig. 2 (continued)

Discussion

The MLHFQ has been used widely, translated to over 34 languages, and approved by the FDA to evaluate the

superiority or non-inferiority of medical devices [7, 30]. Our analyses provide further evidence about the underlying factor structure of the MLHFQ. The factor analysis and item-scale correlation suggested that item 9 (recreational

Fig. 2 (continued)

activities) represented physical health in our sample. Item 20 (difficulty concentrating or remembering things) was found to be equally correlated with all three scales including physical health. Items not currently included in guidelines for existing scales, such as items 1 (swelling in ankles or legs) and 11 (eating less of the foods you like), capture meaningful information and could contribute to their respective scales [5]. Item–scale correlation indicated items 14 (hospitalization) and 20 (difficulty concentrating or remembering things) were associated with all scales. Future research is needed to continue evaluating the structure of the MLHFQ.

This study showed that the intermediate (unlabeled) response choices had suboptimal performance. They were

rarely chosen and were uninformative about underlying HRQOL. Rather than six levels of differentiation, the MLHFQ items primarily have three levels of distinction: the labeled extremes (0 “No,” 5 “Very much”) and unlabeled categories collapsed into one intermediate category along with the third labeled category (1 “Very Little”).

We also examined two other variations in collapsing the 6 categories (collapsing to 4 and 5 categories). Associations of physical and emotional health scale scores estimated using the different scoring options with other variables (self-rated general health, New York Heart Association severity classification, Geriatric Depression Scale, and hospital readmission within 30 days) were similar. Depending on the specific variables being compared, 3–5-category scoring produced

somewhat larger coefficients than the original 6-category scoring. For MLHFQ summary scores, 4-category scoring produced the largest product moment estimates with readmission-based ECI. The different scoring approaches yielded almost identical estimates of the effect of the BEAT-HF intervention on the MLHFQ total score.

While collapsing response options is warranted with the existing measure, future research could be conducted to evaluate the effect of labeling the intermediate response categories (e.g., 0 = *No*, 1 = *Very little*, 2 = *Little*, 3 = *Some*, 4 = *Much*, 5 = *Very much*). Moors, Kieruj, and Vermunt [6] recommended full labeling of categories because it makes all of them “more or less equally clear to respondents,

Fig. 3 Item characteristic curves for five emotional health items using original six response categories and recoded three response categories

Fig. 3 (continued)

which means that no preference for certain categories is facilitated simply by labeling one category and not the other” (p. 391). This approach would maintain continuity with existing scoring guides.

This study has limitations that need to be considered. How users would respond to fully labeled response categories for the MLHFQ is unknown so future research is needed. Moreover, our sample received care at academic teaching hospitals in California. Hence, results may not generalize to other settings or samples. Finally, to satisfy scale scoring assumptions, the physical and mental health scales we evaluated were slightly different than those recommended by the original developer, but the unlabeled response options are equally problematic with the “standard” version of the scales. Also, a few of the statistically significant associations (e.g., the product moment correlations between the MLHFQ physical health scale and 30-day hospital readmission) reported in the study would become non-significant if we had adjusted for multiple comparisons.

Conclusions

ICCs provide valuable information. As Fayers and Machin noted, “overlapping and disordered categories either indicate items with problems that should either be excluded or, at the very least, have categories that should be combined or reworded to produce a revised item with better performance.” [10]. ICCs are now part and parcel of item response theory software and are increasingly used to evaluate response options [31]. Future evaluation of the MLHFQ unlabeled response options is important. We also recommend that those evaluating measures with mixed unlabeled and labeled response options carry out similar analyses as those reported in this paper. Finally, it is also worth noting that some of the MLHFQ items (e.g., “Making your sexual activities difficult”) may not be applicable [32], and the example instructions indicate that respondents should select the “No” response if an item is “not applicable.” [5]. Evaluating the equivalence of “no” and “not applicable” is also needed.

Funding This study was funded in part by the Agency for Healthcare Research and Quality (R01 HS019311), the National Heart Lung and Blood Institute (RC2 HL101811), the National Institute on Aging (P30 AG021684), the Robert Wood Johnson Foundation (66336), and the Sierra Health Foundation, and the participating institutions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval UCLA Institutional Review Board approved the study.

Informed consent All study participants provided informed consent.

References

- Garin, O., Herdman, M., Vilagut, G., Ferrer, M., Ribera, A., Rajmil, L., et al. (2013). Assessing health-related quality of life in patients with heart failure: A systematic, standardized comparison of available measures. *Heart Failure Reviews*, 19(3), 359–367.
- Rector, T. S., & Cohn, J. N. (1992). Assessment of patient outcome with the Minnesota living with heart failure questionnaire: Reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. *American Heart Journal*, 124(4), 1017–1025.
- Garin, O., Ferrer, M., Pont, À., Rué, M., Kotzeva, A., Wiklund, I., et al. (2008). Disease-specific health-related quality of life questionnaires for heart failure: A systematic review with meta-analyses. *Quality of Life Research*, 18(1), 71–85.
- U.S Food & Drug Administration. (2016). Medical device development tool (MDDT) Qualified Tools. Retrieved June 14, 2019, from https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt#Qualified_Tools.
- Pietri, G., Ganse, E. V., Ferrier, M., Garin, O., & Wiklund, I. (2004). *Minnesota living with heart failure questionnaire*. Lyon, France: MAPI Research Institute.
- Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The effect of labeling and numbering of response scales on the likelihood of response bias. *Sociological Methodology*, 44(1), 369–399.
- Bilbao, A., Escobar, A., Garcia-Perez, L., et al. (2016). The Minnesota living with heart failure questionnaire: Comparison of different factor structures. *Health Qual Life Outcomes*, 14, 23.
- Munyombwe, T., Höfer, S., Fitzsimons, D., Thompson, D. R., Lane, D., Smith, K., et al. (2014). An evaluation of the Minnesota living with heart failure questionnaire using Rasch analysis. *Quality of Life Research*, 23(6), 1753–1765.
- Preston, K., Reise, S., Cai, L., & Hays, R. D. (2011). Using the nominal response model to evaluate response category discrimination in the PROMIS emotional distress item pools. *Educational and Psychological Measurement*, 71(3), 523–550.
- Fayers, P. M., & Machin, D. (2016). *Quality of life: The assessment, analysis, and reporting of patient-reported outcomes*. Chichester, UK: Wiley.
- Kaplan, R. M., & Saccuzzo, D. P. (2012). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Wadsworth Publishing.
- Hays, R. D., Spritzer, K. L., Thompson, W. W., & Cella, D. (2015). US general population estimate for “excellent”, to “poor” self-rated health item. *Journal of General Internal Medicine*, 30(10), 1511–1516.
- Black, J. T., Romano, P. S., Sadeghi, B., Auerbach, A. D., Ganiats, T. G., Greenfield, S., et al. (2014). A remote monitoring and telephone nurse coaching intervention to reduce readmissions among patients with heart failure: Study protocol for the Better Effectiveness After Transition-Heart Failure (BEAT-HF) randomized controlled trial. *Trials*, 151, 124.
- Ong, M. K., Romano, P. S., Edgington, S., Aronow, H. U., Auerbach, A. D., Black, J. T., et al. (2016). Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure. *JAMA Internal Medicine*, 176(3), 310.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Spritzer, K. L., Hays, R. D.. A SAS macro (ordalpha) to computer ordinal coefficient alpha. Retrieved June 14, 2019, from https://labs.dgsom.ucla.edu/hays/pages/programs_utilities.
- Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.) *Handbook of modern item response theory*. New York, NY: Springer.
- Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, 14, 867–897.
- Bürkner, P. (2019). Bayesian item response modelling in R with brms and Stan. arXiv preprint [arXiv:1905.09501](https://arxiv.org/abs/1905.09501).
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Dolgin, M., & New York Heart Association Criteria Committee. (1973). *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels*. Boston: Little, Brown.
- Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontology*, 5(1–2), 165–173.
- Michelson, H., Bolund, C., & Brandberg, Y. (2000). Multiple chronic health problems are negatively associated with health-related quality of life (HRQoL) irrespective of age. *Quality of Life Research*, 9(10), 1093–1104.
- Fortin, M., Bravo, G., Hudon, C., Lapointe, L., Almirall, J., Dubois, M. F., et al. (2006). Relationship between multimorbidity and health-related quality of life in patients in primary care. *Quality of Life Research*, 15(1), 83–91.
- Moore, B. J., White, S., Washington, R., Coenen, N., & Elixhauser, A. (2017). Identifying increased risk of readmission and in-hospital mortality using hospital administrative data: The AHRQ Elixhauser comorbidity index. *Medical Care*, 55(7), 698–705.
- Walraven, C. V., Austin, P. C., Jennings, A., Quan, H., & Forster, A. J. (2009). A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical Care*, 47(6), 626–633.
- Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Stan Development Team. (2018). Stan modeling language user's guide and reference manual. Version 2.18.0. Retrieved from <https://mc-stan.org>.
- StataCorp., (2015). *Stata statistical software: release 14*. College Station, TX: StataCorp LP.
- Food & Drug Administration. (2016). Medical device development tool (MDDT) qualification decision summary for Minnesota living with heart failure questionnaire (MLHFQ). Retrieved June 14, 2019, from <https://www.fda.gov/media/112157/download>.
- Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care*, 38(9), I128–I142.
- Hak, T., Willems, D., Wal, G. V., & Visser, F. (2004). A qualitative validation of the Minnesota living with heart failure questionnaire. *Quality of Life Research*, 13(2), 417–426.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Visith Uy¹ · Ron D. Hays^{1,2,3} · Jay J. Xu⁴ · Peter M. Fayers^{5,6} · Andrew D. Auerbach⁷ · Jeanne T. Black⁸ · Lorraine S. Evangelista⁹ · Theodore G. Ganiats¹⁰ · Patrick S. Romano^{11,12} · Michael K. Ong^{1,2,13}

¹ Division of General Internal Medicine & Health Services Research, UCLA Department of Medicine, 1100 Glendon Avenue, Suite 850, Los Angeles, CA 90024, USA

² Department of Health Policy & Management, University of California, Los Angeles, Los Angeles, CA, USA

³ RAND Health, RAND Corporation, Santa Monica, CA, USA

⁴ UCLA Department of Biostatistics, 650 Charles E. Young Drive South, 51-254 CHS, Los Angeles, CA 90024, USA

⁵ Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway

⁶ Emeritus Professor, Institute of Applied Health Sciences, King's College, University of Aberdeen, Aberdeen AB24 3FX, UK

⁷ Division of Hospital Medicine, Department of Medicine, University of California, San Francisco, 533 Parnassus Avenue, Room 131, San Francisco, CA 94117, USA

⁸ Department of Orthopaedics, Cedars-Sinai Medical Center, Los Angeles, 116 N. Robertson Blvd., Suite 800, Los Angeles, CA 90048, USA

⁹ Sue & Bill Gross School of Nursing, University of California, Irvine, 284 Berk Hall, Irvine, CA 92697, USA

¹⁰ Department of Family Medicine and Public Health, University of California, San Diego, 9500 Gilman Drive #0725, La Jolla, CA 92093, USA

¹¹ Department of Internal Medicine, University of California, Davis, 4860 Y Street, Suites 0101 & 0400, Sacramento, CA 95817, USA

¹² Department of Pediatrics, University of California, Davis, Sacramento, CA, USA

¹³ Department of Medicine, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, USA