Reliability, Validity, and Responsiveness of InFLUenza Patient-Reported Outcome (FLU-PRO©) Scores in Influenza-Positive Patients

John H. Powers III, MD[1,*], *Elizabeth D. Bacci, PhD*[2], *M. Lourdes Guerrero, MD, MS*[3], *Nancy Kline Leidy, PhD*[4], *Sonja Stringer, MPH*[4], *Katherine Kim, MPH*[4], *Matthew J. Memoli, MD, MS*[5], *Alison Han, MD, MS*[5], *Mary P. Fairchok, MD*[6,7,8], *Wei-Ju Chen, PhD*[7,8], *John C. Arnold, MD*[9], *Patrick J. Danaher, MD*[10], *Tahaniyat Lalani, MBBS*[8,11], *Michelande Ridoré, MS*[12], *Timothy H. Burgess, MD, MPH*[7,13], *Eugene V. Millar, PhD*[7,8], *Andrés Hernández, MD, MS*[14], *Patricia Rodríguez-Zulueta, MD*[15], *Mary C. Smolskis, RN, MA*[4], *Hilda Ortega-Gallegos, BS*[3], *Sarah Pett, MD, PhD*[16,17], *William Fischer, MD*[18], *Daniel Gillor, DrMed*[19], *Laura Moreno Macias, MD*[20], *Anna DuVal, MPH*[21], *Richard Rothman, MD, PhD*[21], *Andrea Dugas, MD, PhD*[21], *Guillermo M. Ruiz-Palacios, MD, FIDSA*[3]

[1]Clinical Research Directorate/Clinical Monitoring Research Program, Leidos Biomedical Research, Inc., NCI Campus at Frederick, Frederick, MD, USA; [2]Evidera, Seattle, WA, USA; [3]Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico; [4]Evidera, Bethesda, MD, USA; [5]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA; [6]Madigan Army Medical Center, Fort Lewis, WA, USA; [7]Infectious Disease Clinical Research Program, Department of Preventive Medicine and Biostatistics, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; [8]Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, USA; [9]Naval Medical Center, San Diego, CA, USA; [10]Defense Institute for Medical Operations, San Antonio, TX, USA; [11]Naval Medical Center, Portsmouth, VA, USA; [12]Children's National Medical Center, Washington, DC, USA; [13]Walter Reed National Military Medical Center, Bethesda, MD, USA; [14]Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, Mexico City, Mexico; [15]Hospital General Dr. Manuel Gea González, Mexico City, Mexico; [16]University College London, London, UK; [17]Kirby Institute, University of New South Wales, Sydney, New South Wales, Australia; [18]University of North Carolina, Chapel Hill, NC, USA; [19]Cologne, Germany; [20]Hospital General de Agudos JM Ramos Mejia, Buenos Aires, Argentina; [21]Johns Hopkins University School of Medicine, Baltimore, MD, USA (for the INSIGHT Flu 002 Plus Writing Group, the México Emerging Infectious Diseases Clinical Research Network, and the Infectious Diseases Clinical Research Program)

ABSTRACT

Objectives: To assess the reliability, validity, and responsiveness of InFLUenza Patient-Reported Outcome (FLU-PRO©) scores for quantifying the presence and severity of influenza symptoms. **Methods:** An observational prospective cohort study of adults (≥18 years) with influenza-like illness in the United States, the United Kingdom, Mexico, and South America was conducted. Participants completed the 37-item draft FLU-PRO daily for up to 14 days. Item-level and factor analyses were used to remove items and determine factor structure. Reliability of the final tool was estimated using Cronbach α and intraclass correlation coefficients (2-day reliability). Convergent and known-groups validity and responsiveness were assessed using global assessments of influenza severity and return to usual health. **Results:** Of the 536 patients enrolled, 221 influenza-positive subjects comprised the analytical sample. The mean age of the patients was 40.7 years, 60.2% were women, and 59.7% were white. The final 32-item measure has six factors/domains (nose, throat, eyes, chest/respiratory, gastrointestinal, and body/systemic), with a higher order factor representing symptom severity overall (comparative fit index = 0.92; root mean square error of approximation = 0.06). Cronbach α was high (total = 0.92; domain range = 0.71–0.87); test-retest reliability (intraclass correlation coefficient, day 1–day 2) was 0.83 for total scores and 0.57 to 0.79 for domains. Day 1 FLU-PRO domain and total scores were moderately to highly correlated (≥0.30) with Patient Global Rating of Flu Severity (except nose and throat). Consistent with known-groups validity, scores differentiated severity groups on the basis of global rating (total: $F = 57.2$, $P < 0.001$; domains: $F = 8.9$–67.5, $P < 0.001$).

Conflicts of interest: The contents of this publication are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions, or policies of the Department of Health and Human Services, Uniformed Services University of the Health Sciences, the Department of Defense, or the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the US government.

* *Address correspondence to:* John H. Powers III, Leidos Biomedical Research, Inc., in support of Division of Clinical Research, National Institute of Allergy and Infectious Diseases, 5601 Fishers Lane, Room 4D50, Bethesda, MD 20892.

E-mail: john.powers@nih.gov.

http://dx.doi.org/10.1016/j.jval.2017.04.014

Subjects reporting return to usual health showed significantly greater ($P < 0.05$) FLU-PRO score improvement by day 7 than did those who did not, suggesting score responsiveness. **Conclusions:** Results suggest that FLU-PRO scores are reliable, valid, and responsive to change in influenza-positive adults.

Keywords: influenza, patient-reported outcome, psychometric, reliability, responsiveness, validity.

Introduction

Approximately 5% to 20% of the US population is infected with influenza yearly, with 200,000 hospitalizations and 36,000 deaths [1–3]. Worldwide, influenza causes 3 million to 5 million severe cases and 250,000 to 500,000 deaths annually [4]. Symptoms range from mild to severe and include various systemic and respiratory symptoms, with gastrointestinal symptoms occurring less frequently [3].

Despite the prevalence of influenza and many research studies evaluating its natural history and treatment options, there are few validated patient-reported outcome measures for quantifying symptoms. Two previously developed instruments have been published but are limited by the populations studied with smaller numbers of patients with influenza studied compared with influenza-like illness (ILI) [5,6].

A validated, standardized patient-reported influenza symptom scale that comprehensively assesses the symptom experience in influenza across multiple body systems would allow for consistent, accurate assessments of symptoms associated with various viral strains over the course of the disease within and across subgroups. Use would facilitate meta-analyses, cross-product evaluations, and more precise estimates of treatment effects. Standardized measures should be developed using good research practices [7–9]. Instruments intended for use in drug development should address recommendations of the US Food and Drug Administration [10], including attention to content validity and quantitative testing in the target population for designated contexts of use.

The purpose of the InFLUenza Patient-Reported Outcome (FLU-PRO©) measure is to comprehensively assess the presence and severity of influenza symptoms across body systems often affected by these viruses. The ultimate intent was to develop a reliable, valid, and responsive measure for use in profiling the symptomatic manifestations of influenza on any given day, track changes over time, and test the effects of treatments. To ensure content validity, we used a two-stage qualitative instrument development methodology. In stage 1, we conducted concept elicitation interviews in the United States and Mexico to gather information regarding patient experience of influenza symptoms (i.e., type, magnitude, expression, pattern of onset, and recovery) [11,12]. Results informed the development of a draft measure, including content (candidate items), structure (response options, recall, and instructions), and conceptual framework [13]. In stage 2, we conducted cognitive interviews to assess completeness, comprehension, and ease of use of the draft measure from the respondent's perspective [13]. This work resulted in a draft instrument with 37 candidate questions ready for quantitative testing in the target population.

The objectives of this study were to 1) evaluate performance of the 37 candidate items; 2) reduce the number of items as empirically and conceptually appropriate; 3) finalize measurement/domain structure and develop a scoring algorithm for the final instrument, the FLU-PRO; and 4) explore the reliability, validity, and responsiveness of FLU-PRO total and domain scores.

Methods

Study Design and Sample

This was a prospective, observational study of English- and Spanish-speaking hospitalized and nonhospitalized adults 18 years or older with acute influenza. Patients seeking care for influenza symptoms at participating military or civilian clinics in the United States (16 sites), Argentina (2 sites), the United Kingdom (1 site), and Mexico (3 sites) were recruited in influenza seasons in northern and southern hemispheres. Influenza status was assessed through a positive polymerase chain reaction, rapid antigen test, and/or viral culture by nasal or nasopharyngeal swab.

We prespecified subjects testing positive for influenza as the target population and the primary analytical sample, with a goal of 200 or more subjects (100 for confirmatory factor analysis [CFA] and 185 [5 per item] for exploratory factor analysis [EFA]) [14], assuming that 50% of enrolled subjects testing positive for influenza would permit separate analyses on performance of the FLU-PRO in ILI. Given the different context of use, ILI results are presented elsewhere.

A total of 536 English- and Spanish-speaking patients were enrolled in the study; 441 had diary entries on day 1 and at least 1 day thereafter, qualifying them for analyses. Two hundred twenty-one were influenza-positive (see Appendix Figure S2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.014).

Procedures

Clinical research coordinators recruited participants with influenza-like symptoms. Patients providing consent: 1) completed clinic-based baseline assessments of sociodemographic and clinical characteristics; 2) were tested for laboratory-confirmed influenza; and 3) completed daily diaries for up to 14 days after enrollment. This included the 37-item draft FLU-PRO symptom diary and nine additional questions for validation purposes. At Mexico sites, diaries were completed via telephone interviews with data entered directly into a Web-based portal. Participants in the United States, the United Kingdom, and Argentina completed the survey via either an interviewer-administered method or a Web-based system using their personal devices. Translation procedures for Spanish followed the International Society for Pharmacoeconomics and Outcomes Research guidelines [15]. The study was conducted with informed consent, institutional review board approval, and in accordance with the Declaration of Helsinki [15].

Instruments: Patient-Reported Outcomes

InFLUenza Patient-Reported Outcome

The draft FLU-PRO Questionnaire instructed respondents to rate the severity of 37 influenza symptoms over the past 24 hours, including those related to the nose, throat, eye, chest, head, stomach, fatigue, and body aches/pains. Six items measured the same symptom using different wording to select the best performing item for the final instrument. For 32 of the 37 items, respondents rated the severity of each symptom on five-point Likert-type scales, with 0 indicating "Not at all"; 1, "A little bit"; 2, "Somewhat"; 3, "Quite a bit"; and 4, "Very much." For the five remaining items, severity is expressed as frequency of occurrence: vomiting or diarrhea (0 time, 1 time, 2 times, 3 times, or 4 or more times). Sneezing, coughing, and coughed-up mucus or phlegm were expressed on a scale from 0 ("Never") to

4 (“Always”), with higher scores indicating more severe symptoms.

Patient Global Rating of Flu Severity

The Patient Global Rating of Flu Severity assesses patients’ perceptions of overall influenza symptom severity scored as 0 (“No flu symptoms today”), 1 (“Mild”), 2 (“Moderate”), 3 (“Severe”), and 4 (“Very severe”).

Patient Global Assessment of Interference in Daily Activities

The Patient Global Assessment of Interference in Daily Activities assesses interference in daily activities because of influenza symptoms scored as 1 (“Not at all”), 2 (“A little bit”), 3 (“Somewhat”), 4 (“Quite a bit”), and 5 (“Very much”).

Patient Global Assessment of Physical Health

The Patient Global Assessment of Physical Health assesses general physical health scored as 1 (“Poor”), 2 (“Fair”), 3 (“Good”), 4 (“Very good), and 5 (“Excellent”).

Return to “usual” health and activities

Patients were asked to respond (yes/no) to the questions “Have you returned to your usual activities today?” and “Have you returned to your usual health today?”

Statistical Analyses

Analyses were based on the classical test theory [16] performed on the influenza-positive cohort with at least 2 days of FLU-PRO data: day 1 and at least 1 day thereafter. Analyses were conducted in two phases:

> Phase I: Finalize the instrument and scoring algorithm, including item-level descriptive statistics, item-to-item correlations, CFA, and EFA [17].
>
> Phase II: Evaluate the psychometric properties of FLU-PRO scores, including reliability, validity, and responsiveness.

Phase I: Finalize the instrument and scoring algorithm

Item analysis. Day 1 data were used to examine distributional characteristics of the items, including mean, median, range, mode, percentages of minimum and maximum responses for floor and ceiling effects, percentage missing, and the frequency and percentage of each response category. Items were flagged for further consideration if they showed floor effects (minimum response >25%) or ceiling effects (maximum response >25%). Spearman correlations were used to calculate interitem correlations among items.

Confirmatory factor analysis. CFA was used to assess the fit of the hypothesized three-domain structure (see Appendix Figure S1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.014): upper respiratory (nose, throat, and eye symptoms), lower respiratory (chest symptoms), and systemic (head, gastrointestinal, sleep, and body/systemic symptoms). The model was tested using a weighted least squares mean and variance-adjusted estimator, with fit assessed using the comparative fit index (CFI), root mean square error of approximation (RMSEA), and weighted root mean square residual (WRMR). CFI greater than 0.90, RMSEA less than 0.07, and WRMR close to 1 [18,19] were considered acceptable. Items with standardized coefficient less than 0.30 were reviewed for possible deletion.

Exploratory factor analysis. EFA was prespecified for follow-up analysis if there was a misfit of the hypothesized model. This EFA accounted for the ordinal categorical nature of the variables, with no prespecified number of factors. Values for CFI, standardized root mean square residual, and RMSEA were examined to assess model goodness of fit. Acceptable model fit was indicated when values of root mean square residual were less than 0.08 [20] and RMSEA were less than 0.07 [18,19]. Approximation of simple structure with factor loadings of 0.4 or higher was the criterion for accepting a factor solution; oblique rotation was used. CFA and EFA were conducted using Mplus software (Muthén & Muthén, Los Angeles, CA) [21].

Phase II: Evaluate psychometric properties

Reliability (internal and test-retest). Cronbach formula for coefficient α was used to estimate internal consistency reliability of the FLU-PRO total and domain scores as appropriate at day 1. Coefficients of 0.7 to 0.9 were prespecified as “good” internal consistency, 0.4 to less than 0.7 as “moderate,” and less than 0.4 as “low” or “poor” [16,22].

Test-retest reliability was estimated using data from patients reporting “no change” on the Patient Global Rating of Change in Flu Severity on two consecutive days from week 1 (day 1–day 7). Intraclass correlation coefficients (ICCs; fixed-effects model), paired t tests, and effect size (ES) were examined. ICCs were expected to be higher than 0.60 and mean differences insignificant with small ES (<0.20).

Construct validity. *Construct validity* is the degree to which scores from one measure are related to those of other measures in a manner consistent with theory. Relationships between FLU-PRO scores and the three global ratings were assessed using Spearman correlations (r), day 1 and day 3, with the expectation that the strongest relationship would be with the patient rating of flu severity, followed by physical health and interference in daily activities, with all coefficients moderate to high (>0.30) [23].

Known-groups validity. Known-groups validity tests score differences between two or more groups known to differ on the underlying construct [24]. In this study, analysis of variance was used to compare FLU-PRO scores across day 1 in Patient Global Rating of Flu Severity categories: “None” or “Mild,” “Moderate,” and “Severe” or “Very severe.” Scheffe test was used for pairwise comparisons.

Hospitalization status was not used to test known-groups validity. Patients can be admitted to the hospital at any time during an influenza episode. Dates of influenza symptom onset, hospital admission, and discharge were not gathered during the study, precluding “day 1 to day 1” between-group comparisons. Patients with influenza are hospitalized for various reasons, including worsening of underlying conditions [25–27]. Because data on comorbid conditions or admitting diagnoses were not available, these factors could not be controlled. Finally, the extent to which influenza symptoms are actually worse in hospitalized versus outpatients has not been shown. Therefore, although the two groups are “known,” the nature of differences (if any) on underlying constructs is not precluding their use as validity indicators.

Responsiveness. To test responsiveness [28], analysis of covariance compared changes in FLU-PRO scores at day 7 in responders (returned to usual health/activity) and nonresponders (not returned to usual health/activity), adjusting for day 1 scores, expecting that responders would have significantly larger ($P < 0.05$) change scores.

Results

Table 1 and S2 presents sample baseline demographic and clinical characteristics.

Phase I: Item Evaluation, Item Reduction, and Domain Structure

Item analysis

The full range of response options was used for all 37 candidate items; 25 items (68%) were flagged for further evaluation because of floor effects. No item reached the ceiling threshold of 25% (see Appendix Table S1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.014).

No interitem correlation coefficient was higher than 0.80, suggesting no redundancy (data not shown). The strongest correlations ($r_s \geq 0.60$) occurred across pairs of related items within systems, for example, chills, shivering, and felt cold (0.73); swollen throat and difficulty swallowing (0.64); chest tightness and trouble breathing (0.62); and chest congestion (0.67). The weakest correlations ($r_s \leq 0.05$) were between items assessing different body systems or logical inconsistencies, such as teary, watery eyes with wet/loose cough (0.04) and coughing mucus (0.01), or diarrhea frequency with sore or painful throat (0.00) or difficulty swallowing (0.01) (see Table S2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.014).

Confirmatory factor analysis

The hypothesized three-factor model with upper respiratory, lower respiratory, and systemic symptoms demonstrated unacceptable global fit (CFI = 0.836; RMSEA = 0.089; WRMR = 1.722). Modification indices suggested a model with several items loading on more than one factor. Because the statistically optimized model would have been difficult to interpret, EFAs were performed to examine a simpler structure more consistent with qualitative data and clinical evaluations of influenza symptoms across multiple body systems.

EFA and item reduction

EFA models with 4 to 15 factors showed acceptable fit indices. A seven-factor solution best approximated the hypothesized conceptual framework, was clinically interpretable, and achieved the best fit (see Appendix Table S3 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.014). This model included domains assessing nose, throat, eyes, chest, head/body, gastrointestinal, and sleep.

Individual items were discussed by the study team, examining item and factor-level performance, insight from patients from the qualitative research, and clinical considerations. Two sleep items were removed ("Difficulty staying asleep" and "Difficulty falling asleep"). "Sleeping more than usual" was retained as the best question to represent sleep disturbance associated with systemic manifestations of influenza.

Revised conceptual framework and scoring

The final conceptual framework for the FLU-PRO (Fig. 1) is a six-factor structure (nose [4 items], throat [3 items], eyes [3 items], chest/respiratory [7 items], gastrointestinal [4 items], and body/systemic [11 items]), with a higher order factor representing influenza symptom severity (CFI = 0.92; RMSEA = 0.06; WRMR = 1.23).

A mean-based scoring algorithm was selected, with total scores calculated by computing means across all 32 items. This yields weighted total scores, dominated by chest/respiratory (22%) and body/systemic (34%) symptoms. To ensure representation of all

Table 1 – Subjects' demographic and clinical characteristics by region: Influenza-positive patients (N = 221).

Variable	Day 1	
	United States (n = 150)	Other countries[*] (n = 71)
Age (y)		
Mean ± SD	39.4 ± 16.1	43.5 ± 17.5
Median (range)	36.0 (18–86)	41.0 (18–95)
>65 years		
Sex: female, n (%)	92 (61.3)	41 (57.7)
Ethnicity, n (%)[†]		
Hispanic or Latino	16 (10.7)	67 (94.4)
Non-Hispanic or Latino	133 (88.7)	4 (5.6)
Race, n (%)		
American Indian or Alaska Native	4 (2.7)	0
Asian	5 (3.3)	0
African American	74 (49.3)	0
Mestizo	0 (0)	67 (94.4)
White	61 (40.7)	4 (5.6)
Other	6 (4.0)	0
Employment status, n (%)		
Employed, full-time or part-time	81 (54.0)	33 (46.5)
Retired	11 (7.3)	3 (4.2)
Other[‡]	36 (24.0)	31 (43.7)
Missing	22 (14.7)	4 (5.6)
Military status, n (%)		
Never in the military	61 (40.7)	67 (94.4)
Active	40 (26.7)	0
Retired	10 (6.7)	0
Other[§]	16 (10.6)	0
Missing	23 (15.3)	4 (5.6)
Highest level of education, n (%)		
Secondary/high school or less	43 (28.7)	29 (40.8)
Some college	33 (22.0)	4 (5.6)
College degree or more	42 (28.0)	31 (43.7)
Other	32 (21.3)	7 (9.9)
Current treatments, n (%)		
Oseltamivir (Tamiflu)	50 (33.3)	13 (18.3)
Amantadine (Symmetrel)	0 (0.0)	2 (2.8)
Other	56 (37.3)	42 (59.2)
None	57 (38.0)	19 (26.8)
Comorbidities[‖], n (%)		
None	56 (37.3)	29 (40.8)
Asthma	38 (25.3)	11 (15.5)
Chronic obstructive pulmonary disease	9 (6.0)	0 (0.0)
Osteoporosis	1 (0.7)	1 (1.4)
Depression	17 (11.3)	4 (5.6)
Hypertension	20 (13.3)	13 (18.3)
Raised cholesterol	12 (8.0)	10 (14.1)
Stomach ulcers	3 (2.0)	3 (4.2)
Heart attack/angina	2 (1.3)	1 (1.4)
Diabetes	22 (14.7)	8 (11.3)
Kidney disease	6 (4.0)	2 (2.8)
Lung disease	3 (2.0)	2 (2.8)
Tuberculosis	0 (0.0)	2 (2.8)
Other	39 (26.0)	17 (23.9)

[*] Other countries include Mexico (n = 67), Argentina (n = 3), and the United Kingdom (n = 1).

[†] One participant had missing ethnicity.

[‡] Other includes homemaker, student, unemployed, and others.

[§] Other includes reserves and others.

[‖] Not mutually exclusive.

body systems in the overall score, total scores are computed only if there are sufficient data to compute each domain score. Scores range from 0 to 4; higher scores indicate more severe symptoms.

Phase II: Evaluation of Psychometric Properties

Overall results for influenza-positive patients are reported herein; results stratified by hospitalization status are provided in Supplemental Materials Tables S6 through S11 and Figure S3 found at http://dx.doi.org/10.1016/j.jval.2017.04.014.

The analytical sample included influenza-positive participants with FLU-PRO diary data on day 1 and at least 1 day thereafter. Missing data increased over time with variance observed by geographic region. Specifically, 53% of US patients completed diaries on days 1, 2, and 3, compared with 89% outside the United States. By day 7, completion rates were 28% and 82% for US and ex-US subjects, respectively. Further analyses showed high rates of compliance during active influenza symptom days, with 90% of subjects completing the diary to symptom resolution, defined by return to usual health or activity. Therefore, most missing data were not informative regarding symptom course.

Descriptive statistics of FLU-PRO total and domain scores

Distributional characteristics of day 1 FLU-PRO scores are presented in Table 2. Figure 2 displays mean FLU-PRO scores over time.

Reliability (internal and test-retest)

Cronbach α was high for all domains (nose = 0.81, throat = 0.81, eyes = 0.81, chest/respiratory = 0.80, gastrointestinal = 0.71, body/systemic = 0.87) and the total score (0.92).

For test-retest reliability, day 1 to day 2 (n = 44), score reliability values for eyes (ICC = 0.62), chest/respiratory (ICC = 0.76), gastrointestinal (ICC = 0.62), and body/systemic (ICC = 0.65) domains were considered acceptable according to the ES and ICC estimates, whereas the nose values (ICC = 0.79) and total score (ICC = 0.83) were acceptable according to the ICC estimate (the throat values did not meet thresholds; ICC = 0.57). All other 2-day assessment points, FLU-PRO ES, and ICC estimates were acceptable (except for body/systemic at day 2–day 3 and day 6–day 7) (see Appendix Table S4 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.014).

Construct Validity

As hypothesized, at day 1 the strongest association was between the FLU-PRO total scores and the Patient Global Rating of Flu Severity ($r = 0.59$; $P < 0.0001$), followed by Interference in Daily Activities ($r = 0.43$; $P < 0.0001$) and Physical Health ($r = -0.29$; $P < 0.0001$) (see Appendix Table S5 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.04.014). Domain scores displayed moderate to large associations with patient ratings of Flu Severity ($r = 0.34$–0.61) except for nose ($r = 0.27$) and throat ($r = 0.28$), with all coefficients statistically significant ($P < 0.0001$). There was a moderate to large correlation between the body/systemic domain and more distal ratings of Interference in Daily Activities ($r = 0.50$; $P < 0.0001$); correlations between this global rating and other FLU-PRO domains were smaller ($r = 0.11$

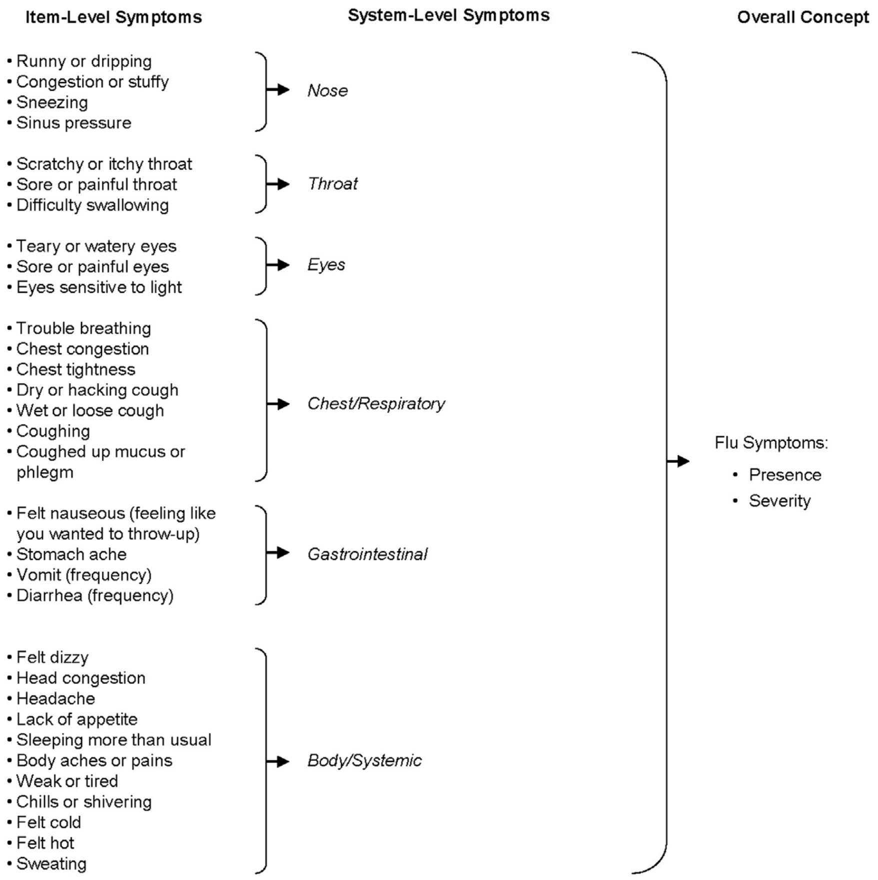

Fig. 1 – Final FLU-PRO conceptual framework. FLU-PRO, inFLUenza Patient-Reported Outcome.

Table 2 – 32-Item FLU-PRO domain and total score descriptive statistics (N = 221), day 1.

Scale	Mean ± SD	Range, median (mode)	Floor effect, n (%)	Ceiling effect, n (%)
Nose	1.7 ± 1.1	0.0–4.0, 1.5 (1.3)	13 (5.9)	4 (1.8)
Throat	1.4 ± 1.1	0.0–4.0, 1.0 (0.0)	41 (18.6)	5 (2.3)
Eyes	1.0 ± 1.1	0.0–4.0, 0.7 (0.0)	67 (30.3)	8 (3.6)
Chest/respiratory	1.9 ± 0.9	0.0–4.0, 1.9 (1.7)	3 (1.4)	2 (0.9)
Gastrointestinal	0.7 ± 0.8	0.0–3.8, 0.3 (0.0)	77 (34.8)	0 (0.0)
Body/systemic	1.8 ± 0.9	0.0–3.8, 1.8 (2.5)	2 (0.9)	0 (0.0)
Total score	1.6 ± 0.7	0.3–3.7, 1.6 (1.3)	0 (0.0)	0 (0.0)

Notes. Higher FLU-PRO scores = more severe symptoms.
FLU-PRO, inFLUenza Patient-Reported Outcome.

[nonsignificant] to 0.29 [$P < 0.0001$]). Similarly, there were weaker associations between domain scores and patients' ratings of Physical Health ($r = 0.06$ [nonsignificant] to 0.28 [$P < 0.0001$]).

Known-Groups Validity

Significant differences in FLU-PRO scores were observed across the known global Flu Severity rating groups ($F = 57.2$; $P < 0.001$). Mean scores were the lowest in the No/Mild symptoms group (0.98 ± 0.47), followed by the Moderate (1.38 ± 0.57) and Severe/very severe (2.01 ± 0.63) groups, with all pairwise comparisons statistically significant ($P < 0.001$). For domain scores, mean values for the No/mild symptoms group were the lowest (mean range = 0.29–1.37), followed by the Moderate (mean range = 0.48–1.75) and Severe/very severe (mean range = 1.06–2.48) groups. Pairwise comparisons showed similar patterns to the total score except for the No/mild symptoms group versus the Moderate group for nose, throat, eyes, and gastrointestinal domains, which were in the correct direction but nonsignificant ($P > 0.05$) (Table 3). This may be due to smaller numbers of patients experiencing these symptoms, reducing variability.

Responsiveness

Mean total and domain change scores were significantly greater for patients reporting return to usual health (responders) by day 7 compared with those who did not, except for the gastrointestinal domain (Table 4). Mean change scores were significantly greater for patients reporting return to usual activities (responders) by day 7 compared with those who did not, except for the eyes domain (Table 4).

Discussion

The purpose of this study was to finalize content, structure, and scoring of the FLU-PRO and assess the performance properties of this new instrument in adults with laboratory-confirmed influenza [10]. A reliable and accurate measurement tool providing a comprehensive profile of influenza symptoms on the basis of concepts that patients have stated are relevant will facilitate conduct of population-level epidemiologic studies, natural history studies, and clinical trials.

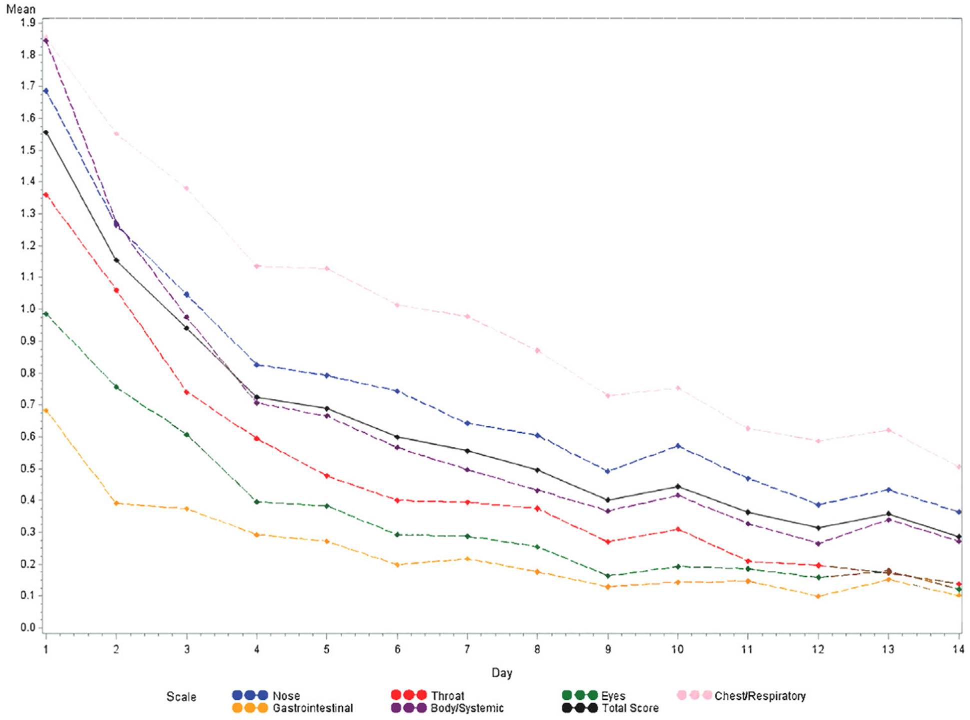

Fig. 2 – FLU-PRO domain and total score by diary day 1–day 14. FLU-PRO, inFLUenza Patient-Reported Outcome.

Table 3 – Known-groups validity: 32-item FLU-PRO scores by Patient Global Rating of Disease Severity, day 1.

Scale	Patient Global Rating of Flu Severity, mean ± SD			F value (P value)	Pairwise comparisons[*]
	No/mild symptoms (n = 50)	Moderate symptoms (n = 77)	Severe/very severe symptoms (n = 94)		
Nose	1.29 ± 0.88	1.56 ± 0.95	2.01 ± 1.15	8.9[†]	2[†], 3[‡]
Throat	0.85 ± 0.83	1.24 ± 1.03	1.73 ± 1.24	11.4[†]	2[†], 3[‡]
Eyes	0.51 ± 0.88	0.82 ± 0.98	1.37 ± 1.19	12.3[†]	2[†], 3[§]
Chest/respiratory	1.37 ± 0.69	1.75 ± 0.86	2.20 ± 0.86	17.5[†]	1[‡], 2[†], 3[§]
Gastrointestinal	0.29 ± 0.43	0.48 ± 0.65	1.06 ± 1.00	19.8[†]	2[†], 3[†]
Body/systemic	1.03 ± 0.64	1.60 ± 0.78	2.48 ± 0.77	67.5[†]	1[†], 2[†], 3[†]
Total score	0.98 ± 0.47	1.38 ± 0.57	2.01 ± 0.63	57.2[†]	1[†], 2[†], 3[†]

Notes. Higher FLU-PRO scores = more severe symptoms.

FLU-PRO, inFLUenza Patient-Reported Outcome.

[*] Pairwise comparisons between means will be performed using Scheffe test adjusting for multiple comparisons: 1 = No/mild symptoms vs. Moderate; 2 = No/mild symptoms vs. Severe and Very severe; and 3 = Moderate symptoms vs. Severe and Very severe.

[†] $P < 0.001$.

[‡] $P < 0.05$.

[§] $P < 0.01$.

A 37-item draft measure was developed on the basis of patient descriptions of influenza and included content-redundant items for evaluation and elimination during quantitative analysis [13]. Five redundant and lower performing items were removed on the basis of qualitative and quantitative information to yield the final 32-item questionnaire. Patients participating in cognitive interviews [13] found the 37-item questionnaire easy to complete with uninterrupted response times of 5 minutes, suggesting that the final instrument should perform similarly. The six-domain/subscale structure of the FLU-PRO is clinically intuitive and consistent with body systems commonly affected by influenza. The mean-based scoring algorithm is easy to use and interpret, with higher scores indicating more severe symptoms. Each body system/domain is represented in the score profile, with the two domains most important to patients and clinicians, chest/respiratory and systemic symptoms, more heavily weighted in the total score.

Results suggest that FLU-PRO scores are reliable, valid, and responsive to improvements in health as patients recover from influenza. Consistent with a priori hypotheses, scores were significantly related to patient global ratings of influenza severity, interference in daily activities, and physical health. Correlations were similar or higher than those reported for the Influenza Intensity and Impact Questionnaire (FluiiQ), in which day 1 Spearman correlations between FluiiQ total score and global ratings of severity, the feeling thermometer, and health were 0.44, −0.27, and 0.23 [6]. For the FluiiQ respiratory domain, correlations with the three criterion measures were 0.35, −0.15, and 0.08; coefficients for the systemic domain were 0.41, −0.27, and 0.24. The FLU-PRO data supported known-groups validity because scores were the lowest in patients rating No/mild symptoms, higher in the Moderate group, and the highest in the Severe/very severe group. Finally, the FLU-PRO demonstrated responsiveness to change from day 1 to day 7, with responders defined by reports of return to usual health and activities.

The purpose of this study was to perform empirically based item reduction, test the factor structure of the measure, develop a scoring algorithm, and evaluate the instrument for reliability and responsiveness associated with recovery. Global ratings were used in the responsiveness analyses, similar to the approach used by others to estimate cross-sectional validity [6]. We did not ask patients to record response to the FluiiQ because we used the same validation method as that instrument, and asking ill patients with influenza to complete 61 questions daily (37-item FLU-PRO pool and 24-item FluiiQ) would have been burdensome, potentially causing missing data and larger dropout rates. FLU-PRO may be more comprehensive, because FluiiQ assesses two domains, respiratory and systemic, whereas the FLU-PRO assesses six body systems. The systemic subscales of the two instruments include some common content, with 4 of the FLU-PRO's 11 systemic symptoms ($\alpha = 0.87$) also represented in the FluiiQ's seven-item systemic scale ($\alpha = 0.85$) [6], suggesting that these subscales may have high correlation. The respiratory scales of the two measures are, however, different. FluiiQ respiratory domain includes three items (cough, sore throat, and nasal congestion) ($\alpha = 0.48$) [6]. The FLU-PRO yields separate scores for nose ($\alpha = 0.81$), throat ($\alpha = 0.81$), and chest/respiratory ($\alpha = 0.80$), with the latter assessing trouble breathing, chest congestion, chest tightness, and four questions related to cough and mucus. The content and internal consistency estimates suggest that the FLU-PRO chest/respiratory domain provides a more precise picture of the respiratory symptoms of influenza than does the FluiiQ.

The present study had several limitations. Hospitalized patients were included in the validation patient population, but specific details about hospitalization (e.g., duration of influenza before hospitalization, acuity level during hospitalization, and concurrent complicating conditions) are unknown. Results suggest that the FLU-PRO performs consistently in hospitalized and clinic-based samples; nevertheless, additional study in both groups is warranted. Missing data increased over time with geographic variance in the rate of missing data. This may be due, in part, to the interviewer-administered methods used in several ex-US sites. The high rate of compliance through symptom resolution suggests that the most relevant days of influenza episodes were captured and missing data were not informative. One of the strengths of the study is international participation but this necessitated several modes of administration to meet local data collection needs and preferences. Because mode was nested under country (e.g., interviewer-administered questionnaires in Mexico and self-administration in the United States), effects for mode and country could not be factored out and tested. Finally, because this was a naturalistic study to validate the instrument and not a randomized trial to determine treatment effects of interventions, the relationship between symptom severity and medication use is indeterminate; that is, subjects

Table 4 – Responsiveness of 32-item FLU-PRO by patient return to usual health (N = 147)* or return to usual activities (N = 126)†, day 1–day 7.

Scale	Responders‡			Nonresponders			P value
	Day 1, mean ± SD	Day 7, mean ± SD	Change score, LS mean ± SD	Day 1, mean ± SD	Day 7, mean ± SD	Change score, LS mean ± SD	
Nose							
Usual health	1.7 ± 1.1	0.4 ± 0.5	1.3 ± 0.1	1.6 ± 1.1	0.8 ± 0.7	0.8 ± 0.1	<0.0001
Usual activities	1.8 ± 1.0	0.6 ± 0.6	1.1 ± 0.1	1.3 ± 1.2	0.7 ± 1.0	0.8 ± 0.1	0.0375
Throat							
Usual health	1.1 ± 1.0	0.1 ± 0.3	1.2 ± 0.1	1.5 ± 1.2	0.5 ± 0.7	0.9 ± 0.1	0.0010
Usual activities	1.4 ± 1.1	0.3 ± 0.5	1.1 ± 0.1	1.6 ± 1.3	0.6 ± 0.9	0.8 ± 0.1	0.0244
Eyes							
Usual health	1.0 ± 1.0	0.1 ± 0.4	0.9 ± 0.1	1.1 ± 1.2	0.4 ± 0.7	0.7 ± 0.1	0.0452
Usual activities	1.0 ± 1.1	0.2 ± 0.6	0.9 ± 0.1	1.3 ± 1.3	0.5 ± 0.8	0.7 ± 0.1	0.1166
Chest/ respiratory							
Usual health	1.4 ± 0.8	0.5 ± 0.6	1.1 ± 0.1	2.0 ± 0.8	1.2 ± 0.7)	0.7 ± 0.1	<0.0001
Usual activities	1.8 ± 0.8	0.8 ± 0.7	1.0 ± 0.1	2.0 ± 0.9	1.4 ± 0.8	0.6 ± 0.1	0.0003
Gastrointestinal							
Usual health	0.5 ± 0.8	0.1 ± 0.5	0.5 ± 0.1	0.7 ± 0.8	0.3 ± 0.4	0.4 ± 0.0	0.2062
Usual activities	0.7 ± 0.8	0.2 ± 0.4	0.5 ± 0.0	0.6 ± 0.9	0.4 ± 0.5	0.3 ± 0.1	0.0169
Body/systemic							
Usual health	1.6 ± 0.9	0.2 ± 0.5	1.5 ± 0.1	1.9 ± 1.0	0.6 ± 0.6	1.2 ± 0.1	0.0004
Usual activities	1.9 ± 0.9	0.4 ± 0.5	1.5 ± 0.1	1.9 ± 1.0	0.9 ± 0.8	1.0 ± 0.1	<0.0001
Total score							
Usual health	1.3 ± 0.6	0.3 ± 0.4	1.2 ± 0.1	1.6 ± 0.8	0.7 ± 0.5	0.8 ± 0.0	<0.0001
Usual activities	1.6 ± 0.7	0.5 ± 0.4	1.1 ± 0.0	1.6 ± 0.8	0.8 ± 0.6	0.7 ± 0.1	<0.0001

FLU-PRO, inFLUenza Patient-Reported Outcome; LS, least squares.

* Responders: N = 51; nonresponders: N = 96.

† Responders: N = 87; nonresponders: N = 39.

‡ Responders are defined as patients responding that they have returned to their usual health or usual activities at day 7.

using one or more drugs could have more or fewer symptoms. This would, however, not affect evaluation of the instrument itself. Differences in symptom occurrence and severity between influenza and ILI are not definitely known. Thus, neither variable could be used to test known-groups validity. Future articles will describe the use of the FLU-PRO in patients who test negative for influenza. Further use and testing of the FLU-PRO in randomized controlled trials, varied treatment settings, and epidemiologic studies is warranted.

The content validity of the FLU-PRO has been established in children and adolescents through qualitative research [13]. Quantitative testing in this population is warranted. Future research using the FLU-PRO in influenza challenge studies will provide data on the full course of influenza, from the pre-influenza asymptomatic state to symptom resolution. The FLU-PRO is also being evaluated for use in disease due to other acute respiratory viruses.

Conclusions

This study used quantitative methods to develop and test the FLU-PRO for evaluating patient-reported symptoms in patients with influenza. This new instrument yields a profile of scores across six body systems with a total score reflecting overall symptom severity. Results suggest that FLU-PRO scores are reliable, valid, and responsive to change in adults with laboratory-confirmed influenza. The profile can be used to understand and compare symptomatic manifestations of various influenza strains, dominant systems across settings or patient subgroups, and patterns of change over time, including differential symptom onset and recovery patterns. The instrument is available for further testing and use as a standardized method for evaluating symptoms of influenza in natural history studies and clinical trials.

Acknowledgments

We extend sincere appreciation to the staff who provided research support at the clinical sites and to Sonja Stringer, Katherine Kim, David Hengerer, Sandra Macker, and Hongyan Shen of Evidera for their assistance with data collection, management, and analyses. We thank Kathryn Miller for her assistance with the editing and formatting of the manuscript. We also express our gratitude to the patients who gave their time to help complete this study.

Source of financial support: This project was funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute of Allergy and Infectious Diseases.

The contents of this publication are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions or policies of Uniformed Services University of the Health Sciences (USUHS), the Department of Defense (DoD), the Departments of the Army, Navy, or Air Force, or the Henry M. Jackson Foundation for the Advancement of Military Medicine. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

The views expressed are those of the author(s) and do not reflect the official policy or position of the U.S. Army Medical Department, Department of the Army, Department of Defense or the U.S. Government. The investigators have adhered to the policies for protection of human subjects as prescribed in 45CRF46.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2017.04.014 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Centers for Disease Control and Prevention. Seasonal influenza (flu) in the workplace. Available from: http://www.cdc.gov/niosh/topics/flu/. [Accessed June 1, 2015].

[2] Centers for Disease Control and Prevention. Seasonal influenza (flu). Available from: http://www.cdc.gov/flu/about/disease/index.htm. [Accessed January 28, 2016].

[3] Centers for Disease Control and Prevention. Key facts about influenza (flu) and flu vaccine. Available from: http://www.cdc.gov/flu/keyfacts.htm. [Accessed January 28, 2016].

[4] World Health Organization. Influenza (seasonal): fact sheet no. 211 2014. http://www.who.int/mediacentre/factsheets/fs211/en/. [Accessed June 1, 2015].

[5] Osborne R, Hawthorne G, Papanicolaou M, Wegmueller Y. Measurement of rapid changes in health outcomes in people with influenza symptoms. J Outcomes Res 2000;4:15–30.

[6] Osborne RH, Norquist JM, Elsworth GR, et al. Development and validation of the Influenza Intensity and Impact Questionnaire (FluiiQ). Value Health 2011;14:687–99.

[7] Frost MH, Reeve BB, Liepa AM, et al. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health 2007;10(Suppl. 2):S94–105.

[8] Revicki DA. Regulatory Issues and Patient-Reported Outcomes Task Force for the International Society for Quality of Life Research. FDA draft guidance and health-outcomes research. Lancet 2007;369:540–2.

[9] Rothman M, Burke L, Erickson P, et al. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. Value Health 2009;12:1075–83.

[10] Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. Fed Regist 2009;74:65132–3.

[11] Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 2—assessing respondent understanding. Value Health 2011;14:978–88.

[12] Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 1—eliciting concepts for a new PRO instrument. Value Health 2011;14:967–77.

[13] Powers JH, Guerrero ML, Leidy NK, et al. Development of the Flu-PRO: a patient-reported outcome (PRO) instrument to evaluate symptoms of influenza. BMC Infect Dis 2015;16:1.

[14] Anthoine E, Moret L, Regnault A, et al. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes 2014;12:176.

[15] Wild D, Grove A, Martin M, et al. ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8:94–104.

[16] Nunnally JC, Bernstein IH. Psychometric Theory (3rd ed.). New York, NY: McGraw-Hill, 1994.

[17] Gorsuch RL. Factor Analysis. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

[18] Steiger JH. Understanding the limitations of global fit assessment in structural equation modeling. Pers Individ Dif 2007;42:893–8.

[19] Yu C-Y. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Los Angeles, CA: Doctoral dissertation, University of California, 2002.

[20] Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6:1–55.

[21] Muthén LK, Muthén BO,editors. Mplus User's Guide. (3rd ed.). Los Angeles, CA: Muthén & Muthén, 2004.

[22] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.

[23] Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

[24] Stewart AL, Hays RD, Ware JE. Methods of Validating MOS Health Measures. Durham, NC: Duke University Press, 1992.

[25] Arriola CS, Anderson EJ, Baumbach J, et al. Does influenza vaccination modify influenza severity? Data on older adults hospitalized with influenza during the 2012–2013 season in the United States. J Infect Dis 2015;212:1200–8.

[26] Ono S, Ono Y, Matsui H, Yasunaga H. Factors associated with hospitalization for seasonal influenza in a Japanese nonelderly cohort. BMC Public Health 2016;16:922.

[27] van Essen GA, Beran J, Devaster JM, et al. Influenza symptoms and their impact on elderly adults: randomised trial of AS03-adjuvanted or non-adjuvanted inactivated trivalent seasonal influenza vaccines. Influenza Other Respir Viruses 2014;8:452–62.

[28] Hays RD, Revicki D. Reliability and validity (including responsiveness). In: Fayers P, Hays RD, eds. Assessing Quality of Life in Clinical Trials: Methods and Practice. New York, NY: Oxford University Press, 2005. p. 25–39.