Evaluation of the Performance Properties of the InFLUenza Patient-Reported Outcome (FLU-PRO©) Instrument

Authors: 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]; Erin A Hansen, MS[8,12]; Michelande Ridoré, MS[13]; Timothy H. Burgess, MD, MPH[7,14]; Eugene V. Millar, PhD[7,8]; Andrés Hernández, MD, MS[15]; Patricia Rodríguez-Zulueta, MD[16]; Mary C. Smolskis, RN, MA[4]; Hilda Ortega-Gallegos, BS[3]; Sarah Pett, BSc, MBBS[17]; William Fischer, MD[18]; Daniel Gillor, Dr. Med[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]

Affiliations: [1]Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD in support of the Collaborative Clinical Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, Bethesda, 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 Institutes of Allergy and Infectious Diseases, NIH, Bethesda, MD, USA; [6]Madigan Army Medical Center, Fort Lewis, WA, USA; [7]Infectious Disease Clinical Research Program, Department of Preventive Medicine and Biostatistics, Uniformed Services University, Bethesda MD, USA; [8]Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, 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]Naval Health Research Center, San Diego, CA, USA; [13]Children's National Medical Center, Washington, DC, USA; [14]Walter Reed National Military Medical Center, Bethesda, MD, USA; [15]Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas , Mexico City, Mexico; [16]Hospital General Dr. Manuel Gea González, Mexico City, Mexico; [17]University College London, London, UK; [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

***Corresponding Author**
John H. Powers, MD
5601 Fishers Lane
Room 4D50
Rockville, MD 20852
Phone: 301-768-6711
Fax: 240-465-0396
E-mail: john.powers@nih.gov

Funding: This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. This research was supported in part by the Intramural Research Program of the NIH, NIAID. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Keywords: influenza; patient-reported outcome; psychometric; reliability; validity; responsiveness

Running title: FLU-PRO Validation

Acknowledgements: We would like to extend sincere appreciation to staff providing 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; and to Kathryn Miller for her assistance with manuscript text edit and formatting. We also express our gratitude to the patients who gave of their time to help complete this study.

HIGHLIGHTS

i. What is already known about the topic?

i. Influenza causes substantial burden to patients. Most patients who are infected with influenza do not die or experience serious complications but nearly all experience a bothersome array of symptoms.

ii. There is no standardized measure of influenza symptoms developed according to current best practices and scientific methodology.

ii. What does the paper add to existing knowledge?

i. A standardized measure of influenza symptoms could be used in clinical trials evaluating medical interventions for treatment or prevention, epidemiology and natural history studies, and transmission and challenge studies. A standardized symptom measure could also be combined with other baseline variables.

ii. Prior work has evaluated the content validity and understandability of the FLU-PRO symptom scale. The evidence from this study on the measurement properties of the FLU-PRO shows that it is reliable, has known-groups and construct validity, and demonstrates responsiveness to change over time as symptoms change in outpatients and hospitalized patients.

iii. (optional) What insights does the paper provide for informing health care-related decision making?

i. FLU-PRO is ready for use in clinical trial and in epidemiology studies. FLU-PRO provides a valid, reliable and responsive standardized measure of influenza symptoms that can be used to evaluate patient-reported outcomes either alone or in combination with other outcomes to accurately assess the burden of illness in epidemiological studies and treatment effects in clinical trials of medical interventions.

ABSTRACT (MAX: 250 WORDS) – CURRENT: 249

Objectives: Assess the reliability, validity, and responsiveness of FLU-PRO© scores for quantifying the presence and severity of influenza symptoms.

Methods: Observational prospective cohort study of adults (≥18 years) with influenza-like illness in US, UK, Mexico, and South America. 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's alpha and intraclass correlation coefficients (ICC; 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 536 enrolled, N=221 influenza-positive subjects comprised the analytical sample. Mean age=40.7, 60.2% female, 59.7% white. The final 32-item measure has 6-factors/domains (Nose Throat, Eyes, Chest/Respiratory, Gastrointestinal, Body/Systemic) with a higher-order factor representing symptom severity overall (comparative fit index [CFI]=0.92; root mean square error of approximation [RMSEA]=0.06). Cronbach's alpha was high (Total=0.92; domain range=0.71-0.87; test-retest reliability (ICC, Day 1–2) 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 based on global rating (Total; $F=57.2$, $p<0.001$; domains: $F=8.9$-67.5, $p<0.001$). Subjects reporting return to usual health showed significantly greater ($p<0.05$) FLU-PRO score improvement Day 7 than those who did not, suggesting score responsiveness.

Conclusions: Results suggest FLU-PRO scores are reliable, valid, and responsive to change in influenza positive adults.

INTRODUCTION (Total Word Limit: 4000; Current: 4,003)

Influenza (flu) is an acute illness caused by influenza viruses [1]. Symptoms can range from mild to severe, and include fever/chills, cough, sore throat, runny or stuffy nose, fatigue, muscle/body aches, with gastrointestinal symptoms (diarrhea and vomiting) occurring less frequently [2].

In the United States (US), approximately 5%–20% of the population is infected with influenza yearly, with approximately 200,000 hospitalizations and 36,000 deaths [1-3]. Worldwide, influenza causes approximately 3–5 million severe cases and 250,000–500,000 deaths annually [4]. While most patients recover, during their illness these patients experience symptoms that impair their daily functioning.

Currently, no standardized, validated patient-reported outcome (PRO) measure for influenza and influenza-like symptoms has been developed using good research practices for scale development methods [6-8], such as those recommended by the US Food and Drug Administration (FDA) [5]. A standardized patient-reported influenza symptom scale would allow for consistent, accurate assessments of the characteristic symptoms associated with various viral strains and their corresponding severity in population-level epidemiologic studies, natural history studies on the course of influenza, studies comparing influenza subtypes within and across years, and clinical trials. For treatment trials, a precise, standardized patient-reported influenza symptom scale will allow clinical trials to detect differences between interventions with greater accuracy and facilitate cross-product evaluations and meta-analysis.

The InFLUenza Patient-Reported Outcome (FLU-PRO©) measure was designed to assess the occurrence and severity of influenza symptoms; used as a daily diary, the FLU-PRO can track changes over time, during the course of an influenza episode. Two-stage qualitative instrument development methodology was used to create this new measure [9]. Stage I involved concept elicitation interviews in the US and Mexico to gather information regarding patient experience of influenza symptoms (i.e., type, magnitude, expression, pattern of onset, and recovery). Results informed the development of the draft FLU-PRO, including content, structure (item phrasing, length, response options, recall, instructions), and conceptual framework. Stage II consisted of cognitive interviews to assess completeness,

comprehension, interpretability, and ease of use of the draft measure from the respondent's perspective, refining the instrument to assure content validity.

This study assessed the performance properties of the FLU-PRO in adults ≥18 years of age with acute laboratory-confirmed influenza. Specifically, the objectives of this study were to: 1) evaluate individual item performance and measurement/domain structure; 2) reduce the number of items as empirically and conceptually appropriate; 3) develop a scoring algorithm; and 4) assess the reliability, construct and known-groups 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 non-hospitalized adults ≥18 years of age with acute influenza. Patients seeking care for influenza symptoms at participating clinics in the US (16 sites), Argentina (two sites), United Kingdom (one site), and Mexico (three sites) were recruited during clinic visits. Influenza status was assessed through a positive PCR, rapid antigen test, and/or viral culture by nasal or nasopharyngeal swab.

Procedures

Clinical research coordinators recruited all participants with influenza-like symptoms and tested for laboratory confirmed influenza diagnosis to determine the primary analytical sample. The FLU-PRO study sample was recruited as part of a larger outpatient study. Consented patients completed clinic-based baseline assessments of sociodemographic and clinical characteristics. Patients completed a daily diary for 14 days following enrollment that included the 37-item draft FLU-PRO symptom diary and nine additional questions for validation purposes. At the Mexico site, the diary was completed via personal telephone interview with data entered directly into a web-based portal. Patients in 16 US sites, one UK site, and two Argentina sites completed the survey either via interviewer-administration or a via web-based system using the subject's personal web-enabled device,

The study was conducted with informed consent, under institutional review board approval, and in accordance with the Declaration of Helsinki.

Instruments: Patient-reported Outcomes (PROs)

InFLUenza Patient-Reported Outcome (FLU-PRO©)

The draft FLU-PRO Questionnaire instructed respondents to rate the severity of 37 influenza symptoms over the past 24 hours. Symptoms included those related to the nose, throat, eye, chest, head, stomach, fatigue, and body aches/pains based on concepts elicited from patients in Stage I. Six items measured the same symptom using different wording to select the best performing item for the final instrument. For 32 of 37 items, respondents rated the severity of each symptom on a 5-point Likert-type scale from 0 ("Not at all), 1 ("A little bit"), 2 ("Somewhat"), 3 ("Quite a bit"), to 4 ("Very much"). For the five remaining items, symptom severity is expressed in terms of frequency of occurrence: vomiting or diarrhea (0 times, 1 time, 2 times, 3 times, or 4 or more times), and sneezing, coughing, and coughed up mucus or phlegm on a scale from 0 ("Never") to 4 ("Always"), with higher scores indicating more severe symptoms.

The questionnaire was developed for self-report or interviewer-administration, with slight differences in the instructions applicable for each administration.

Patient Global Rating of Flu Severity

The Patient Global Rating of Flu Severity is a single item to assess participants' overall influenza symptom severity. Participants were asked to rate severity on the following scale: 0 ("No flu symptoms today"), 1 ("Mild"), 2 ("Moderate"), 3 ("Severe"), and 4 ("Very severe").

Patient Global Assessment of Interference with Daily Activities

The Patient Global Assessment of Interference in Daily Activities is a single item to assess interference in daily activities due to influenza symptoms during that day. Participants rated interference on the following scale: 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 Health is a single item to assess general physical health during that day. Participants rated their physical health on the following scale: 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 following questions: “Have you returned to your usual activities today?” and “Have you returned to your usual health today?”

Statistical Analyses

Statistical tests were performed in accordance with classical test theory [10]. Analyses were conducted in two phases:

> Phase I: Item evaluation and item reduction, including descriptive item statistics, floor and ceiling effects, item-to-item correlations, confirmatory factor analysis (CFA), and exploratory factor analysis (EFA) [11]. Analyses were performed on data from the entire influenza positive cohort. Results from these analyses were used to inform item deletion or retention and determine the scoring algorithm for the final FLU-PRO.
>
> Phase II: Evaluation of psychometric properties of the FLU-PRO total and domain scores, including reliability, construct and known-groups validity, and responsiveness. These analyses were performed on the entire influenza positive cohort and stratified by hospitalization status.

Phase I: Item Evaluation, Item Reduction, and Domain Structure of the FLU-PRO

Item Analysis

Day 1 data were used to examine distributional characteristics of the 37 items comprising the draft FLU-PRO, including mean, median, range, mode, percentages of minimum and maximum responses for floor and ceiling effects, percentage missing, and the frequency and percent of each response category. An item was flagged for potential problems if it showed a floor (minimum response >25%) or ceiling effect

(maximum response >25%). Spearman correlations were used to calculate inter-item correlations among all 37 FLU-PRO items at day 1.

Confirmatory Factor Analysis (CFA)

CFA was used to assess fit of FLU-PRO according to a hypothesized 3-domain structure (Supplement Figure S1), including Upper Respiratory (items measuring nose, throat, and eye symptoms), Lower Respiratory (items measuring chest symptoms), and Systemic (items measuring head, gastrointestinal, sleep, and body/systemic symptoms) domains on day 1.

The hypothesized factor model was tested using a weighted least squares mean and variance adjusted (WLSMV) estimator. The CFA model fit was assessed with the comparative fit index (CFI), root mean square error of approximation (RMSEA), and weighted root mean square residual (WRMR). CFI greater than 0.90 was considered an acceptable fit, RMSEA <0.07, and WRMR close to 1 [12, 13]. Items with standardized coefficient <0.30 were reviewed for possible deletion. CFA was conducted using Mplus software [14].

Exploratory Factor Analysis (EFA)

EFA was conducted in the case of misfit of the hypothesized model according to the CFA. For the EFA, there was no pre-specified number of factors. Values for CFI, standardized root mean square residual (SRMR), and RMSEA were examined to assess model goodness-of-fit. Acceptable model-fit was indicated when values of SRMR <0.08 [15] and RMSEA <0.07 [12, 13]. Approximation of simple structure with factor loadings ≥0.4 was the criterion for accepting a factor solution; oblique rotation was used. EFA was conducted using Mplus software [14].

Phase II: Evaluation of Psychometric Properties

Reliability (Internal and Test-retest)

Cronbach's formula for coefficient alpha 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 pre-specified as "good" internal consistency, 0.4 to <0.7 as moderate, and <0.4 as low or poor [10, 16].

Data from patients whose influenza severity state was unchanged over time were used to estimate the test-retest reliability of FLU-PRO Total and domain scores. Stable subjects were defined as those with "no change" on the Patient Global Rating of Change in Flu Severity using two consecutive days from Week 1 (day 1 to day 7). If a subject was missing FLU-PRO scores for one of the days in the planned comparison, data for this subject was excluded from that pair of days' evaluation. Intraclass correlation coefficients (ICC from a fixed effects model) along with paired *t*-tests, and effect size (ES) were performed to evaluate score stability. ICCs were expected to be at least moderate, exceeding 0.60. Mean differences between the two observations were expected to be minimal with small ES (<0.20);

Construct Validity

Construct validity is the degree to which a measure is related to other measures or constructs in a manner that is consistent with theory. The relationship between the FLU-PRO Total and domain scores and three global ratings were assessed using Spearman correlations to test for construct validity at day 1 and day 3. Correlations between the FLU-PRO were anticipated to be the strongest with the Patient Global Rating of Flu Severity, followed by the Patient Global Rating of Physical Health, and the Patient Global Assessment of Interference with Daily Activities, which were hypothesized to be the more distal constructs. However, it was hypothesized that correlations between the FLU-PRO and all three global ratings would be moderate to high (>0.30) to support construct validity [17].

Known-groups Validity

Known-groups validity involves evaluating an instrument in relation to score differences between two or more groups known to differ on the underlying construct. [18]. In this case, analysis of variance (ANOVA) was used to compare FLU-PRO Total and domain scores across three Patient Global Rating of Flu Severity categories at day 1: "None" or "Mild"; "Moderate"; and "Severe" or "Very severe". Mean (SD), *F*-scores, and p-values were reported to determine the magnitude of the differences. Pairwise comparisons between means were performed using Scheffe's test adjusting for multiple comparisons.

Responsiveness

Ability to detect change refers to the extent to which the PRO instrument can detect change in patients whose clinical status has changed [19]. Analysis of covariance (ANCOVA) was used to compare changes in FLU-PRO scores at day 7 in the responders (those returning to usual health or activity) and non-responders (those not returning to usual health or activity) adjusting for day 1 scores. Responders were defined using the two different anchors in two separate analyses. It was expected that responders would have significantly larger ($p < 0.05$) change scores than non-responders.

RESULTS

Sample

A total of 536 English and Spanish-speaking hospitalized and non-hospitalized patients were enrolled, with 441 included in the analytic dataset (had day 1 diary assessment and $\geq$1 post-day 1 diary entry). Of these 441, 221 were influenza positive and included in the psychometric analyses (Supplement Figure S2). Table 1 presents baseline demographic and clinical characteristics for influenza positive patients.

Phase I: Item Evaluation, Item Reduction, and Domain Structure of the FLU-PRO

Item Analysis

The distributional characteristics of FLU-PRO items were examined at day 1. The full range of response options was utilized for all 37 items; 25 items were flagged for further evaluation due to floor effects. (Supplement Table S1).

Spearman inter-item correlation coefficients among FLU-PRO items at day 1 correlated as expected; no correlations were above 0.80 to indicate a high level of redundancy between items (data not shown).

Confirmatory Factor Analysis

A CFA was conducted to assess the fit of the original hypothesized 3-factor model, with factors for Upper Respiratory, Lower Respiratory, and Systemic symptoms (Supplement Table S2). This model demonstrated unacceptable global fit (CFA = 0.836; RMSEA = 0.089; WRMR = 1.722) and one item demonstrated misfit (Q23: Sleeping more than usual).

Exploratory Factor Analysis and Item Reduction

Given the CFA result, EFA was conducted to allow the data to drive the factor structure. Models with 4–15 factors showed acceptable fit indices. However, the 7-factor solution best approximated the hypothesized conceptual framework, was clinically interpretable, and achieved the best fit (Supplement Table S3). The 7-factor solution was composed of Nose, Throat, Eyes, Chest, Head/Body, Gastrointestinal, and Sleep domains.

Individual items were further examined relative to their item-level performance, this factor structure, and qualitative data gathered during Phase I. The sleep domain was removed, with the remaining sleep item incorporated into the Body/Systemic domain. The items Q24: Difficulty staying asleep and Q25: Difficulty falling asleep were removed due to poor fit in modeling as these items represent opposite problems with sleep.

Revised Conceptual Framework and Scoring

The final conceptual framework for the FLU-PRO is shown in Figure 1. The final scaling model is composed of a 6-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 and has the following fit indices: CFI = 0.92; RMSEA = 0.06, and WRMR = 1.23.

A mean-based scoring algorithm was selected, where domain and total scores can be calculated by computing the mean within each domain or across all 32 items, respectively. Scores range from 0 to 4 with higher scores indicating more severe symptoms.

Phase II: Evaluation of Psychometric Properties

Results for the influenza positive patient sample overall are reported below; results of stratified analyses are provided in the online supplement.

Descriptive Statistics of FLU-PRO Total and Domain Scores

The distributional characteristics of the FLU-PRO domain and Total scores were examined at day 1 (Table 2). Mean domain scores ranged from 0.7 (SD=0.8) for the Gastrointestinal domain to 1.9 (SD=0.9) for the Chest/Respiratory domain. Floor effects were observed for the Eyes (30%) and Gastrointestinal (35%) domains, but no ceiling effects were evident. Figure 2 shows the decline in mean FLU-PRO Total and domain scores day 1 to day 14.

Reliability (Internal and Test-retest)

Cronbach's alpha 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). Only the removal of two items would have increased domain score alpha (Q6: Difficulty swallowing; Q34: How many times did you have diarrhea?); however, this change was insufficient to warrant removal from the final instrument.

Data from patients whose global rating of influenza severity was unchanged over the analytical time period were used to evaluate reliability. From day 1 to day 2 (n=44), score reliability 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, while the Nose (ICC=0.79) and Total score (ICC=0.83) were acceptable according to the ICC estimate (Throat values did not meet thresholds; ICC=0.57). At all other two-day assessment points, FLU-PRO ES and ICCs estimates were acceptable (except Body/Systemic at day 2 to 3 and 6 to 7) (Supplement Table S4).

Construct Validity

As hypothesized, at day 1 the strongest association was evidenced between the FLU-PRO Total score and the Patient Global Rating of Flu Severity ($r=0.59$, $p<0.0001$), followed by the Patient Global Assessment of Interference in Daily Activities ($r=0.43$, $p<0.0001$) and Patient Global Rating of Physical Health ($r=-0.29$, $p<0.0001$) (Supplement Table S5). Domain scores displayed moderate to large associations with the Patient Global Rating of Flu Severity ($r=0.34$ to 0.61) with the exception Nose ($r=0.27$) and Throat ($r=0.28$), with all coefficients statistically significant ($p<0.0001$). A moderate to large correlation was demonstrated between the Body/Systemic domain and the more distal Patient Global

Assessment of Interference with Daily Activities (r=0.50, p<0.0001); correlations between this global rating and the other FLU-PRO domains were smaller (r=0.11 (n.s.) to 0.29 (p<0.0001)). Similarly, weaker associations were demonstrated between the FLU-PRO domains and the Patient Global Rating of Physical Health with coefficients ranging from 0.06 (n.s.) to 0.28 (p<0.0001).

Known-groups Validity

Significant differences in FLU-PRO scores were observed across the patient global symptom severity rating groups (F=57.2, p<0.001). Mean [SD] scores were lowest in the No/Mild Symptoms group (0.98 [0.47]), followed by the Moderate (1.38 [0.57]) and Severe/Very Severe groups (2.01 [0.63]) with all pairwise comparisons statistically significant (p<0.001). For the FLU-PRO domain scores, the mean values for the No/Mild group were the lowest (mean range = 0.29 to 1.37), followed by the Moderate group (mean range = 0.48 to 1.75) and the Severe/Very Severe group (mean range = 1.06 to 2.48). Pairwise comparisons for each domain score showed a similar pattern to the Total score with the exception of the No/Mild symptoms versus Moderate for the Nose, Throat, Eyes, and Gastrointestinal domains, which were in the correct direction but nonsignificant (p>0.05) (Table 3).

Responsiveness

Mean total and domain change scores were significantly greater for patients reporting a return to usual health (responders) by day 7, compared to those who did not, with the exception of the Gastrointestinal domain (Table 4). Mean change scores were also significantly greater for patients reporting return to usual activities (responders) by day 7 compared to those who did not, with the exception of 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 [5]. A reliable and accurate measurement tool to quantify symptoms of influenza in hospitalized and non-hospitalized

patients will facilitate the conduct of population-level epidemiologic studies, natural history studies, and clinical trials.

The 37-item draft FLU-PRO was developed based on patient descriptions of influenza and included content-redundant items for evaluation and elimination during quantitative analysis [9]. Five redundant and lower-performing items were removed to yield a 32-item questionnaire. Through a series of factor analyses, 6-domains were identified, with content consistent with body systems commonly affected by influenza. A mean-based scoring algorithm is used to represent the average symptom severity across symptoms within each domain/body system, with a total score representing symptom severity overall. Given the FLU-PRO's relatively high internal consistency levels, this allows calculation of a domain score in the presence of up to 50% item-level missing data. To assure representation of all body systems in the overall score, total scores are computed across all items comprising the measure only if there is also sufficient data to compute each domain score.

Results suggest FLU-PRO scores are reliable, demonstrate construct and known-groups validity, and are responsive to improvements in health as patients recover from influenza. Floor effects were seen for some items due to low prevalence for those symptoms with the circulating strain of influenza. Internal consistency and 2-day test-retest reliability were strong for the total and domain scores. Consistent with *a priori* hypotheses, FLU-PRO scores were significantly related to patient global ratings of influenza severity, interference with activities, and physical health. The data supported known-groups validity as FLU-PRO scores were lowest in patients rating their symptoms as None/Mild, higher in the Moderate, and highest with Severe/Very Severe. Finally, the FLU-PRO demonstrated responsiveness to change from day 1 to 7, with responders defined by reports of return to usual health and activities.

In FLU-PRO Stage I, concept elicitation interviews with adults and children with influenza were used to develop a comprehensive list of symptoms. Stage II confirmed the ease of understanding item phrasing and ensured content coverage [9]. The 37-item draft instrument could be completed in five minutes or less, showing low respondent burden. The final 32-item FLU-PRO is more comprehensive than existing influenza symptom measures which are limited to the most prevalent symptoms of influenza, such as the 11-item Influenza Symptom Severity Scale (ISS; [20]) and the 10-item FluiiQ™ [21], with the latter

assessing two domains, systemic and respiratory. Quantitative analyses in the present study indicated that the performance of the FLU-PRO was optimized after the removal of only five items, thus the final FLU-PRO remains an inclusive instrument with a broad range of bodily symptoms experienced by patients, enabling greater precision and accuracy for evaluating symptom severity and recovery.

Although FluiiQ [21] was developed in a manner generally consistent with the 2009 FDA PRO Guidance [5] and has been used in clinical trials [22], it has several limitations that are addressed in the FLU-PRO. Specifically, FluiiQ has shown low reliability in the respiratory domain (3 items) [21]. This single domain consists of items that measure cough, sore throat, and nasal congestion symptoms. However, in the domain structure analysis of the FLU-PRO using CFA and EFA, best fit with high reliability was achieved by utilizing three separate domains to assess symptoms in nose, throat, and chest/respiratory systems. Further, each of these three constructs are measured by more than one symptom, in order to more precisely measure changes in symptoms in these body systems. The FLU-PRO also includes an entire 4-item domain to assess a variety of gastrointestinal symptoms, which is absent from the FluiiQ and noted with some strains of influenza. Finally, in terms of the patient populations, the FLU-PRO was developed and evaluated in patients in the US and Mexico—where content coverage and performance was found to be similar in both countries—whereas the FluiiQ was only developed in the US. Further, both hospitalized and non-hospitalized patients were included in the FLU-PRO development samples. To date, no other measure of influenza symptoms has been evaluated in hospitalized patients.

The current study had several limitations worth noting. First, although hospitalized patients were included in the validation patient population, specific details about the event (e.g., duration of influenza prior to hospitalization, acuity level during hospitalization, concurrent complicating conditions) are unknown. Results suggest the FLU-PRO performs consistently in hospitalized and clinic-based samples, however additional study in acute care settings is warranted. Second, missing data increased over time, particularly after the day 2 observation (i.e., approximately 25.2% did not complete the FLU-PRO on days 1, 2, and 3). Three- and 7-day compliance was higher in patients outside the US compared to US sites. Specifically, 52.7% of patients in the US completed the diary on all days from day 1 through day 3 compared to 88.7% of patients outside the US. By day 7, this value dropped to 28.0% and 81.7% for US

and ex-US sites, respectively. This may be due, in part, to the interviewer administered methods used in several ex-US sites. An examination of patterns of missing data indicated patients discontinued the daily survey as their symptoms resolved; 90.0% of patients completed the diary to symptom resolution indicating high rates of compliance during the most relevant days of data collection..

The content validity of the FLU-PRO has been established in children and adolescents through qualitative research. Next steps are to conduct quantitative validation in these patient groups. Future research using the FLU-PRO in influenza challenge studies in healthy adults will provide data on the full course of influenza, from the pre-influenza asymptomatic state to symptom resolution. The FLU-PRO also is also being evaluated for use in influenza-like illness, such as acute respiratory viruses.

CONCLUSION

This paper describes the quantitative methods used to develop and test the FLU-PRO for evaluating patient-reported symptoms in patients with influenza. Results suggest FLU-PRO scores are reliable, valid, and responsive to change in hospitalized and non-hospitalized adults with laboratory-confirmed influenza. The instrument is available for use as a standardized method for evaluating symptoms of influenza in natural history studies and clinical trials.

REFERENCES

1. Centers for Disease Control and Prevention (CDC). Seasonal Influenza (Flu). Available from: http://www.cdc.gov/flu/about/disease/index.htm. [Accessed January 28, 2016].
2. Centers for Disease Control and Prevention (CDC). Key Facts about Influenza (Flu) & Flu Vaccine. Available from: http://www.cdc.gov/flu/keyfacts.htm. [Accessed January 28, 2016].
3. Centers for Disease Control and Prevention (CDC). Seasonal Influenza (Flu) in the Workplace. Available from: http://www.cdc.gov/niosh/topics/flu/. [Accessed January 28, 2016].
4. World Health Organization (WHO). Influenza (Seasonal): Fact sheet N°211. Available from: http://www.who.int/mediacentre/factsheets/fs211/en/. [Accessed January 28, 2016].
5. Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. Fed Regis 2009;74:65132-3.
6. 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.
7. 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.
8. 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-S105.
9. 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 Infectious Diseases 2015;16:1.
10. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York, NY: McGraw-Hill, 1994.
11. Gorsuch RL. Factor Analysis. Hillsdale, NJ: L. Erlbaum Associates, 1983.
12. Yu C-Y. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Doctoral dissertation. University of California, Los Angeles, 2002.

13. Steiger JH. Understanding the limitations of global fit assessment in structural equation modeling. Personality and Individual Differences 2007;42:893-8.
14. Muthén LK, Muthén BO. Mplus user's guide. 3rd ed. Los Angeles, CA: Muthén & Muthén, 1998-2010.
15. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6:1-55.
16. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.
17. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates Inc,, 1988.
18. Stewart AL, Hays RD, Ware JE. Methods of Validating MOS Health Measures. Durham, NC: Duke University Press, 1992.
19. Hays R, Revicki DA. Reliability and validity (including responsiveness). In Fayers P and Hays R (eds.). *Assessing quality of life in clinical trial*s. 2nd ed. New York: Oxford University Press; 2005.
20. Osborne R, Hawthorne G, Papanicolaou M, et al. Measurement of rapid changes in health outcomes in people with influenza symptoms. J Outcomes Res 2000;4:15-30.
21. 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.
22. 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.

TABLES

Table 1. Subject Demographic and Clinical Characteristics by Region: Influenza-Positive Patients (N=221)

Variable	Day 1	
	USA (n=150)	Other Countries[1] (n=71)
Age (years)		
Mean (SD)	39.4 (16.1)	43.5 (17.5)
Median (Range)	36.0 (18–86)	41.0 (18–95)
>65		
Sex, n (%)		
Female	92 (61.3%)	41 (57.7%)
Ethnicity, n (%)[2]		
Hispanic or Latino	16 (10.7%)	67 (94.4%)
Not 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
Black or 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[3]	36 (24.0%)	31 (43.7%)
Missing	22 (14.7%)	4 (5.6%)

Variable	Day 1	
	USA (n=150)	Other Countries[1] (n=71)
Military Status, n (%)		
Never in the military	61 (40.7%)	67 (94.4%)
Active	40 (26.7%)	0
Retired	10 (6.7%)	0
Other[4]	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%)
Co-morbidities[5], n (%)		
None	56 (37.3%)	29 (40.8%)
Asthma	38 (25.3%)	11 (15.5%)
Chronic Obstructive Pulmonary Disease (COPD)	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%)

Variable	Day 1	
	USA (n=150)	Other Countries[1] (n=71)
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%)

[1]Other countries include Mexico (n=67), Argentina (n=3), and UK (n=1)

[2]One participant had missing ethnicity

[3]Other includes homemaker, student, unemployed, and other

[4]Other includes reserves and other

[5]Not mutually exclusive

Table 2. 32-Item FLU-PRO Domain and Total Score Descriptive Statistics (N=221) – Day 1

Scale	Mean (SD)	Range, Median (Mode)	Floor n (%)	Ceiling 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%)

Note: higher FLU-PRO scores = more severe symptoms.

Abbreviations: SD=Standard Deviation

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)[1]	Pairwise Comparisons[2]
	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***

Note: higher FLU-PRO scores = more severe symptoms

[1]p values are: *<0.05, **<0.01, ***<0.001;

[2]Pairwise comparisons between means will be performed using Scheffe's 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.

Abbreviations: SD=Standard Deviation

Table 4. Responsiveness of 32-Item FLU-PRO by Patient Return to Usual Health (N=147)[1] or Return to Usual Activities (N=126)[2], Day 1 to Day 7

Scale	Responders[3]			Non-Responders			p-value
	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (SD)	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (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

Scale	Responders[3]			Non-Responders			p-value
	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (SD)	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (SD)	
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

[1]Responders: N=51; Non-Responders: N=96

[2]Responders: N=87; Non-Responders: N=39

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

Abbreviations: SD=Standard Deviation

Figure 1. Final FLU-PRO Conceptual Framework

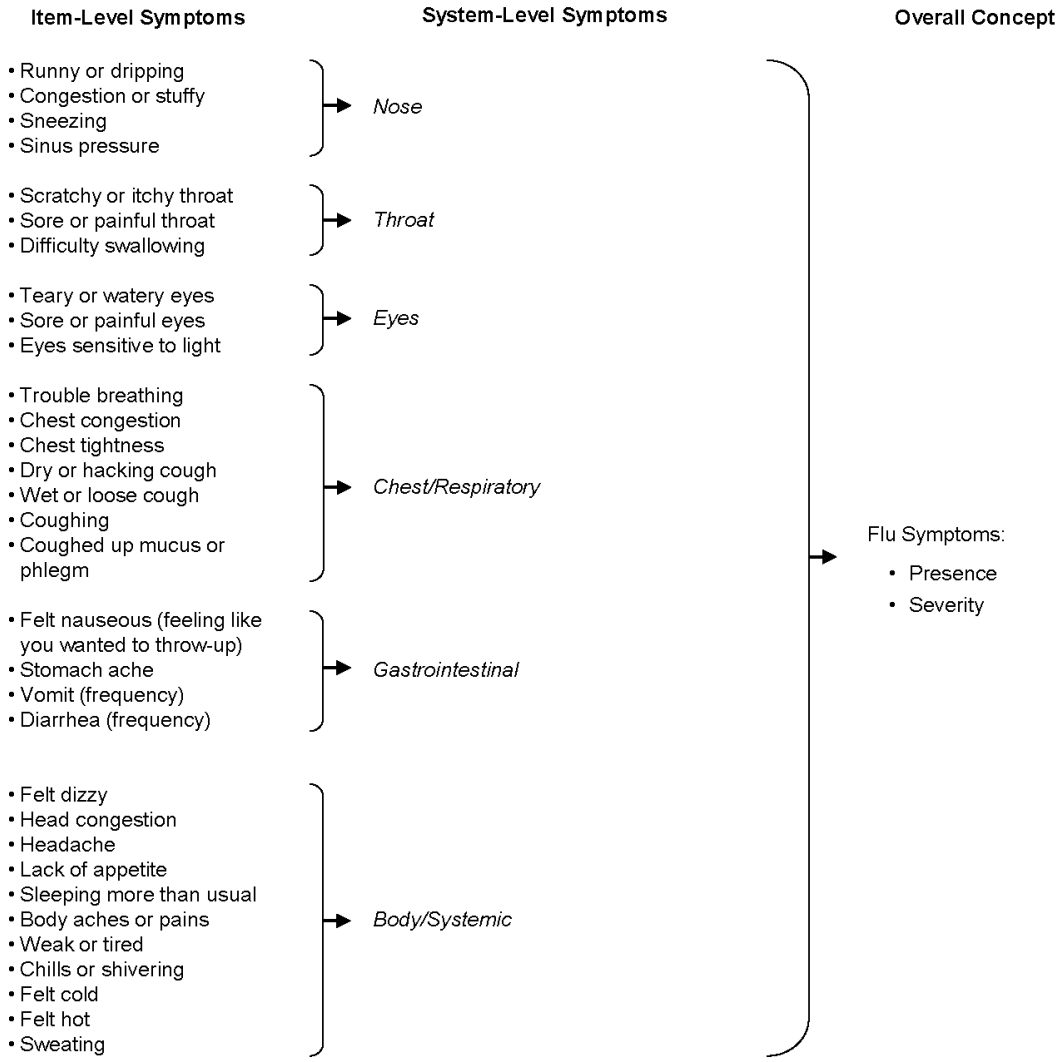
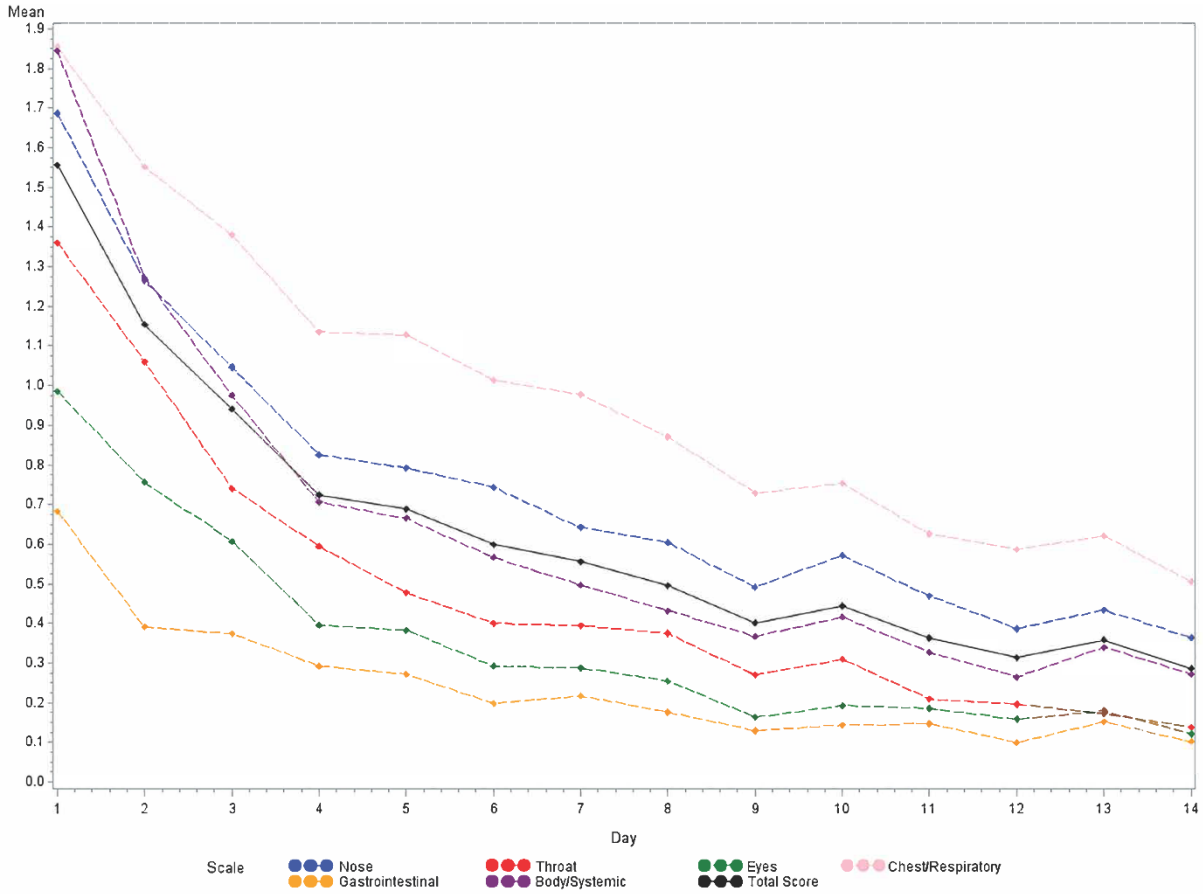


Figure 2. FLU-PRO Domain and Total Score by Diary Days 1 to 14



Evaluation of the Performance Properties of the InFLUenza Patient-Reported Outcome (FLU-PRO©) Instrument

ONLINE SUPPLEMENT

Authors: 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]; Erin A Hansen, MS[8,12]; Michelande Ridoré, MS[13]; Timothy H. Burgess, MD, MPH[7,14]; Eugene V. Millar, PhD[7,8]; Andrés Hernández, MD, MS[15]; Patricia Rodríguez-Zulueta, MD[16]; Mary C. Smolskis, RN, MA[4]; Hilda Ortega-Gallegos, BS[3]; ; Sarah Pett, BSc, MBBS[17]; William Fischer, MD[18]; Daniel Gillor, Dr. Med[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]

Affiliations: [1]Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD in support of the Collaborative Clinical Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, Bethesda, 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 Institutes of Allergy and Infectious Diseases, NIH, Bethesda, MD, USA; [6]Madigan Army Medical Center, Fort Lewis, WA, USA; [7]Infectious Disease Clinical Research Program, Department of Preventive Medicine and Biostatistics, Uniformed Services University, Bethesda MD, USA; [8]Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, 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]Naval Health Research Center, San Diego, CA, USA; [13]Children's National Medical Center, Washington, DC, USA; [14]Walter Reed National Military Medical Center, Bethesda, MD, USA; [15]Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas , Mexico City, Mexico; [16]Hospital General Dr. Manuel Gea González, Mexico City, Mexico; [17]University College London, London, UK; [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

***Corresponding Author**
John H. Powers, MD
5601 Fishers Lane
Room 4D50
Rockville, MD 20852
Phone: 301-768-6711
Fax: 240-465-0396
E-mail: john.powers@nih.gov

Funding: This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. This research was supported in part by the Intramural Research Program of the NIH, NIAID. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Keywords: influenza; patient-reported outcome; psychometric; reliability; validity; responsiveness

Running title: FLU-PRO Validation

Acknowledgements: We would like to extend sincere appreciation to staff providing 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; and to Kathryn Miller for her assistance with manuscript text edit and formatting. We also express our gratitude to the patients who gave of their time to help complete this study.

PHASE I SUPPLEMENTARY TABLES: ALL INFLUENZA POSITIVE PATIENTS

Table S1. Item Analysis: FLU-PRO Item Descriptive Statistics in Influenza-Positive Patients (N=221) – Day 1[1]

	Mean ± SD	Median	Mode	Floor n (%)	Ceiling n (%)
1. Runny or dripping nose	1.8 ± 1.4	2.0	1	44 (19.9%)	35 (15.8%)
2. Congested or stuffy nose	2.0 ± 1.4	2.0	1	37 (16.7%)	38 (17.2%)
3. Scratchy or itchy throat	1.7 ± 1.4	1.0	0	60 (27.1%)	35 (15.8%)
4. Sore or painful throat	1.6 ± 1.4	1.0	0	65 (29.4%)	34 (15.4%)
5. Swollen throat	0.9 ± 1.2	0.0	0	115 (52.0%)	13 (5.9%)
6. Difficulty swallowing	0.8 ± 1.1	0.0	0	124 (56.1%)	11 (5.0%)
7. Teary or watery eyes	1.1 ± 1.3	1.0	0	98 (44.3%)	19 (8.6%)
8. Sore or painful eyes	0.9 ± 1.3	0.0	0	128 (57.9%)	18 (8.1%)
9. Eyes sensitive to light	1.0 ± 1.3	0.0	0	126 (57.0%)	19 (8.6%)
10. Trouble breathing	1.4 ± 1.4	1.0	0	80 (36.2%)	24 (10.9%)
11. Chest congestion	1.8 ± 1.4	2.0	0	51 (23.1%)	35 (15.8%)
12. Chest tightness	1.4 ± 1.4	1.0	0	88 (39.8%)	26 (11.8%)
13. Dry or hacking cough	2.3 ± 1.3	3.0	3	31 (14.0%)	48 (21.7%)
14. Wet or loose cough	1.5 ± 1.4	1.0	0	69 (31.2%)	23 (10.4%)
15. Headache	2.1 ± 1.5	2.0	4	48 (21.7%)	60 (27.1%)
16. Head congestion	1.5 ± 1.4	1.0	0	82 (37.1%)	26 (11.8%)
17. Sinus pressure	1.4 ± 1.4	1.0	0	89 (40.3%)	23 (10.4%)
18. Felt dizzy	1.1 ± 1.3	1.0	0	102 (46.2%)	15 (6.8%)
19. Felt lightheaded	0.9 ± 1.2	0.0	0	117 (52.9%)	13 (5.9%)
20. Lack of appetite	1.8 ± 1.5	2.0	0	58 (26.2%)	45 (20.4%)
21. Felt nauseous	1.0 ± 1.3	0.0	0	116 (52.5%)	23 (10.4%)
22. Stomach ache	0.8 ± 1.2	0.0	0	130 (58.8%)	12 (5.4%)
23. Sleeping more than usual	1.6 ± 1.5	1.0	0	82 (37.1%)	32 (14.5%)
24. Difficulty staying asleep	1.7 ± 1.4	2.0	0	69 (31.2%)	34 (15.4%)
25. Difficulty falling asleep	1.4 ± 1.4	1.0	0	88 (39.8%)	27 (12.2%)
26. Body aches or pains	2.5 ± 1.5	3.0	4	31 (14.0%)	79 (35.7%)
27. Weak or tired	2.6 ± 1.3	3.0	4	18 (8.1%)	79 (35.7%)
28. Chills or shivering	1.9 ± 1.5	2.0	0	63 (28.5%)	46 (20.8%)
29. Felt cold	2.1 ± 1.5	2.0	3	40 (18.1%)	52 (23.5%)
30. Felt hot	1.6 ± 1.4	1.0	0	64 (29.0%)	30 (13.6%)

	Mean ± SD	Median	Mode	Floor n (%)	Ceiling n (%)
31. Sweating	1.5 ± 1.4	1.0	0	71 (32.1%)	27 (12.2%)
32. Felt uncomfortable	2.4 ± 1.4	3.0	4	33 (14.9%)	68 (30.8%)
33. How many times did you vomit	0.4 ± 0.9	0.0	0	185 (83.7%)	6 (2.7%)
34. How many times did you have diarrhea?	0.5 ± 1.1	0.0	0	169 (76.5%)	13 (5.9%)
35. Sneezing	1.6 ± 1.2	1.0	1	48 (21.7%)	11 (5.0%)
36. Coughing	2.7 ± 1.0	3.0	3	5 (2.3%)	48 (21.7%)
37. Coughed up mucus or phlegm	1.9 ± 1.3	2.0	2	42 (19.0%)	29 (13.1%)

[1]Range for all items: 0.0–4.0

Abbreviations: SD=Standard Deviation

Table S2. Item Analysis: Confirmatory Factor Analysis for FLU-PRO Upper Respiratory, Lower Respiratory, and Systemic Domains

Item Number	Standardized Coefficient (SE)[1]
Upper Respiratory	
Nose	
1. Runny or dripping nose	0.634 (0.042)
2. Congested or stuffy nose	0.681 (0.041)
35. Sneezing	0.453 (0.052)
Throat	
3. Scratchy or itchy throat	0.760 (0.034)
4. Sore or painful throat	0.785 (0.032)
5. Swollen throat	0.720 (0.043)
6. Difficulty swallowing	0.768 (0.040)
Eyes	
7. Teary or watery eyes	0.678 (0.043)
8. Sore or painful eyes	0.812 (0.034)
9. Eyes sensitive to light	0.803 (0.038)
Lower Respiratory	
Chest	
10. Trouble breathing	0.743 (0.045)
11. Chest congestion	0.860 (0.030)
12. Chest tightness	0.859 (0.032)
13. Dry or hacking cough	0.655 (0.046)
14. Wet or loose cough	0.479 (0.060)
36. Coughing	0.594 (0.053)
37. Coughed up mucus or phlegm	0.424 (0.065)
Systemic	
Head	
15. Headache	0.744 (0.034)
16. Head congestion	0.713 (0.040)
17. Sinus pressure	0.662 (0.043)
18. Felt dizzy	0.814 (0.027)
19. Felt lightheaded	0.776 (0.031)
Gastrointestinal	
20. Lack of appetite	0.599 (0.047)
21. Felt nauseous	0.671 (0.044)
22. Stomach ache	0.647 (0.047)

Item Number	Standardized Coefficient (SE)[1]
33. How many times did you vomit?	0.676 (0.056)
34. How many times did you have diarrhea?	0.321 (0.083)
Sleep	
23. Sleeping more than usual	0.265 (0.065)
24. Difficulty staying asleep	0.594 (0.043)
25. Difficulty falling asleep	0.526 (0.048)
Body/Systemic	
26. Body aches or pains	0.820 (0.027)
27. Weak or tired	0.754 (0.034)
28. Chills or shivering	0.800 (0.029)
29. Felt cold	0.721 (0.034)
30. Felt hot	0.579 (0.044)
31. Sweating	0.537 (0.047)
32. Felt uncomfortable	0.802 (0.029)

Comparative Fit Index (CFI) = 0.836, Root Mean Square Error of Approximation (RMSEA) = 0.089, and Weighted Root Mean Residual (WRMR) = 1.722.

[1]All parameter estimates are $p<0.001$.

Abbreviations: SE=Standard error

Table S3. Item Analysis: Exploratory Factor Analysis Standardized Factor Loadings for FLU-PRO Items - Seven Factor Solution

Item Number	F1	F2	F3	F4	F5	F6	F7
1. Runny or dripping nose	0.821	-0.006	-0.024	-0.044	0.087	0.018	0.066
2. Congested or stuffy nose	0.744	0.081	0.105	0.074	0.065	0.016	-0.048
35. Sneezing	0.716	0.007	0.127	0.000	-0.083	-0.053	0.049
3. Scratchy or itchy throat	0.073	0.764	0.027	-0.021	0.047	-0.104	0.084
4. Sore or painful throat	0.056	0.946	-0.117	-0.091	0.036	0.004	0.039
5. Swollen throat	-0.003	0.787	0.019	0.171	-0.064	0.155	-0.037
6. Difficulty swallowing	-0.106	0.731	0.117	0.271	0.094	0.038	-0.083
7. Teary or watery eyes	0.424	0.196	0.050	0.451	0.030	-0.044	0.045
8. Sore or painful eyes	0.315	0.292	0.012	0.758	-0.040	0.015	0.082
9. Eyes sensitive to light	0.134	0.139	0.018	0.636	0.231	0.022	0.059
10. Trouble breathing	0.011	0.042	0.564	0.248	0.034	0.087	0.058
11. Chest congestion	-0.004	-0.011	0.801	0.024	0.243	0.016	-0.092
12. Chest tightness	-0.044	-0.041	0.745	0.247	0.180	-0.046	0.010
13. Dry or hacking cough	0.063	0.046	0.464	-0.008	0.144	-0.138	0.179
14. Wet or loose cough	-0.014	0.036	0.554	-0.309	-0.104	0.391	0.050
36. Coughing	0.156	0.027	0.595	-0.237	0.070	-0.049	0.080
37. Coughed up mucus or phlegm	0.072	0.028	0.538	-0.427	-0.092	0.371	0.007
15. Headache	0.147	0.011	-0.046	0.247	0.640	0.043	-0.026
16. Head congestion	0.194	0.010	0.137	0.087	0.484	0.209	-0.050
17. Sinus pressure	0.483	-0.029	0.136	0.005	0.425	0.122	-0.113
18. Felt dizzy	-0.019	-0.003	-0.049	0.220	0.464	0.630	-0.051
19. Felt lightheaded	-0.178	-0.058	0.015	0.162	0.405	0.610	0.119
20. Lack of appetite	0.024	-0.051	0.064	0.118	0.346	0.180	0.222
21. Felt nauseous	0.103	0.019	-0.151	-0.072	0.036	0.616	0.585
22. Stomach ache	0.046	0.144	0.015	0.004	0.035	0.502	0.394
33. How many times did you vomit?	0.020	0.046	-0.007	-0.171	0.096	0.417	0.643
34. How many times did you have diarrhea?	0.061	-0.187	0.125	-0.001	0.097	0.155	0.280
23. Sleeping more than usual	0.063	-0.107	0.066	-0.135	0.467	0.122	-0.250
24. Difficulty staying asleep	-0.170	-0.103	0.069	0.077	0.072	0.052	0.770
25. Difficulty falling asleep	-0.299	0.001	0.118	0.124	-0.123	-0.026	0.902
26. Body aches or pains	0.065	0.138	-0.114	0.052	0.698	-0.025	0.180
27. Weak or tired	-0.030	0.052	0.021	0.101	0.667	0.105	0.029
28. Chills or shivering	0.062	0.017	-0.030	-0.032	0.912	-0.293	0.054

Item Number	F1	F2	F3	F4	F5	F6	F7
29. Felt cold	0.055	-0.029	-0.034	-0.016	0.815	-0.302	0.116
30. Felt hot	-0.306	0.135	0.090	-0.191	0.615	0.003	0.149
31. Sweating	-0.207	0.104	0.075	-0.203	0.706	-0.023	-0.051
32. Felt uncomfortable	-0.079	0.137	0.019	0.075	0.711	0.015	0.077

Comparative Fit Index (CFI) = .92, Root Mean Square Error of Approximation (RMSEA) = .058, and Standardized Root Mean Square Residual (SRMR) = .046, Test of Close Fit p value = .033

Color coding represents items within separate body systems.

PHASE II SUPPLEMENTARY TABLES: ALL INFLUENZA POSITIVE PATIENTS

Table S4. Two-day Reproducibility of 32-Item FLU-PRO Days 1–7 in Influenza Positive Patients

FLU-PRO Scores	N[1]	Mean Difference (SD)[2]	T Statistic	p-value	Effect Size	ICC
Day 1 to Day 2						
Nose	44	0.3 (0.7)	2.78	0.0079	0.25	0.79
Throat	44	0.3 (1.0)	2.14	0.0382	0.28	0.57
Eyes	44	0.2 (0.9)	1.27	0.2092	0.17	0.62
Chest/Respiratory	44	0.1 (0.6)	0.93	0.3591	0.10	0.76
Gastrointestinal	44	0.2 (0.9)	1.27	0.2092	0.17	0.62
Body/Systemic	44	0.2 (0.7)	1.53	0.1327	0.17	0.65
Total Score	44	0.3 (0.6)	3.40	0.0015	0.29	0.83
Day 2 to Day 3						
Nose	27	0.0 (0.5)	0.11	0.9170	0.01	0.87
Throat	27	0.1 (0.6)	0.50	0.6209	0.05	0.84
Eyes	27	-0.0 (0.4)	-0.36	0.7223	0.04	0.83
Chest/Respiratory	27	0.0 (0.4)	0.06	0.9503	0.01	0.85
Gastrointestinal	27	-0.0 (0.4)	-0.36	0.7223	0.04	0.83
Body/Systemic	27	-0.2 (0.7)	-1.48	0.1517	0.47	0.24
Total Score	27	0.1 (0.5)	0.93	0.3600	0.09	0.86
Day 3 to Day 4						
Nose	29	-0.1 (0.7)	-0.87	0.3891	0.12	0.74
Throat	29	0.0 (0.5)	0.12	0.9023	0.01	0.91
Eyes	29	0.1 (0.5)	0.70	0.4892	0.06	0.88
Chest/Respiratory	29	0.1 (0.5)	0.61	0.5486	0.06	0.86
Gastrointestinal	29	0.1 (0.5)	0.70	0.4892	0.06	0.88
Body/Systemic	29	-0.1 (0.6)	-0.75	0.4582	0.12	0.64
Total Score	29	0.0 (0.4)	0.30	0.7637	0.02	0.92
Day 4 to Day 5						
Nose	18	-0.0 (0.5)	-0.37	0.7168	0.04	0.88
Throat	18	0.2 (0.6)	1.11	0.2840	0.18	0.70
Eyes	18	0.0 (0.5)	0.15	0.8811	0.02	0.85
Chest/Respiratory	18	-0.2 (0.5)	-1.65	0.1175	0.24	0.77
Gastrointestinal	18	0.0 (0.5)	0.15	0.8811	0.02	0.85
Body/Systemic	18	-0.1 (0.3)	-0.68	0.5084	0.11	0.76
Total Score	18	-0.1 (0.4)	-1.22	0.2380	0.18	0.79
Day 5 to Day 6						
Nose	23	-0.0 (0.5)	-0.30	0.7633	0.04	0.80
Throat	23	0.3 (0.8)	1.52	0.1424	0.24	0.60

FLU-PRO Scores	N[1]	Mean Difference (SD)[2]	T Statistic	p-value	Effect Size	ICC
Eyes	23	0.1 (0.8)	0.53	0.6041	0.10	0.52
Chest/Respiratory	23	0.1 (0.4)	1.04	0.3099	0.12	0.84
Gastrointestinal	23	0.1 (0.8)	0.53	0.6041	0.10	0.52
Body/Systemic	23	-0.0 (0.3)	-0.64	0.5285	0.10	0.76
Total Score	23	-0.0 (0.3)	-0.18	0.8606	0.02	0.90
Day 6 to Day 7						
Nose	23	-0.1 (0.2)	-1.30	0.2077	0.06	0.97
Throat	23	0.0 (0.2)	0.81	0.4264	0.05	0.96
Eyes	23	0.1 (0.4)	1.50	0.1479	0.16	0.82
Chest/Respiratory	23	0.0 (0.4)	0.17	0.8692	0.01	0.92
Gastrointestinal	23	0.1 (0.4)	1.50	0.1479	0.16	0.82
Body/Systemic	23	0.1 (0.3)	1.19	0.2451	0.25	0.31
Total Score	23	0.0 (0.2)	0.60	0.5557	0.05	0.93

[1]Number of study participants.

[2]Mean difference = average day X FLU-PRO score - average day Y FLU-PRO score (e.g., day 1 score - day 2 score)

Table S5. Construct Validity: 32-Item FLU-PRO Scale Correlations with Other PRO Measures at Day 1

Day 1	Domains and Total Score[1]						
	Nose	Throat	Eyes	Chest/ Respiratory	Gastrointestinal	Body/ Systemic	Total Score
Patient Global Rating of Flu Severity[2]	0.27***	0.28***	0.34***	0.39***	0.34***	0.61***	0.59***
Patient Global Rating of Physical Health[3]	-0.09	-.021*	-0.26***	-0.24**	-0.06	-0.28***	-0.29***
Patient Global Assessment of Interference in Daily Activities[4]	0.11	0.14*	0.28***	0.29***	0.22*	0.50***	0.43***

[1]Spearman correlation coefficients: ***p<0.0001, **p<0.001, *p<0.05

[2]Greater values indicate greater disease severity

[3]Greater values indicate better patient health

[4]Greater values indicate greater interference with daily activities

PHASE II SUPPLEMENTARY TABLES: STRATIFIED ANALYSES BY HOSPITALIZATION STATUS

Table S6. Patient Demographic and Clinical Characteristics by Hospitalization Status: Influenza-Positive Patients (N=221)

Variable	Day 1	
	Hospitalization[1] (n=53)	No Hospitalization (n=168)
Age, Years		
Mean (SD)	49.7 (18.6)	37.9 (15.0)
Median (Range)	50.0 (18–95)	33.0 (18–81)
>65	9 (17.0%)	10 (6.0%)
Sex, n (%)		
Female	35 (66.0%)	98 (58.3%)
Ethnicity, n (%)[2]		
Hispanic or Latino	29 (54.7%)	54 (32.1%)
Not Hispanic or Latino	23 (43.4%)	114 (67.9%)
Race, n (%)		
American Indian or Alaska Native	0	4 (2.4%)
Asian	0	5 (3.0%)
Black or African American	17 (32.1%)	57 (33.9%)
Mestizo	29 (54.7%)	38 (22.6%)
White	7 (13.2%)	58 (34.5%)
Other	0	6 (3.6%)
Employment Status, n (%)		
Employed, full time or part-time	21 (40.0%)	93 (55.3%)
Retired	5 (9.4%)	9 (5.4%)
Other[3]	27 (50.9%)	40 (23.8%)
Missing	0	26 (15.5%)
Military Status, n (%)		
Never in the military	53 (100.0%)	75 (44.6%)
Active	0	40 (23.8%)
Retired	0	10 (6.0%)
Other	0	16 (9.5%)
Missing	0	27 (16.1%)
Highest Level of Education		
Elementary/primary school	5 (9.4%)	7 (4.2%)
Secondary/high school or less	24 (45.3%)	48 (28.6%)

Variable	Day 1	
	Hospitalization[1] (n=53)	No Hospitalization (n=168)
Some college	8 (15.1%)	29 (17.3%)
College degree or more	14 (26.4%)	59 (35.1%)
Other	7 (13.2%)	32 (19.0%)
Current Treatments, n (%)		
None	13 (24.5%)	63 (37.5%)
Oseltamivir (Tamiflu)	18 (34.0%)	45 (26.8%)
Amantadine (Symmetrel)	0	2 (1.2%)
Other	31 (58.5%)	67 (39.9%)
Acetaminophen	10 (18.9%)	34 (20.2%)
Antibiotic	5 (9.4%)	5 (3.0%)
Antihistamine	0	3 (1.8%)
Aspirin	3 (5.7%)	1 (0.6%)
Codeine	1 (1.9%)	0
Cough suppressant or expectorant	3 (5.7%)	11 (6.5%)
Decongestant	0	7 (4.2%)
IV fluids	0	1 (0.6%)
Inhaled corticosteroid	6 (11.3%)	1 (0.6%)
Mucolytic	4 (7.5%)	0 (0%)
NSAID	3 (5.7%)	13 (7.7%)
OTC: symptom relief	0	3 (1.8%)
Opioid	1 (1.9%)	0
Short-acting beta agonist	2 (3.8%)	0
Co-morbidities[5], n (%)		
None	11 (20.8%)	74 (44.0%)
Asthma	17 (32.1%)	32 (19.0%)
Chronic Obstructive Pulmonary Disease (COPD)	3 (5.7%)	6 (3.6%)
Osteoporosis	0	2 (1.2%)
Depression	5 (9.4%)	16 (9.5%)
Hypertension	15 (28.3%)	18 (10.7%)
Raised cholesterol	5 (9.4%)	17 (10.1%)
Stomach ulcers	1 (1.9%)	5 (3.0%)
Heart attack/angina	2 (3.8%)	1 (0.6%)
Diabetes	14 (26.4%)	16 (9.5%)

Variable	Day 1	
	Hospitalization[1] (n=53)	No Hospitalization (n=168)
Kidney disease	5 (9.4%)	3 (1.8%)
Lung disease	3 (5.7%)	2 (1.2%)
Tuberculosis	0	2 (1.2%)
Other	16 (30.2%)	40 (23.8%)

[1]53 patients were hospitalized in the influenza-positive group and 61 patients were hospitalized in the influenza-negative group

[2]One person had missing ethnicity

[3]Other includes homemaker, student, unemployed, and other

[4]Other includes reserves and other

[5]Not mutually exclusive

Table S7a. 32-Item FLU-PRO Domain and Total Score Descriptive Statistics in Non-Hospitalized Influenza-Positive Patients (N=168) - Day 1

Scale	Mean (SD)	Range, Median (Mode)	Floor n (%)	Ceiling n (%)
Nose	1.8 ± 1.0	0.0–4.0, 1.8 (1.3)	6 (3.6%)	3 (1.8%)
Throat	1.5 ± 1.1	0.0–4.0, 1.3 (0.0)	23 (13.7%)	4 (2.4%)
Eyes	1.0 ± 1.0	0.0–4.0, 0.7 (0.0)	45 (26.8%)	4 (2.4%)
Chest/Respiratory	1.8 ± 0.9	0.0–3.7, 1.8 (1.7)	3 (1.8%)	0 (0.0%)
Gastrointestinal	0.6 ± 0.8	0.0–3.5, 0.3 (0.0)	61 (36.3%)	0 (0.0%)
Body/Systemic	1.9 ± 0.9	0.0–3.8, 2.0 (2.5)	2 (1.2%)	0 (0.0%)
Total Score	1.6 ± 0.7	0.3–3.4, 1.7 (1.3)	0 (0.0%)	0 (0.0%)

Table S7b. 32-Item FLU-PRO Domain and Total Score Descriptive Statistics in Hospitalized Influenza-Positive Patients (N=53) - Day 1

Scale	Mean (SD)	Range, Median (Mode)	Floor n (%)	Ceiling n (%)
Nose	1.2 ± 1.0	0.0–4.0, 1.0 (1.0)	7 (13.2%)	1 (1.9%)
Throat	1.1 ± 1.1	0.0–4.0, 0.7 (0.0)	18 (34.0%)	1 (1.9%)
Eyes	1.0 ± 1.3	0.0–4.0, 0.3 (0.0)	22 (41.5%)	4 (7.5%)
Chest/Respiratory	1.9 ± 0.9	0.4–4.0, 1.9 (1.9)	0 (0.0%)	2 (3.8%)
Gastrointestinal	0.8 ± 1.0	0.0–3.8, 0.5 (0.0)	16 (30.2%)	0 (0.0%)
Body/Systemic	1.6 ± 0.9	0.1–3.6, 1.5 (0.3)	0 (0.0%)	0 (0.0%)
Total Score	1.4 ± 0.8	0.3–3.7, 1.3 (0.5)	0 (0.0%)	0 (0.0%)

Table S8a. Two-day Reproducibility of 32-Item FLU-PRO Days 1 to Day 7 in Non-Hospitalized Influenza-Positive Patients

FLU-PRO Scores	N[1]	Day X Mean (SD)	Day Y Mean (SD)	Mean Difference (SD)[2]	T Statistic	p value	Effect Size	ICC
Day 1 to Day 2								
Nose	36	1.8 (1.1)	1.6 (1.1)	0.2 (0.6)	2.29	0.0281	0.20	0.84
Throat	36	1.9 (1.2)	1.6 (1.1)	0.3 (1.0)	1.84	0.0740	0.25	0.61
Eyes	36	1.0 (1.0)	0.9 (1.1)	0.1 (0.9)	0.71	0.4845	0.11	0.65
Chest/Respiratory	36	1.7 (0.8)	1.7 (0.9)	-0.0 (0.6)	-0.17	0.8664	0.02	0.79
Gastrointestinal	36	0.7 (0.9)	0.6 (0.6)	0.2 (0.7)	1.50	0.1424	0.21	0.52
Body/Systemic	36	1.9 (1.0)	1.7 (1.1)	0.2 (0.6)	2.49	0.0177	0.23	0.84
Total Score	36	1.6 (0.7)	1.5 (0.7)	0.2 (0.4)	2.60	0.0135	0.22	0.85
Day 2 to Day 3								
Nose	21	1.2 (0.9)	1.2 (0.9)	0.0 (0.5)	0.11	0.9147	0.01	0.86
Throat	21	1.2 (1.2)	1.2 (1.2)	0.1 (0.7)	0.50	0.6236	0.07	0.82
Eyes	21	0.6 (0.7)	0.6 (0.7)	-0.0 (0.3)	-0.21	0.8333	0.02	0.87
Chest/Respiratory	21	1.7 (0.8)	1.6 (0.8)	0.1 (0.5)	0.55	0.5915	0.06	0.84
Gastrointestinal	21	0.3 (0.5)	0.5 (0.6)	-0.1 (0.6)	-0.76	0.4554	0.24	0.30
Body/Systemic	21	1.3 (1.0)	1.2 (0.8)	0.1 (0.5)	1.11	0.2813	0.12	0.85
Total Score	21	1.2 (0.6)	1.1 (0.6)	0.0 (0.3)	0.75	0.4617	0.07	0.89
Day 3 to Day 4								
Nose	24	1.4 (0.9)	1.5 (1.0)	-0.1 (0.7)	-0.76	0.4575	0.11	0.77
Throat	24	1.3 (1.1)	1.3 (1.2)	0.0 (0.5)	0.13	0.9013	0.01	0.90
Eyes	24	1.0 (1.1)	0.8 (0.9)	0.1 (0.5)	1.31	0.2031	0.13	0.87
Chest/Respiratory	24	1.9 (1.0)	1.9 (1.0)	0.0 (0.5)	0.11	0.9151	0.01	0.86
Gastrointestinal	24	0.5 (0.6)	0.5 (0.6)	-0.1 (0.6)	-0.45	0.6591	0.09	0.57
Body/Systemic	24	1.5 (1.0)	1.4 (0.9)	0.1 (0.4)	0.88	0.3888	0.07	0.93
Total Score	24	1.4 (0.7)	1.4 (0.8)	0.0 (0.3)	0.34	0.7396	0.03	0.93
Day 4 to Day 5								
Nose	18	1.5 (1.0)	1.6 (0.9)	-0.0 (0.5)	-0.37	0.7168	0.04	0.88
Throat	18	1.1 (0.9)	0.9 (0.7)	0.2 (0.6)	1.11	0.2840	0.18	0.70
Eyes	18	0.8 (0.9)	0.8 (0.9)	0.0 (0.5)	0.15	0.8811	0.02	0.85
Chest/Respiratory	18	1.6 (0.8)	1.8 (0.7)	-0.2 (0.5)	-1.65	0.1175	0.24	0.77
Gastrointestinal	18	0.4 (0.5)	0.5 (0.5)	-0.1 (0.3)	-0.68	0.5084	0.11	0.76
Body/Systemic	18	1.1 (0.6)	1.2 (0.6)	-0.1 (0.4)	-1.22	0.2380	0.18	0.79

FLU-PRO Scores	N[1]	Day X Mean (SD)	Day Y Mean (SD)	Mean Difference (SD)[2]	T Statistic	p value	Effect Size	ICC
Total Score	18	1.2 (0.6)	1.2 (0.5)	-0.1 (0.3)	-1.02	0.3207	0.13	0.84
Day 5 to Day 6								
Nose	18	0.9 (0.8)	0.9 (0.8)	0.0 (0.4)	0.00	1.0000	0.00	0.87
Throat	18	0.6 (1.0)	0.5 (0.8)	0.1 (0.4)	1.30	0.2097	0.11	0.92
Eyes	18	0.3 (0.5)	0.3 (0.5)	0.0 (0.4)	0.22	0.8260	0.04	0.75
Chest/Respiratory	18	1.1 (0.7)	1.2 (0.7)	-0.0 (0.2)	-0.32	0.7557	0.02	0.95
Gastrointestinal	18	0.3 (0.5)	0.3 (0.5)	-0.0 (0.4)	-0.50	0.6260	0.08	0.76
Body/Systemic	18	0.6 (0.6)	0.6 (0.6)	0.0 (0.2)	0.45	0.6594	0.03	0.95
Total Score	18	0.7 (0.5)	0.7 (0.5)	0.0 (0.1)	0.37	0.7124	0.02	0.97
Day 6 to Day 7								
Nose	19	0.9 (1.1)	0.9 (1.2)	-0.1 (0.3)	-1.10	0.2871	0.06	0.97
Throat	19	0.4 (0.6)	0.3 (0.6)	0.0 (0.2)	1.00	0.3306	0.05	0.97
Eyes	19	0.4 (0.8)	0.3 (0.5)	0.1 (0.4)	1.57	0.1341	0.18	0.82
Chest/Respiratory	19	1.1 (0.9)	1.1 (0.9)	0.0 (0.4)	0.09	0.9332	0.01	0.92
Gastrointestinal	19	0.2 (0.3)	0.1 (0.2)	0.1 (0.2)	0.94	0.3597	0.20	0.48
Body/Systemic	19	0.5 (0.6)	0.5 (0.7)	0.1 (0.2)	1.36	0.1895	0.10	0.95
Total Score	19	0.6 (0.7)	0.6 (0.6)	0.0 (0.2)	1.07	0.2985	0.06	0.97

[1]Number of study participants with no change in flu symptom at day Y.

[2]Mean difference = average Day X FLU-PRO score - average Day Y FLU-PRO score (ex. Day 1 score - Day 2 score); p value from paired t-test.

Table S8b. Two-day Reproducibility of 32-Item FLU-PRO Days 1 to Day 7 in Hospitalized Influenza-Positive Patients

FLU-PRO Scores	N[1]	Day X Mean (SD)	Day Y Mean (SD)	Mean Difference (SD)[2]	T Statistic	p value	Effect Size	ICC
Day 1 to Day 2								
Nose	8	1.6 (1.4)	1.1 (1.2)	0.6 (1.0)	1.58	0.1580	0.40	0.66
Throat	8	1.1 (1.1)	0.7 (0.6)	0.5 (1.3)	1.03	0.3375	0.44	-0.04
Eyes	8	0.8 (1.1)	0.4 (0.5)	0.5 (0.9)	1.49	0.1806	0.41	0.44
Chest/Respiratory	8	1.9 (0.8)	1.4 (0.7)	0.5 (0.5)	3.07	0.0180	0.63	0.68
Gastrointestinal	8	0.5 (1.2)	0.5 (1.0)	0.0 (0.3)	0.31	0.7627	0.03	0.97
Body/Systemic	8	1.5 (1.0)	0.9 (1.1)	0.6 (0.6)	2.75	0.0286	0.56	0.75
Total Score	8	1.4 (0.9)	0.9 (0.8)	0.5 (0.5)	2.67	0.0322	0.51	0.74
Day 2 to Day 3								
Nose	6	0.6 (0.6)	0.6 (0.5)	0.0 (0.3)	0.00	1.0000	0.00	0.89
Throat	6	0.3 (0.4)	0.3 (0.4)	0.0 (0.0)			0.00	1.00
Eyes	6	0.3 (0.4)	0.3 (0.3)	-0.1 (0.4)	-0.31	0.7711	0.13	0.35
Chest/Respiratory	6	0.9 (0.4)	1.1 (0.3)	-0.2 (0.3)	-1.23	0.2722	0.47	0.49
Gastrointestinal	6	0.1 (0.2)	0.6 (0.9)	-0.5 (0.8)	-1.55	0.1820	2.39	0.20
Body/Systemic	6	0.4 (0.2)	0.4 (0.4)	-0.0 (0.3)	-0.36	0.7327	0.21	0.54
Total Score	6	0.5 (0.1)	0.6 (0.2)	-0.1 (0.2)	-1.67	0.1566	0.99	0.50
Day 3 to Day 4								
Nose	5	0.8 (0.6)	0.9 (0.6)	-0.2 (0.8)	-0.40	0.7102	0.26	-0.06
Throat	5	0.5 (0.9)	0.5 (1.0)	-0.0 (0.2)	-0.00	1.0000	0.00	0.98
Eyes	5	0.7 (1.3)	1.0 (1.4)	-0.3 (0.5)	-1.21	0.2943	0.21	0.93
Chest/Respiratory	5	1.3 (0.7)	1.1 (0.8)	0.3 (0.4)	1.58	0.1890	0.39	0.83
Gastrointestinal	5	0.8 (0.9)	1.0 (0.6)	-0.2 (0.5)	-0.87	0.4320	0.22	0.80
Body/Systemic	5	0.9 (1.0)	1.1 (0.9)	-0.2 (0.5)	-0.80	0.4669	0.19	0.87
Total Score	5	0.9 (0.7)	1.0 (0.7)	-0.1 (0.2)	-0.68	0.5314	0.10	0.95
Day 5 to Day 6								
Nose	5	1.1 (0.8)	1.3 (0.9)	-0.2 (0.8)	-0.40	0.7102	0.18	0.59
Throat	5	0.9 (1.6)	0.1 (0.2)	0.8 (1.6)	1.09	0.3375	0.52	-0.10
Eyes	5	1.1 (1.7)	0.7 (1.2)	0.3 (1.7)	0.44	0.6808	0.20	0.37
Chest/Respiratory	5	1.4 (0.8)	1.0 (0.5)	0.4 (0.6)	1.56	0.1939	0.54	0.52

FLU-PRO Scores	N[1]	Day X Mean (SD)	Day Y Mean (SD)	Mean Difference (SD)[2]	T Statistic	p value	Effect Size	ICC
Gastrointestinal	5	0.4 (0.3)	0.5 (0.4)	-0.1 (0.2)	-0.53	0.6213	0.18	0.82
Body/Systemic	5	1.3 (0.8)	1.5 (0.8)	-0.1 (0.6)	-0.46	0.6710	0.16	0.74
Total Score	5	1.1 (0.8)	1.0 (0.5)	0.1 (0.7)	0.40	0.7098	0.17	0.44
Day 6 to Day 7								
Nose	4	0.3 (0.4)	0.3 (0.5)	-0.1 (0.1)	-1.00	0.3910	0.18	0.96
Throat	4	0.3 (0.3)	0.3 (0.4)	0.0 (0.3)	0.00	1.0000	0.00	0.73
Eyes	4	0.1 (0.2)	0.1 (0.2)	0.0 (0.3)	0.00	1.0000	0.00	-0.50
Chest/Respiratory	4	0.9 (0.5)	0.9 (0.4)	0.0 (0.2)	0.33	0.7608	0.07	0.92
Gastrointestinal	4	0.3 (0.5)	0.1 (0.3)	0.2 (0.6)	0.68	0.5472	0.40	-0.09
Body/Systemic	4	0.2 (0.1)	0.3 (0.4)	-0.1 (0.3)	-0.96	0.4058	1.00	0.44
Total Score	4	0.4 (0.2)	0.4 (0.2)	-0.0 (0.1)	-0.45	0.6807	0.11	0.90

[1]Number of study participants with no change in flu symptom at day Y. No data available for Day 4 to Day 5.

[2]Mean difference = average Day X FLU-PRO score - average Day Y FLU-PRO score (ex. Day 1 score - Day 2 score); p value from paired t-test.

Table S9a. Construct Validity: 32-Item FLU-PRO Scale Correlations with Other PRO Measures at Day 1 in Non-Hospitalized Influenza-Positive Patients

	Domains and Total Score						
Day	**Nose**	**Throat**	**Eyes**	**Chest/ Respiratory**	**Gastrointestinal**	**Body/ Systemic**	**Total Score**
Day 1							
Patient Global Rating of Flu Severity	0.31***	0.28**	0.31***	0.39***	0.31***	0.63***	0.62***
Patient Global Rating of Physical Health	-.17*	-.25**	-.35***	-.23*	-.05	-.33***	-.35***
Patient Global Assessment of Interference in Daily Activities	0.26**	0.21*	0.40***	0.37***	0.26**	0.64***	0.58***

[1]Spearman correlation coefficients: [1]p<0.0001, [2]p<0.001, [3]p<0.05

Table S9b. Construct Validity: 32-Item FLU-PRO Scale Correlations with Other PRO Measures at Day in Hospitalized Influenza-Positive Patients

	Domains and Total Score						
Day	**Nose**	**Throat**	**Eyes**	**Chest/ Respiratory**	**Gastrointestinal**	**Body/ Systemic**	**Total Score**
Day 1							
Patient Global Rating of Flu Severity	0.20	0.31*	0.44**	0.38*	0.42*	0.60***	0.54***
Patient Global Rating of Physical Health	0.12	-.13	-.05	-.28*	-.09	-.12	-.13
Patient Global Assessment of Interference in Daily Activities	-.22	0.04	-.08	-.01	0.05	0.13	0.02

[1]Spearman correlation coefficients: [1]p<0.0001, [2]p<0.001, [3]p<0.05

Table S10a. Known-Groups Validity: 32-Item FLU-PRO Scores by Patient Global Rating of Disease Severity, Day 1 in Non-Hospitalized Influenza-Positive Patients

Scale	Patient Global Rating of Flu Severity Mean (SD)			F Value (p value)[1]	Pairwise Comparisons[2]
	No/Mild Symptoms (N=38)	Moderate Symptoms (N=59)	Severe/Very Severe Symptoms (N=71)		
Nose	1.40 (0.90)	1.68 (0.92)	2.20 (1.09)	9.3***	2***,3*
Throat	1.01 (0.85)	1.25 (1.01)	1.85 (1.24)	9.0***	2***,3**
Eyes	0.61 (0.92)	0.80 (0.94)	1.31 (1.07)	7.7***	2**,3*
Chest/Respiratory	1.40 (0.65)	1.68 (0.85)	2.21 (0.85)	14.1***	2***,3**
Gastrointestinal	0.26 (0.43)	0.47 (0.64)	0.98 (0.94)	13.6***	2***,3***
Body/Systemic	1.10 (0.66)	1.70 (0.77)	2.57 (0.74)	54.5***	1***,2***,3***
Total Score	1.04 (0.45)	1.41 (0.56)	2.06 (0.57)	49.5***	1**,2***,3***

[1]p values are: *<0.05, **<0.01, ***<0.001.

[2]Pairwise comparisons between means will be performed using Scheffe's 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.

Table S10b. Known-Groups Validity: 32-Item FLU-PRO Scores by Patient Global Rating of Disease Severity, Day 1 in Hospitalized Influenza-Positive Patients

Scale	Patient Global Rating of Flu Severity Mean (SD)			F Value (p value)[1]	Pairwise Comparisons[2]
	No/Mild Symptoms (N=12)	Moderate Symptoms (N=18)	Severe/Very Severe Symptoms (N=23)		
Nose	0.92 (0.76)	1.15 (0.97)	1.39 (1.16)	0.9	
Throat	0.36 (0.52)	1.19 (1.13)	1.36 (1.20)	3.7*	2*
Eyes	0.22 (0.67)	0.89 (1.11)	1.55 (1.52)	4.7*	2*
Chest/Respiratory	1.27 (0.84)	1.95 (0.87)	2.17 (0.92)	4.1*	2*
Gastrointestinal	0.38 (0.42)	0.50 (0.70)	1.30 (1.14)	6.2**	2*,3*
Body/Systemic	0.80 (0.56)	1.28 (0.74)	2.18 (0.80)	15.9***	2***,3**
Total Score	0.77 (0.49)	1.27 (0.62)	1.83 (0.77)	10.4***	2***,3*

[1]p values are: *<0.05, **<0.01, ***<0.001.

[2]Pairwise comparisons between means will be performed using Scheffe's 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.

Table S11a. Responsiveness of 32-Item FLU-PRO by Patient Return to Usual Health (N=111) or Return to Usual Activities (N=89), Day 1 to Day 7 in Non-Hospitalized Influenza-Positive Patients

Scale	Responders[1]			Non-Responders			p-value[2]
	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (SD)	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (SD)	
Nose							
Usual Health	1.7 (1.1)	0.4 (0.4)	1.4 (0.1)	1.9 (1.0)	0.9 (0.8)	0.9 (0.1)	<.0001
Usual Activities	1.9 (1.0)	0.6 (0.6)	1.2 (0.1)	1.8 (1.3)	1.0 (1.1)	0.8 (0.1)	0.0114
Throat							
Usual Health	1.2 (1.0)	0.1 (0.3)	1.4 (0.1)	1.7 (1.2)	0.6 (0.7)	1.0 (0.1)	0.0013
Usual Activities	1.5 (1.1)	0.3 (0.6)	1.2 (0.1)	2.0 (1.3)	0.8 (1.0)	0.9 (0.1)	0.0466
Eyes							
Usual Health	1.0 (0.9)	0.1 (0.4)	0.9 (0.1)	1.1 (1.2)	0.3 (0.7)	0.7 (0.1)	0.0955
Usual Activities	1.1 (1.0)	0.2 (0.6)	0.9 (0.1)	1.3 (1.2)	0.4 (0.7)	0.8 (0.1)	0.4852
Chest/Respiratory							
Usual Health	1.4 (0.8)	0.5 (0.6)	1.1 (0.1)	2.1 (0.8)	1.3 (0.7)	0.7 (0.1)	<.0001
Usual Activities	1.9 (0.8)	0.9 (0.7)	1.0 (0.1)	2.2 (1.0)	1.6 (1.0)	0.5 (0.1)	0.0006
Gastrointestinal							
Usual Health	0.4 (0.8)	0.0 (0.2)	0.6 (0.0)	0.7 (0.8)	0.2 (0.4)	0.4 (0.0)	0.0094
Usual Activities	0.7 (0.8)	0.1 (0.3)	0.6 (0.0)	0.7 (0.9)	0.4 (0.5)	0.3 (0.1)	0.0006
Body/Systemic							
Usual Health	1.6 (0.9)	0.1 (0.3)	1.7 (0.1)	2.1 (0.9)	0.6 (0.6)	1.3 (0.1)	0.0001
Usual Activities	2.0 (0.9)	0.3 (0.5)	1.7 (0.1)	2.3 (0.8)	1.0 (0.8)	1.1 (0.1)	<.0001
Total Score							
Usual Health	1.3 (0.6)	0.2 (0.3)	1.3 (0.1)	1.7 (0.7)	0.7 (0.5)	0.9 (0.0)	<.0001
Usual Activities	1.7 (0.7)	0.5 (0.4)	1.2 (0.1)	1.9 (0.6)	1.0 (0.7)	0.8 (0.1)	<.0001

[1]Responders are defined as patients responding that they have returned to their usual activities at Day 7.

[2]From ANCOVA to compare LSMean change scores between responders and non-responders adjusting for Day 1 scores.

Table S11b. Responsiveness of 32-Item FLU-PRO by Patient Return to Usual Health (N=36) or Return to Usual Activities (N=37), Day 1 to Day 7 in Hospitalized Influenza-Positive Patients

Scale	Responders[1]			Non-Responders			p-value[2]
	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (SD)	Day 1 Mean (SD)	Day 7 Mean (SD)	Change Score LSMean (SD)	
Nose							
Usual Health	1.4 (0.7)	0.4 (0.7)	0.8 (0.2)	1.0 (0.9)	0.5 (0.6)	0.5 (0.1)	0.2860
Usual Activities	1.3 (0.8)	0.6 (0.6)	0.5 (0.1)	0.9 (0.9)	0.5 (0.7)	0.6 (0.1)	0.9580
Throat							
Usual Health	0.8 (1.1)	0.1 (0.3)	0.8 (0.2)	1.1 (1.1)	0.4 (0.8)	0.6 (0.1)	0.3852
Usual Activities	0.8 (1.0)	0.2 (0.3)	0.7 (0.1)	1.2 (1.1)	0.5 (0.9)	0.6 (0.1)	0.3645
Eyes							
Usual Health	1.0 (1.3)	0.1 (0.3)	0.9 (0.2)	1.1 (1.4)	0.4 (0.8)	0.6 (0.1)	0.3096
Usual Activities	0.6 (1.0)	0.1 (0.3)	0.8 (0.1)	1.4 (1.4)	0.6 (0.9)	0.6 (0.1)	0.3756
Chest/Respiratory							
Usual Health	1.4 (0.6)	0.6 (0.7)	1.1 (0.2)	1.9 (0.9)	1.1 (0.7)	0.8 (0.1)	0.2676
Usual Activities	1.6 (0.9)	0.6 (0.6)	1.1 (0.1)	1.9 (0.9)	1.2 (0.7)	0.6 (0.1)	0.0410
Gastrointestinal							
Usual Health	0.6 (0.9)	0.5 (1.1)	0.1 (0.2)	0.6 (0.7)	0.4 (0.5)	0.2 (0.1)	0.4178
Usual Activities	0.6 (0.7)	0.4 (0.8)	0.2 (0.1)	0.6 (0.8)	0.3 (0.5)	0.2 (0.1)	0.6773
Body/Systemic							
Usual Health	1.2 (0.8)	0.6 (0.8)	0.7 (0.2)	1.5 (1.1)	0.7 (0.7)	0.7 (0.1)	0.9550
Usual Activities	1.2 (1.0)	0.5 (0.6)	0.8 (0.1)	1.5 (1.0)	0.8 (0.7)	0.7 (0.1)	0.5442
Total Score							
Usual Health	1.2 (0.7)	0.5 (0.7)	0.7 (0.1)	1.3 (0.8)	0.7 (0.5)	0.6 (0.1)	0.5134
Usual Activities	1.1 (0.7)	0.5 (0.5)	0.7 (0.1)	1.4 (0.8)	0.7 (0.6)	0.6 (0.1)	0.3487

[1]Responders are defined as patients responding that they have returned to their usual health at Day 7.

[2]From ANCOVA to compare LSMean change scores between responders and non-responders adjusting for Day 1 scores.

Figure S1. Preliminary Hypothesized Conceptual Framework: Symptoms of Influenza

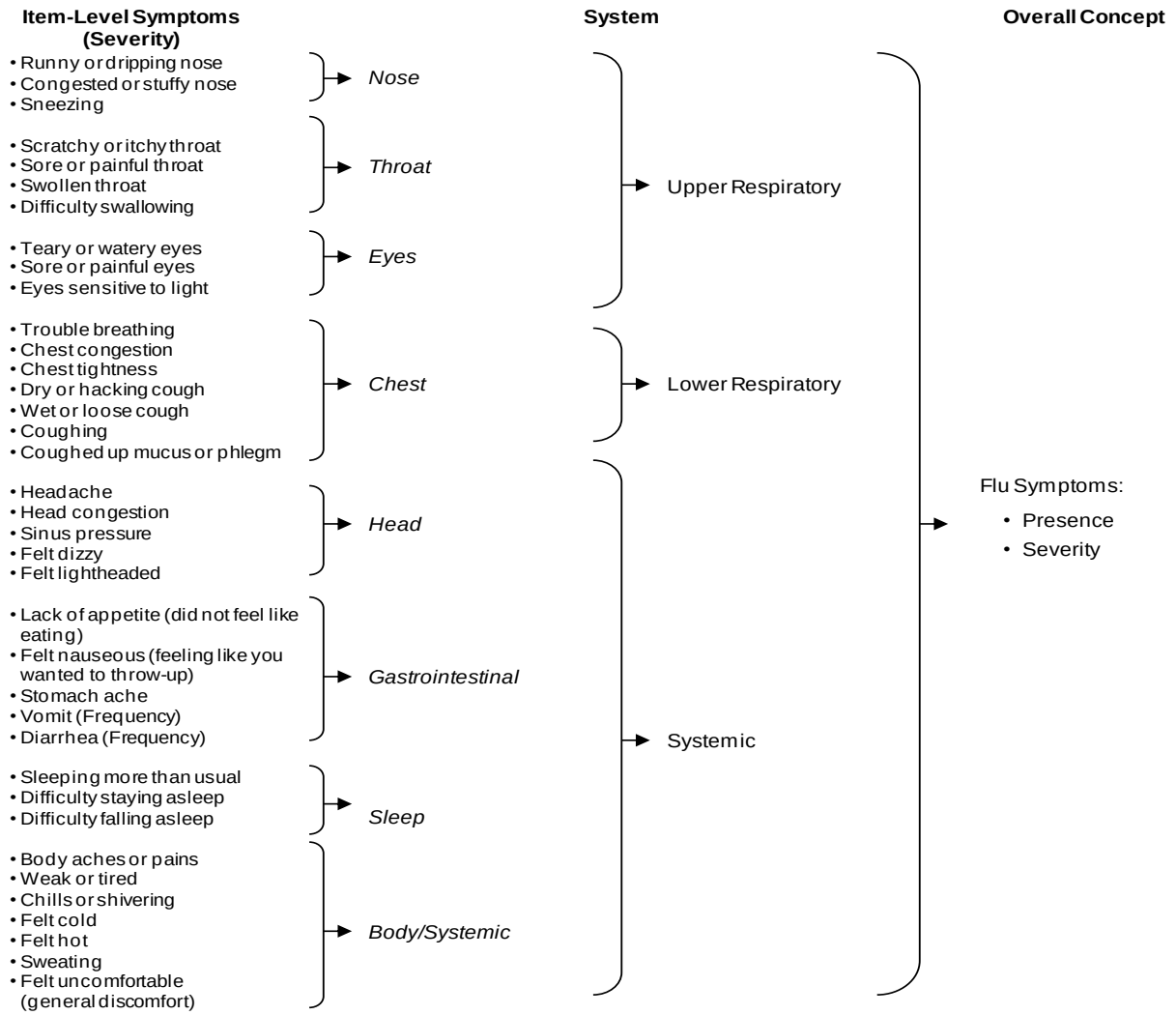
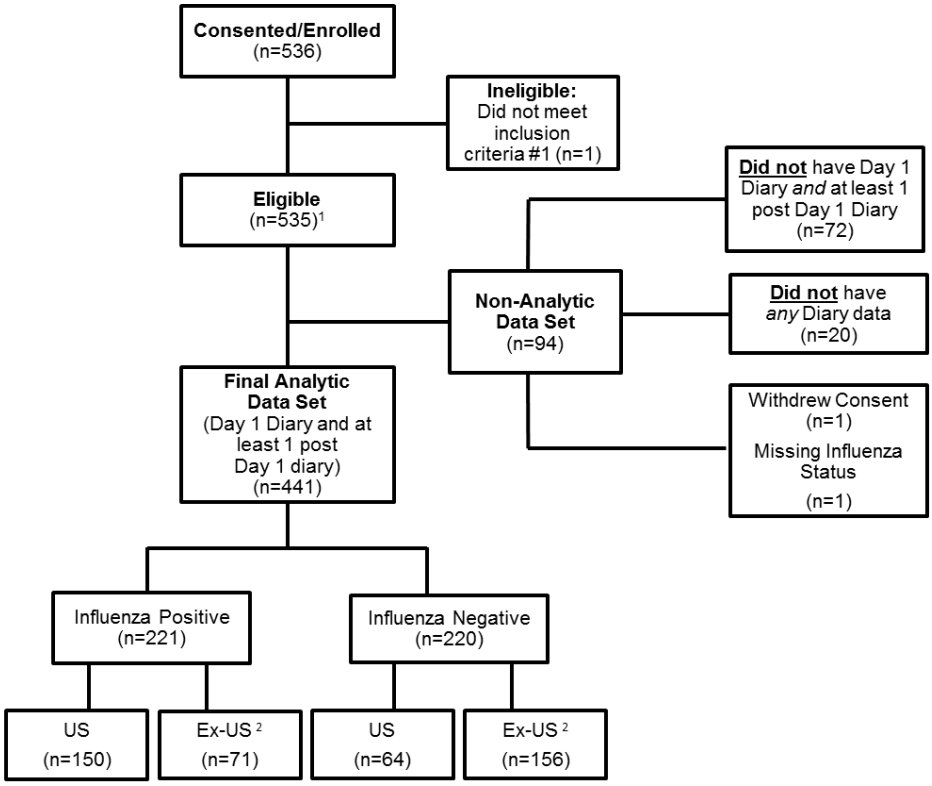


Figure S2. FLU PRO Stage 3: Subject Disposition Chart



[1]28 of 535 participants were eligible but did not successfully complete the study due to LTFU, death, hospitalization, etc. Since these participants had Day 1 Diary and at least 1 post Day 1 Diary, they were included in the final analysis set.
[2]Ex-US countries include Mexico, Argentina, and UK

Figure S3a. FLU-PRO Domain and Total Score by Diary Days 1 to 14: Hospitalized, Influenza-Positive Patients

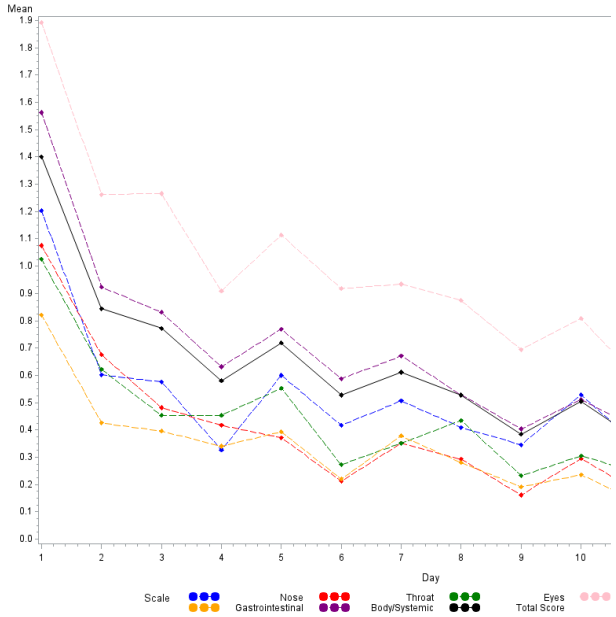


Figure S3b. FLU-PRO Domain and Total Score by Diary Days 1 to 14: Non-Hospitalized, Influenza-Positive Patients