Development and Validation of the Influenza Intensity and Impact Questionnaire (FluiiQ™)

Richard H. Osborne, BSc, PhD1,∗, Josephine M. Norquist, MS2, Gerald R. Elsworth, BSc, PhD3, Lucy Busija, BA (Hons), MSci1, 3, Vinay Mehta, PhD, MS2, Tim Herring, MPH2, Swati B. Gupta, DrPH, MPH2

1Public Health Innovation, Population Health Strategic Research Centre, Faculty of Health, Deakin University, Melbourne, Victoria, Australia; 2Merck Sharpe & Dohme Corp. UGID-60, North Wales, PA, USA; 3Population Health and Macular Research Units, Center for Eye Research Australia, The University of Melbourne, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, Australia

ABSTRACT

Objective: Clinical trials of new agents to reduce the severity and impact of influenza require accurate assessment of the effect of influenza infection. Because there are limited high-quality adult influenza Patient Reported Outcomes (PRO) measures, the aim was to develop and validate a simple but comprehensive questionnaire for epidemiological research and clinical trials. Methods: Construct and item generation was guided by the literature, concept mapping, focus groups, and interviews with individuals with laboratory-confirmed influenza and expert physicians. Items were administered to 311 people with influenza-like illness (ILI) across 25 US sites. Analyses included classical psychometrics, structural equation modeling (SEM), and Rasch analyses. Results: Concept mapping generated 149 concepts covering the influenza experience and clustered into symptoms and impact on daily activities, emotions, and others. Items were drafted using simplicity and brevity criteria. Eleven symptoms from the literature underwent review by physicians and patients, and two were removed and one added. The symptoms domain factored into systemic and respiratory symptoms, whereas the impact domains were unidimensional. All domains displayed good internal consistency (Cronbach α > 0.8) except the three-item respiratory domain (α = 0.48). A five-factor SEM indicated excellent fit where systemic, respiratory, and daily activities domains differentiated patients with ILI or confirmed influenza. All scales were responsive over time. Conclusions: Patient and clinician consultations resulted in an influenza PRO measure with high validity and good overall evidence of reliability and responsiveness. The Influenza Intensity and Impact Questionnaire (FluiiQ™) will improve the evaluation of existing and future agents designed to prevent or control influenza infection by increasing the breadth and depth of measurement in this field. Keywords: clinical trial, influenza, patient-reported outcomes, psychometrics, surveillance, questionnaire validation.

Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Influenza clinical trials usually focus on outcomes such as laboratory-documented infection using viral culture and/or increase in antibody titers to influenza antigens [1–3]. In drug and vaccine trials, however, new agents are not expected to prevent infection; rather, they are expected to reduce severity and duration of symptoms (e.g., fever, cough, malaise) functional impact, and time to return to daily activities. To determine the efficacy or effectiveness of investigational agents on influenza severity, patient-reported outcomes (PROs) are used as primary and secondary end points in clinical trials. Psychometric properties of PRO measures are now commonly evaluated by regulatory authorities to ensure that they generate reliable evidence to inform decisions regarding the safety and efficacy of a new product [4].

The release of the US Food and Drug Administration guidance (FDA) “Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims” [4] in December 2009 significantly raised the level of rigor required for the development and validation for all PRO measures included in clinical trials to support labeling claims. Questionnaires accepted by regulatory bodies before the release of the FDA guidance are now facing new scrutiny because they may be lacking in several of the required development and validation factors. One such factor, content validity, has been repeatedly recognized as a critical step in establishing the validity of the items used in measures based on patient self-report. The guidance points to the need for direct patient input during the development and refinement items in PRO measures. It requires all items, response options, and recall periods to be demonstrated as relevant and important to patients with the particular disease or illness assessed by the measure.

There are limited high-quality adult PRO measures available and even fewer validated PRO measures on influenza symptom...
severity and impact. After an extensive literature review, although there is one well-validated acute respiratory illness scale in children [5], to the knowledge of the authors, the only validated adult PROs in this area are the Influenza Symptom Severity Scale (ISS) and the Influenza Impact Wellbeing Scale (IIWS) questionnaires [6]. These measures were developed in 2000 to measure the symptoms and impact of influenza. These questionnaires were developed based on literature review and physician input and were shown to have evidence of internal consistency and construct validity. Given, however, the FDA guidance documents discussed previously [4,7], these questionnaires had some shortcomings, including the absence of consultation with patients in the development of the questions, absence of cognitive testing of questions, and inadequate testing in individuals with laboratory-confirmed influenza.

The aim of this study was to develop and validate a comprehensive measure of the symptoms and impact of influenza on individuals with influenza-like illness (ILI) and laboratory-confirmed influenza for use in epidemiological studies and trials of vaccine and pharmaceutical agents.

Methods

To improve on the previously developed ISS and IIWS [6], questionnaire development studies were conducted from May 2005 through November 2006 and consisted of the following: 1) influenza expert physician consultations; 2) conduct of the Item and Questionnaire Development Protocol, which consisted of patient rankings of bothersomeness of individual symptoms, patient consultation using concept mapping [8] in Australia, and a traditional focus group in the United States; 3) development of a conceptual model; and 4) patient cognitive testing of newly developed scales. Methods for each activity are briefly outlined below in the following. An area of impact that we did not seek to cover in the construct development was the impact of influenza on working and associated economic impacts. These impacts were omitted because we regarded them as being adequately covered by other tools (e.g., employment and cost diaries).

Expert physician consultations

Expert physician views were sought to elicit feedback on symptoms that patients with influenza experience and the impact that influenza has on patient well-being. We also sought feedback on the relevance and wording of items included in the two candidate measures (ISS and IIWS). Semistructured telephone interviews were conducted by an experienced interviewer with six general physicians in the United States with familiarity with treating patients with influenza and one physician with extensive influenza academic research experience. An interview guide was used to ensure that key topics were covered, but it was not adhered to verbatim. Experts were chosen based on recommendations made by the Merck Research Laboratories clinical team, by field monitors, or by geographic location. Experts were not compensated for their time, and interviews were not audio-recorded. Field notes were taken by two observers as well as by the interviewer.

Item and questionnaire development study protocol

Participants 18 years of age and older with ILI (defined as fever ≥37.8°C plus at least one of the following symptoms: cough, sore throat, runny nose, or muscle aches on one day or, in the absence of fever, the presence of at least two of these symptoms of moderate severity for two consecutive days) were enrolled from general practitioners and health clinics in the United States and Australia. Independent ethical review committees at each study site approved the study protocol, and all participants provided written informed consent. At the time of enrollment, to explore the content validity of the symptoms within the ISS and the degree to which patients endorse the items, patients with ILI rated each symptom on both importance and bothersomeness on 6-point scales (i.e., from 0 [not at all bothersome] to 5 [very bothersome]). Patients were able to list additional symptoms in an open-ended “other” category. Bothersomeness was only assessed for the ISS as it was specific to symptom severity and therefore was not conducted for the IIWS components. Bothersomeness scores were analyzed and reported as ranks, frequencies, and means with SD. After enrollment in the study was completed, those patients found to have laboratory-confirmed influenza infection took part in a separate focus group. Laboratory-confirmed influenza infection was defined as any participant who was positive for influenza infection via viral culture and/or by a fourfold or more increase in hemagglutinin inhibition antibody titer. Patient focus groups were one traditional focus group conducted in the United States and three concepts mapping workshops (see later), conducted in Australia. The recall period between enrollment in the study and focus groups (traditional and concept mapping) did not exceed 5 months for all patients.

Traditional focus group (United States)

Based on enrollment into the questionnaire development protocol conducted in the United States, and described previously, patients were identified as having laboratory-confirmed influenza at the University of Virginia student health clinic and subsequently were enrolled in a focus group. The objectives of the focus group were to ask patients to describe the constellation of symptoms caused by their current influenza episode, identify areas of their lives affected by influenza as well as obtain direct feedback on the ISS and IIWS questionnaires. The focus group was facilitated by an external vendor using a standard focus group guide.

Concept mapping workshops (Australia)

Based on enrollment in the questionnaire development protocol conducted in Australia, and described previously, patients were identified as having laboratory-confirmed influenza at two general practitioner clinics in Brisbane, Melbourne, and Sydney and subsequently enrolled in concept mapping workshops. Concept mapping, which includes a structured nominal group process and is well suited to questionnaire development [8,9], was undertaken to identify broad potential impacts of influenza on people and generate potential questionnaire items [9–11]. Participants were asked to respond to the following seeding statement: “Thinking as broadly as possible, generate statements to describe how your recent episode of flu affected you and your life.” Responses were printed on individual cards, and participants were then required to intuitively sort them into categories. Participants were also asked to rate each response according to two dimensions: impact and duration. The sort data were then analyzed during the workshop using specialized software (Concept System software, Concept Systems Inc., New York, and ClustanGraphics, Clustan Limited, Edinburgh), which uses two multivariate statistical methods (multidimensional scaling and cluster analysis) to analyze the patterns among statements generated and subsequently sorted by participants. The outcome of this process provides a visual map that groups responses into clusters. The final step involves displaying the map to participants who are asked to come to a consensus around the meaningfulness of statement groupings (i.e., clusters) and identify overarching descriptors of the underlying theme of each cluster.
Item generation
The concepts within each broad cluster informed the development of items. Where possible, the patient's original wording was preserved for content validity. Item generation was highly structured with constant reference to how a person with influenza may respond to each item. Each item had to be succinct and concise because patients completing questions might be quite ill. The content had to be relevant across disease severity and to the broad target population (e.g., across age, sex, education, and language groups).

Concepts were submitted to a restricted set of experts, including three psychometricians, five epidemiologists, two statisticians, and a medical editor for review. After numerous revisions, a final set of items for the PRO measure within each target construct was derived through consensus by the experts in a full-day face-to-face workshop.

Development of a conceptual model of symptoms and impact of influenza
Based on the literature review, physician consultations, traditional focus group, and concept mapping results, a causal model was conceived to guide questionnaire development and the subsequent examination of construct validity. It was proposed that an infection would result in symptoms being caused by the pathophysiology of the disease. These symptoms then affect function in several ways, namely, through physical activities, emotions, and psychosocial factors (Fig. 1). The focus of the measures being developed was limited to the inclusion of proximal effects (symptoms and immediate impact of these symptoms) rather than further downstream effects such as economic burden. With a model established a priori, a subsequent finding of robust factorial separation of items into factors together with empirical support for the causal model would provide evidence of construct validity [12].

Cognitive testing
Once items were derived using the described procedures, each item and the overall format of the questionnaire underwent cognitive testing. As the target population of the questionnaire is patients with ILI or confirmed influenza infection, a general sample of people who reported having had physician-diagnosed influenza infection within the past 3 years were recruited for participation. Subjects were asked to review the study materials and then participate in in-depth, individual semistructured interviews. Participants were asked to provide their understanding of words and phrases and comment on the instructions, the ease of completion, and the appropriateness of response options. The interviews were not recorded; however, comprehensive notes were taken.

Questionnaire validation
Questionnaire validation protocol
People 18 years of age and older were recruited across 25 US sites including general practitioner offices, outpatient health clinics (hospital or community), and university health clinics. Patients were included if they presented within 48 hours of onset of ILI as defined previously. Using a computer-generated series of random numbers, patients were assigned with equal probability to receive the questionnaire in a paper-based format or in an electronic interactive voice response system format once per day for 14 days. The equivalence of these delivery modes will be the subject of future reports. Data from day 1 were used for structural equation modeling (SEM) and Rasch analyses (described in the following). A central institutional review board was used to obtain ethics committee approval across all sites, and all participants provided written informed consent.

Patients assessed and recorded the level of their symptoms daily throughout the 14 days of the study. Additionally, patients and physicians were asked to assess overall severity of symptoms using the Patient or Physician Assessment of Global Severity Question (GSQ) (4-point scale ranging from 0 [none] to 4 [severe]). The global severity questions were completed at days 1, 7, and 14 (days 1 and 14 only for physicians). Patients completed the EQSD thermometer (www.euroqol.org/home.html), a widely used generic visual analogue measure of health-related quality of life (QoL) with anchors from 0 (worst imaginable health state) to 100 (best imaginable health state). This was measured on days 1 and 14. A generic global health question (GGHQ) was also completed by patients; it asked patients to rate their current health status on a 5-point scale (i.e., excellent, very good, good, fair, poor; coded 0 to 4). This was completed on days 1, 7, and 14.

Statistical methods
Descriptive statistics were generated for each item to determine the extent of missing values and floor and ceiling effects across domains. Extreme scores were classified as the bottom decile of a domain score (i.e., <0.3 where the scale range is from 0.0 to 3.0); conversely ceiling effects were the top decile of a domain (i.e., score > 2.7).

Missing data
A small number of discrete data points were missing for the questionnaire data. To maximize data for the SEM, values for missing data were imputed using the Expectation-Maximization algorithm (SPSS version 12.0) and were rounded back to the nearest discrete value. For the remaining analyses, missing data were treated using pairwise deletion, in which cases with missing values were deleted on a variable-by-variable basis.

Factor analysis
Confirmatory factor analysis (CFA) was undertaken with Mplus (version 6.1) to determine whether hypothesized latent variables representing dimensions loaded on the specified items. A sequence of one-factor models and a final multifactor model were fitted to the data. The models were fitted following the general method for the analysis of ordinal data using polychoric correlations recommended by Jöreskog [13] and the robust mean- and variance-adjusted weighted least squares (WLSMV) estimator available in Mplus. Preliminary studies of this approach to CFA with categorical or ordinal data have suggested that it performs well with relatively small samples [14].

Judgments of how well the data fitted the CFA and the resulting SEM models were evaluated using goodness-of-fit statistics. For a

Fig. 1 – Causal model of the impact of influenza infection on affected individuals.
reasonably good fit, the following criteria were used [15,16]: root mean square error of approximation (RMSEA) < 0.08 for an acceptable fit and < 0.05 for a good fit; confirmatory fit index (CFI) and Tucker-Lewis index (TLI) > 0.95; χ²/df < 2.0; and weighted root mean square residual (WRMR) < 1.00.

Criteria to judge whether the estimates of association between factors, items, and error terms were substantive included:

1. Loadings of factors on items < 0.50 were regarded as low. Further assessment included whether the loadings of each item within a factor were approximately similar.

2. Individual correlations between residual errors of items were regarded as substantial if they were > 0.15. Correlations between residuals suggesting that there may be a more complex factor structure than that hypothesized and subfactors were then explored.

3. Correlations between factors were regarded as excessive if they were > 0.90.

Rasch analysis

Rasch analyses using RUMM2020 [17] were conducted to explore item characteristics and further test construct validity of domains [18]. The Rasch model summarizes the relationship between the characteristics of an item and the characteristics of an individual in terms of probability of endorsing an item of a given difficulty (the extent to which the item is difficult to agree with) by a person of certain ability (level of underlying trait). Construct validity of a scale is supported when the data have good fit with the Rasch model. Assessment of fit of a domain was based on item-trait interaction χ² and Person Separation Index (PSI). Good model fit is indicated by a nonsignificant χ² value and is achieved when the endorsement of higher response options is uniformly associated with higher levels of an underlying trait. A PSI indicates a domain’s ability to discriminate between the levels of an underlying trait, with values of > 0.80 indicating good discriminability. Fit of individual items was assessed using fit residuals and item χ², both of which represent the differences between the observed pattern of responses for the item and the pattern of responses predicted from the Rasch model. Poor item fit is suggested by significant χ² and fit residual values < -2.5 (indicative of redundant items) and >2.5 (suggesting item content is incongruent within a domain).

Rasch analysis was also used to assess item threshold ordering and differential item functioning (DIF). Thresholds define the boundaries between the item response option categories and represent the level of the trait being measured such that the probability of choosing either of the adjacent response categories is 0.5 [19]. Disordered thresholds are undesirable because they tend to occur when respondents have difficulty in discriminating between response options such as when response options are ambiguous or do not clearly represent different intensity of symptoms. Disordered thresholds can also be due to item content that is ambiguous or double-barreled. DIF, indicated by significant F-ratios from analysis of variance, suggests that an item is not unidimensional and that characteristics other than levels of the underlying trait may be influencing the respondent’s answers. DIF was assessed across subgroups based on age, sex, race, and influenza status (confirmed influenza vs. ILI). Bonferroni corrections (α = 0.05/number of domain items) were applied to all Rasch model analyses.

To investigate whether one or more items do not relate to the intended construct within a scale, principal components analysis of residual correlations (PCARC) was conducted [20]. PCARC involves analysis of item residuals to identify the presence of secondary variables that could potentially account for the correlations among items after removing variance due to the primary construct (i.e., influenza symptoms).

Reliability, discriminant and convergent validity, known groups validity, and responsiveness

Once the items and domains were confirmed, internal consistency was assessed with Cronbach’s α coefficients where good internal consistency is defined as an α coefficient > 0.70 [19]. Discriminant and convergent validity (elements of construct validity) was assessed by examining the association between the Influenza Intensity and Impact Questionnaire (FluiiQ™) domains and the other measures included in the validation study. Specifically, to assess construct validity, Spearman’s correlation coefficients were obtained between the FluiiQ™ and the Patient and Physician GSQ, the EQ5D thermometer, and the GGHQ. It was hypothesized that correlations would be moderate with the FluiiQ™ positively correlated with the Patient and Physician GSQ, and the GGHQ and negatively correlated with the EQ5D thermometer. The range used to qualify the strength of correlations between the FluiiQ™ and other measures was: r < 0.25 = weak; 0.25 to 0.75 = moderate; r ≥ 0.75 = strong [21,22].

Known group validity was evaluated by comparing participants with confirmed influenza with those with ILI across all domains using means and 95% confidence intervals (CIs). It was expected that those with confirmed influenza would report higher scores across all domains than those with ILI. To assess responsiveness, change (Δ) from day 1 to day 14 was used to calculate effect size (ES: mean Δ/SD at day 1) and standardized response mean (SRM: mean Δ/SD of Δ). An effect size with an absolute value > 0.8 is indicative of good responsiveness [22].

Results

Item development

Physician consultations

Six general practitioners and one expert influenza researcher were consulted and identified common influenza symptoms including febrile illness (sudden onset), body aches (both myalgia and arthralgia), headache, sore throat, neck pain, chills, extreme fatigue (significant disability), cough, dizziness, malaise, as well as lack of appetite. Gastrointestinal symptoms (nausea, vomiting) were usually regarded as secondary to other infections. All physicians commented that although there is a “typical” constellation of influenza symptoms, these may vary in frequency and severity, depending on the particular strain and host factors. All physicians indicated that the level of prostration was a good indicator of symptom severity. All noted that influenza involved a gradual reduction of each symptom until a return to daily activities. Several physicians noted that symptoms such as a nagging cough may persist for 5 to 20 days after initial infection.

Physicians also reviewed the ISS and IIWS questionnaires [6]. They indicated that a 4-point scale was adequate to capture the symptom intensity range (absent to severe). Of the 11 symptoms from the ISS, it was indicated that the symptom lack of concentration was not sufficiently specific to influenza and was likely to be secondary to other infections. Therefore, lack of concentration and nausea were removed. Although the remainder of the items was endorsed, neck pain was raised as a new item and included. Finally, they suggested the term nasal congestion was potentially difficult for some patients. There was agreement that the three-item IIWS questionnaire (visual analogue scales covering general health, global usual activities, and sleep) poorly captured the impact of influenza. Some anchors for the scales were regarded as potentially objectionable, for example one anchor was “worse than being dead.” It was agreed
that this scale needed redevelopment to capture the true impact of influenza on patients.

**Bothersomeness rating of symptoms**

A total of 67 patients provided a bothersomeness rating of candidate symptoms. Of these, 23 were culture-confirmed influenza cases and 25 were recruited in the United States. The four top-ranked items in both groups (ILI and confirmed influenza) based on bother score included feeling feverish, fatigue, aches and pains, and disturbed sleep, suggesting similar symptom profiles for both groups. Nausea was the lowest ranked item in both groups and was least frequently endorsed. Bother scores were higher in the top-ranked items for the confirmed influenza group, suggestive of more severe symptoms in this group compared with the ILI group (Table 1).

### Table 1 - Bothersomeness of influenza symptoms derived from US and Australian samples with confirmed influenza or influenza-like illness.

<table>
<thead>
<tr>
<th>Draft item</th>
<th>Confirmed influenza (n=23)</th>
<th>Influenza-like illness (n=44)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency endorsed</td>
<td>Bothersomeness, mean (SD)</td>
</tr>
<tr>
<td>Feverish</td>
<td>23</td>
<td>3.87 (0.97)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>22</td>
<td>3.95 (1.13)</td>
</tr>
<tr>
<td>Aches and pains</td>
<td>23</td>
<td>3.70 (1.34)</td>
</tr>
<tr>
<td>Disturbed sleep</td>
<td>21</td>
<td>3.67 (1.01)</td>
</tr>
<tr>
<td>Nasal congestion</td>
<td>22</td>
<td>3.41 (1.18)</td>
</tr>
<tr>
<td>Cough</td>
<td>22</td>
<td>3.41 (1.50)</td>
</tr>
<tr>
<td>Headache</td>
<td>22</td>
<td>3.27 (1.20)</td>
</tr>
<tr>
<td>Sore throat</td>
<td>22</td>
<td>2.90 (1.15)</td>
</tr>
<tr>
<td>Poor concentration</td>
<td>20</td>
<td>2.80 (1.15)</td>
</tr>
<tr>
<td>Loss of appetite</td>
<td>20</td>
<td>2.20 (1.39)</td>
</tr>
<tr>
<td>Nausea</td>
<td>14</td>
<td>2.43 (1.34)</td>
</tr>
</tbody>
</table>

**Traditional focus group (United States)**

Six participants, 19 to 31 years of age, identified a list of influenza symptoms experienced during their recent influenza illness. They identified fatigue and fever as the symptoms that bothered them the most. Additionally, they identified the most significant impact of influenza as their inability to do daily activities (such as schoolwork and social activities) and concentrate on tasks. Although they thought that all symptoms in the ISS were relevant and the questionnaire was easy to answer and clear, this was not the case for the IIWS. They thought that the IIWS questions were vague and difficult to answer due to unclear wording. There was general agreement that it needed modification and additional constructs added if it were to adequately capture the impacts of influenza beyond direct symptoms, consistent with the physi-

---

Fig. 2 – Concept map generated from people recently experiencing influenza symptoms with the group’s interpretation of the domains of influenza impact.
Maps (see Figure 2 for an example) enabled overarching concepts to be identified, and work was undertaken in Australia (2 in Brisbane and 1 in Sydney). A total of 149 statements regarding laboratory-confirmed influenza in the past 5 months, were undertaken by 16 people, 21 to 55 years of age who had laboratory-confirmed influenza. Three workshops with 16 people, 21 to 55 years of age who had laboratory-confirmed influenza were undertaken during the upcoming Southern Hemisphere influenza season in Australia in the form of concept mapping workshops to comprehensively redevelop the ISS and IIWS.

Concept mapping workshops (Australia)
Three workshops with 16 people, 21 to 55 years of age who had laboratory-confirmed influenza in the past 5 months, were undertaken in Australia (2 in Brisbane and 1 in Sydney). A total of 149 statements were obtained in response to the seeding statement. The generated maps (see Figure 2 for an example) enabled overarching concepts to be identified, Concatenation of the three workshops involved consideration of individual statement content, commonalities between the descriptors generated, and consideration of the conceptual model (i.e., symptoms or the consequence of symptoms). Key dimensions to emerge included symptoms, impact on routine physical activities, and psychosocial aspects including emotional, social relationships, and economic concerns. Statements that did not reflect the seeding statement or were nonspecific were removed including “put neck out from sneezing” and “unexpected, didn’t see it coming.” Statements were also removed if they were obscure or not direct impacts (“blister on face” and “irritation of teeth”). Statements were also removed if they related directly to economic and work effects because these dimensions had been chosen a priori to not be part of the current questionnaire.

Draft items, scales, and response options
The final domains and items are shown in Table 2, and in summary they included:

**Influenza symptoms (10 items).** This domain covered systemic symptoms (headache, fever, body aches, fatigue, neck pain, sleep, and appetite) and upper respiratory symptoms (cough, sore throat, and nasal congestion). Response options (0–3) were none, mild, moderate, and severe.

**Impact on daily activities (6 items).** This domain covered the physical impact of influenza symptoms on an individual’s ability to perform daily activities.

**Impact on others (5 items).** This domain covered the impact on others caused by influenza symptoms.

**Systemic symptoms:**
- Headache
- Feeling feverish
- Body aches
- Fatigue...
- Neck pain
- Interrupted sleep
- Loss of appetite
- Feeling feverish
- Fatigue...
- Body aches

**Upper respiratory symptoms:**
- Cough
- Sore throat
- Nasal congestion
- Concentrate on...
- Perform usual...
- Take care of...
- Sore throat
- Nasal congestion
- Concentrate on...
- Perform usual...
- Take care of...

**Lower respiratory symptoms:**
- Sputum
- Shortness of breath
- Wheezing
- Hoarseness
- Snake in throat

**Impact on daily activities:**
- Get out of bed
- Prepare meals...
- Perform usual...
- Leave the home...
- Concentrate on...
- Take care of...

**Impact on emotions:**
- Irritable
- Feel helpless
- Worried
- Frustrated

**Impact on others:**
- People worrying...
- Being a burden
- Needing to depend...
- People having to...
- Having to depend...

CFA fit statistics (two-factor solution): $\chi^2 = 62.78, 33 \text{ df, } P < 0.0013; \chi^2/\text{df} = 1.90; \text{CFI} = 0.983; \text{TLI} = 0.977; \text{RMSEA} = 0.057; \text{WRMR} = 0.712.$

Rasch fit statistics: respiratory symptoms: $\chi^2(8) = 5.82, P = 0.4, \text{PSI} = 0.51.$ Systemic symptoms: $\chi^2(14) = 17.57, P = 0.2, \text{PSI} = 0.85.$

Impact on daily activities: $\text{Bonferroni corrected } \alpha = 0.008.$

Impact on emotions: $\text{Bonferroni corrected } \alpha = 0.013.$

Impact on others: $\text{Bonferroni corrected } \alpha = 0.01.$

CFA fit statistics: $\chi^2 = 14.08, 8 \text{ df, } P = 0.0631; \chi^2/\text{df} = 1.85; \text{CFI} = 0.999; \text{TLI} = 0.998; \text{RMSEA} = 0.055; \text{WRMR} = 0.412.$

Rasch fit statistics: $\chi^2(12) = 22.28, P = 0.035, \text{PSI} = 0.92.$

CFA fit statistics: $\chi^2 = 1.562, 2 \text{ df, } P = 0.458; \chi^2/\text{df} = 0.78; \text{CFI} = 1.000; \text{TLI} = 1.002; \text{RMSEA} = 0.000; \text{WRMR} = 0.196.$

Rasch fit statistics: $\chi^2(8) = 21.53, P = 0.006, \text{PSI} = 0.82.$

Table 2 – Confirmatory factor analysis (CFA) and Rasch analysis of FluiQ™ domains.

<table>
<thead>
<tr>
<th>Domain*</th>
<th>CFA</th>
<th>Rasch analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor loading</td>
<td>Fit residual</td>
<td>$\chi^2$ (df = 2)</td>
</tr>
<tr>
<td>Symptoms: systemic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Headache</td>
<td>0.608</td>
<td>1.61</td>
</tr>
<tr>
<td>e. Feeling feverish</td>
<td>0.763</td>
<td>-0.14</td>
</tr>
<tr>
<td>f. Body aches</td>
<td>0.848</td>
<td>-2.13</td>
</tr>
<tr>
<td>g. Fatigue ...</td>
<td>0.782</td>
<td>-1.17</td>
</tr>
<tr>
<td>h. Neck pain</td>
<td>0.736</td>
<td>-0.07</td>
</tr>
<tr>
<td>i. Interrupted sleep</td>
<td>0.647</td>
<td>1.90</td>
</tr>
<tr>
<td>j. Loss of appetite</td>
<td>0.694</td>
<td>1.56</td>
</tr>
<tr>
<td>Symptoms: respiratory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Cough</td>
<td>0.537</td>
<td>0.21</td>
</tr>
<tr>
<td>b. Sore throat</td>
<td>0.477</td>
<td>1.01</td>
</tr>
<tr>
<td>d. Nasal congestion</td>
<td>0.617</td>
<td>0.97</td>
</tr>
</tbody>
</table>

CFA fit statistics (two-factor solution): $\chi^2 = 62.78, 33 \text{ df, } P < 0.0013; \chi^2/\text{df} = 1.90; \text{CFI} = 0.983; \text{TLI} = 0.977; \text{RMSEA} = 0.057; \text{WRMR} = 0.712.$

Rasch fit statistics: respiratory symptoms: $\chi^2(8) = 5.82, P = 0.4, \text{PSI} = 0.51.$ Systemic symptoms: $\chi^2(14) = 17.57, P = 0.2, \text{PSI} = 0.85.$

Impact on daily activities: $\text{Bonferroni corrected } \alpha = 0.008.$

Impact on emotions: $\text{Bonferroni corrected } \alpha = 0.013.$

Impact on others: $\text{Bonferroni corrected } \alpha = 0.01.$

CFI, confirmatory fit index; RMSEA, root mean square error of approximation; TLI, Tucker-Lewis index; WRMR, weighted root mean square residual.

* Items truncated; full items available from the author.
form routine or normal activities. This encompasses an individual’s ability to look after themselves, mobility, and concentration. Response options (0–3) included no difficulty, some difficulty, moderate difficulty, and great difficulty.

Impact on emotions (4 items). This domain covered the range of negative emotions related to having influenza and included irritability, helplessness, worry, and frustration. Response options (0–3) included not at all, somewhat, moderately, and extremely.

Impact on others (5 items). This domain covered the concerns, strain, or burden present in relationships between the individual and people around them (e.g., family/friends/neighbors) as a result of in-fluenza. Ratings (0–3) included not at all concerned, somewhat concerned, moderately concerned, and extremely concerned.

Cognitive testing
The content of the final set of questions was tested by patients participating in focus groups. The participants further judged the content to be relevant and important to people with influenza. Although almost all questions were easily understood, there was some concern that sore throat and neck pain may have had conceptual overlap, but ultimately these were regarded as distinct symptoms by all subjects. Sore throat was described as internal and related to swallowing, whereas neck pain was depicted as muscular in nature. Neck pain also was differentiated from aches and pains as a distinct symptom. About half of the subjects indicated that adding “body” to “aches and pains” would improve the specificity the item. The fatigue symptom contained the parenthetical expression “tiredness,” and all subjects thought that it helped clarify the question. Within the Impact scales, the intent of the item “have your meals” was interpreted slightly differently by some subjects. Some thought that it related to simply eating meals, whereas others thought that the question was about preparing meals. Additionally, some subjects noted that one might not cook or prepare foods but rely on prepackaged meals, especially when ill. This item was therefore expanded to include “get your own food.” There was some discussion about the “being a burden” item, and it was deemed to be more strongly worded than “needing to depend on people.” Those two items, however, were considered to be independent and different, and no revision was regarded as necessary. All response options and instructions were found to be clear and appropriate.

Validation study
A total of 311 participants were enrolled across 25 sites in the United States. The median age was 40 years (range, 18–84 years), and 64% were female. The majority had completed high school (~60%), were working full time (~67%), and were white (72% white 13% Hispanic, 11% black, and 4% Asian/other) (Table 3). Demographic factors were evenly distributed across those with ILI versus those with confirmed influenza. Influenza status was known for 263 (85%) of the sample; 57% had ILI and 28% had confirmed influenza. True influenza status could not be confirmed for 15% of patients with ILI due to missing serology and/or culture results. Of the 311 enrolled, 18 (5.8%) did not complete any of the questionnaires and were removed from the analyses, leaving 293 cases. At day 1, missing data points for each items ranged from 11 (3.8%) to 14 (4.8%) across items. At day 14, the proportion of cases with missing data ranged from 9.6% to 10.9% across items.

At day 1, the only domain with substantial floor effects was impact on others where 28.4% of respondents were in the low range. There were minimal ceiling effects for all domains, indicating that most respondents did not score “severe” across all items. Systemic symptoms were the most intense for respondents, where 12.4% reported scores in the upper decile (Table 4).

| Table 3 – Validation study: demographic background of people with influenza-like illness and confirmed influenza. |
| Age, y, mean (SD) | 43.3 (15.8) | 41.1 (12.3) | 42.1 (14.9) |
| Sex, female, no. (%) | 119 (67.2) | 52 (60.5) | 199 (64.0) |
| Race, no. (%) | | | |
| White | 126 (71.2) | 67 (77.9) | 193 (73.4) |
| Hispanic American | 25 (14.1) | 1 (1.2) | 26 (9.9) |
| Black | 18 (10.2) | 8 (9.3) | 26 (9.9) |
| Asian | 6 (3.4) | 7 (8.1) | 13 (4.9) |
| Indian (subcontinent) | 1 (0.6) | 2 (2.3) | 3 (1.1) |
| Native American | 1 (0.6) | 1 (1.2) | 2 (0.8) |
| Income, no. (%) | | | |
| <$15,000 | 15 (8.5) | 5 (5.9) | 20 (7.6) |
| $15,000–$24,999 | 22 (12.4) | 5 (5.8) | 27 (10.3) |
| $25,000–$34,999 | 22 (12.4) | 15 (12.8) | 37 (14.1) |
| $35,000–$49,999 | 38 (21.5) | 18 (17.4) | 56 (21.3) |
| $50,000–$74,999 | 32 (18.1) | 12 (20.9) | 44 (16.7) |
| $75,000–$99,999 | 18 (10.2) | 11 (14.0) | 29 (11.0) |
| ≥$100,000 | 10 (4.0) | 8 (9.3) | 18 (6.8) |
| Not available | 20 (11.3) | 11 (12.8) | 31 (11.8) |

CFAs
Symptoms domain. CFA of the 10 items of the symptoms domain resulted in a modest fit. Three items were only weakly loaded by the single latent variable (cough, sore throat, and nasal congestion), and correlated errors among these items suggested a subfactor (data not shown). A two-factor solution (respiratory symptoms and systemic symptoms) resulted in a very good fit (Table 2). The correlation between the factors was moderate (r = 0.58). All items had significant loadings from their respective latent variables, and the lowest loading item was sore throat (r = 0.48). There was one modest correlated residual between headache and neck pain (r = 0.24), a logical connection given the close physical proximity of these sensations.

Impact on daily activities domain. CFA of the impact on activities domain resulted in an excellent fit. All items were loaded strongly by the latent variable (r = 0.81). There was one notable correlated residual between “perform usual activities” and “concentrate on tasks,” r = 0.36, again a logical connection suggesting that loss of concentration may affect task completion.

Impact on emotions domain. CFA of the impact on emotions domain also resulted in an excellent fit. All items loaded satisfactorily (r = 0.59), and there were no statistically significant correlated residuals.

Impact on others domain. CFA of the impact on others domain yielded a problematic fit. Although all items had satisfactorily high loadings (r = 0.78), there was one large correlated residual between “needing to depend on people” and “people having to do extra things for you,” r = 0.63, suggesting a significant level of redundancy between these two items.

SEM
A measurement model with five factors was specified as follows: 1) each construct was constrained to load only on the items that were hypothesized to be associated with it (i.e., no cross-loadings were allowed); 2) there were no correlations among the item residuals; and 3) correlations between factors were freely estimated.
This model fitted the data reasonably well (CFA fit statistics: $\chi^2 = 486.06, 265 \, df, P < 0.0001; \chi^2/df = 1.83; CFI = 0.984; TLI = 0.982; RMSEA = 0.057; WRMR = 0.928$). The fit improved with the addition of correlations between the residuals of headache and neck pain, perform normal activities and concentrate on tasks, and needing to depend on people and having to do extra things for you ($\chi^2 = 448.06, 262 \, df, P < 0.0001; \chi^2/df = 21.71; CFI = 0.986; TLI = 0.985; RMSEA = 0.052; WRMR = 0.871$) but the fit could not be appreciably improved further. Correlations among the factors ranged from $r = 0.46$ between respiratory symptoms and impact on emotions to $r = 0.86$ between impact on emotions and impact on others.

A finding of robust factorial separation of items into factors following theoretical specification, as shown previously, provides strong support, but is not sufficient for construct validity [12]. What is also required is functional separation in the nomological network of the constructs, i.e., differential prediction of important outcomes. To test this causal model based on the five factors with paths between them guided by the conceptual model was specified as follows: 1) one exogenous source of variation in reported respiratory and systemic symptoms was hypothesized: confirmed influenza versus ILI; 2) respiratory and systemic symptoms were subsequently hypothesized to be causally related to daily activities, impact on emotions, and impact on others; 3) no causal paths or correlations between indicators and factors or indicators and indicators with the exception of three residual correlations specified previously were allowed. After estimating the model with all causal paths specified previously, those that were not statistically significant ($P > 0.05$) were deleted, and the model was estimated again. The resulting statistics indicated a very good fit ($\chi^2 = 498.23, 288 \, df, P < 0.0001, \chi^2/df = 1.73, CFI = 0.985, TLI = 0.983, RMSEA = 0.053, WRMR = 0.915$). Statistically significant ($P < 0.05$) causal associations were found between: confirmed influenza versus ILI and systemic symptoms ($0.19$; those with confirmed influenza reported more severe systemic symptoms); systemic symptoms and impact on daily activities ($0.19$; those with confirmed influenza reported more severe systemic symptoms); systemic symptoms and impact on emotions ($0.71$; systemic symptoms and impact on others ($0.51$); and respiratory symptoms and impact on others ($0.18$). Notably, no causal associations between confirmed influenza and respiratory symptoms and respiratory symptoms and impact on others (daily activities or impact on emotions were observed. Additionally, any causal association between confirmed influenza and impact on daily activities, emotions, and others was mediated through influenza symptoms.

The error variances associated with the outcome factors showed that the model accounted for 66% of the variance in impact on daily activities, 51% in impact on emotions, and 40%
in impact on others. The amount of variance explained in these domains is quite substantial and indicates that the symptom domains are indeed pertinent to an individual’s health state and, in particular, strongly determine the activities, emotions, and relationships of affected individuals.

Flu status (confirmed vs. ILI) was only weakly and positively associated with systemic symptoms and not related to respiratory symptoms. The data indicate that influenza status explains only about 4% of the variance in systemic symptoms (and no variance in respiratory symptoms). The final causal model of influenza impact is shown in Figure 3.

Further analysis indicated that the same measurement model fitted the data from both confirmed influenza and ILI groups. This indicates that data collected from these samples could be pooled and that the questionnaire is appropriate for both ILI and confirmed influenza. This analysis was a simple two-group measurement model split by influenza status with all factor loadings, factor correlations, and correlated residuals constrained to be equal across groups. Fit statistics for this two-group model (N = 258) were \( \chi^2 = 761.91, 608 \text{ df, } P < 0.0001; \chi^2/df = 1.25; \text{CFI} = 0.989; \text{TLI} = 0.992; \text{RMSEA} = 0.044; \text{WRMR} = 1.289. \)

**Rasch analysis**

**Symptom domain.** Using Rasch analysis, the 10-item symptom domain had satisfactory threshold ordering, but the overall fit between the domain and the Rasch model was poor, \( \chi^2(20) = 95.78, P < 0.001. \) Items headache and fatigue also showed evidence of DIF, with female subjects obtaining higher scores on these items than male subjects with comparable levels of symptom severity (\( P < 0.001 \)). Items cough, sore throat, and nasal congestion had residuals ranging between 2.52 (nasal congestion) and 4.15 (sore throat) and were also misfitting on the \( \chi^2 \) test, suggesting that these items were not measuring the same construct as the rest of the scale items. One additional item (body aches and pains) had a high negative residual (\( -2.53 \)), suggesting redundancy.

Consistent with the results of CFA, PCARC showed that the items cough, sore throat, and nasal congestion had high positive loadings on the first principal component, whereas the remaining items had low or negative loadings, thus indicating nonunidimensionality of the symptoms domain. The items were then split into the systemic and respiratory symptom domains. The systemic symp-
Impact on daily activities and impact on emotions domains. The impact on daily activities and impact on emotions domains performed very well across most fit indices (Table 2). The only exceptions were “prepare meals/get your own food” (fit residual −3.07) and “get out of bed” (fit residual 2.73) items of the impact on daily activities domain. Removing these items did not substantively improve fit; therefore, they were retained.

Impact on others domain. Although the overall Rasch model fit for the impact on others domain was weak (Table 2), two critical factors were very good; PSI was 0.90 and there were no disordered thresholds.

“Being a burden” and “people being annoyed with you” had misfitting residuals and significant χ² test result. “People worrying about you” had a fit residual of 3.12 and “people having to do extra things for you” misfitted on a direction (positive with GGHQ and negative with EQ5D thermometer) and in the moderate range (r > 0.25). Overall, weak associations (r < 0.25) were observed between the FluiiQ™ and the Physician GSQ, especially at day 1. In general, for all measures used, the correlations were higher on days 7 and 14 compared to day 1.

Known groups validity
Known groups validity was supported because those with confirmed influenza had a higher score than those with ILI across all domains, as expected (Table 4). The 95% CI for difference, however, included 0 for impact on emotions and impact on others. The largest mean difference between confirmed influenza and ILI was for impact on daily activities (mean 0.38; 95% CI 0.12–0.60).

Responsiveness to change
Mean (SD) scores at day 1 across domains ranged from 1.85 (0.63) for respiratory symptoms to 1.00 (0.85) for impact on others. At day 14, the mean (SD) score ranged from 0.42 (0.57) for respiratory symptoms to 0.28 (0.25) for both impact on daily activities and impact on others. Large and rapid changes in mean scores across all domains were observed across the 14 days (Fig. 4). The largest mean individual change was for symptoms (all) and respiratory symptoms, both with changes >1.5 units over 14 days. The symptoms (all) domain had the largest observed SRM (2.28) and effect size (2.66). The SRM was >1.0 for all domains, indicating the domains were highly responsive to change.

Discussion
These analyses provided evidence of the validity of the FluiiQ™ as measures of symptoms severity and impact of influenza. The FluiiQ™ comprises five new scales that are now available for researchers to assess influenza symptom severity and impact in epidemiological research as well as to explore efficacy and effectiveness of influenza prophylactic and/or therapeutic agents in clinical trials. The procedures used to develop the FluiiQ™ followed the FDA guidance for PRO development and validation including patient and clinician input, explicit conceptual model development, cognitive testing, validation in the target population (confirmed influenza patients) with clear evidence of construct validity, reliability, and responsiveness [4].

An important constraint considered during the development of the questionnaire was that it should be short and require minimal cognitive capacity to understand and answer the questions reliably given that people with influenza have substantial malaise. Consequently, each question and response option generated were conceptually straightforward so that affected individuals are able to quickly find a “place” for themselves across response options.

### Table 5 – Spearman’s correlation coefficients between FluiiQ™ domains and other measures.

<table>
<thead>
<tr>
<th></th>
<th>Symptoms</th>
<th>Impact on daily activities</th>
<th>Impact on emotions</th>
<th>Impact on others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Systemic</td>
<td>Respiratory</td>
<td></td>
</tr>
<tr>
<td><strong>Patient global severity question</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td>0.44</td>
<td>0.41</td>
<td>0.35</td>
<td>0.38</td>
</tr>
<tr>
<td>Day 7</td>
<td>0.60</td>
<td>0.53</td>
<td>0.58</td>
<td>0.48</td>
</tr>
<tr>
<td>Day 14</td>
<td>0.68</td>
<td>0.56</td>
<td>0.71</td>
<td>0.46</td>
</tr>
<tr>
<td><strong>Physician global severity question</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td>0.21</td>
<td>0.22</td>
<td>0.07</td>
<td>0.19</td>
</tr>
<tr>
<td>Day 14</td>
<td>0.40</td>
<td>0.32</td>
<td>0.44</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>EQ5D thermometer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td>-0.27</td>
<td>-0.27</td>
<td>-0.15</td>
<td>-0.29</td>
</tr>
<tr>
<td>Day 14</td>
<td>-0.41</td>
<td>-0.39</td>
<td>-0.40</td>
<td>-0.35</td>
</tr>
<tr>
<td><strong>Generic global health question</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td>0.23</td>
<td>0.24</td>
<td>0.08</td>
<td>0.24</td>
</tr>
<tr>
<td>Day 7</td>
<td>0.48</td>
<td>0.45</td>
<td>0.40</td>
<td>0.42</td>
</tr>
<tr>
<td>Day 14</td>
<td>0.43</td>
<td>0.40</td>
<td>0.41</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Notes:**
- All correlation coefficients are based on the Spearman’s rank correlation coefficient (r).
- The table includes correlations between the FluiiQ™ domains and other measures: Patient GSQ, Physician GSQ, GGHQ, EQ5D thermometer, and impact on daily activities and impact on others.
- The table shows the correlation coefficients for each domain at different time points (Day 1, Day 7, and Day 14).
- The table also includes the impact on daily activities and impact on others domains.
- The table includes correlation coefficients for all measures used, the correlations were higher on days 7 and 14 compared to day 1.
The fit indices in the SEM and Rasch analyses indicate that this was achieved for almost every item. The high item endorsement, including minimal missing data, high relevance (bothersomeness) scores for the symptoms questions, and absence of disordered thresholds indicate that the questions are pertinent and people with influenza are able to respond to the questionnaires in a congruent manner. The exception was some redundancy between the “needing to depend on people” and “people having to do extra things for you” items in the impact on others domain. There was also some evidence of weakness in the “being a burden” item. These items, however, were retained to ensure breadth of the impact on others domain; additionally, removal of any item did not significantly improve the model fit. Revision and further testing of items may improve future versions of the questionnaire.

The results largely confirm the symptom profile presented in previous reports [23–25]. The interviews with clinicians and consultation with patients with confirmed influenza and ILI enabled important refinements to be made on the measures and content validity to be established. Nausea and lack of concentration were removed because these were regarded as not being specific to influenza. Neck pain was added and was found to be a highly congruent item in the systemic symptoms scale. The careful consultations that we undertook with patients enabled us to identify and validate novel measures of the impact of influenza, and our SEM analysis provides evidence that they are causally related to the symptoms.

Importantly, the symptoms and the impact scales performed well across several indices of validity (both within the FluiiQ™ items and between the FluiiQ™ and other measures included in the study) and reliability. Each is linked to different rates of change over time, which probably reflects the pathophysiological development and resolution of the disease. Impact on daily activities appears to resolve most quickly, whereas respiratory symptoms begin with the highest burden and at day 14 remain the highest and may reflect a lingering nagging cough and gradual resolution of nasal congestion. In contrast, systemic symptoms and other impacts have mostly resolved by day 14. Although the respiratory symptom domain has only 3 items and low reliability/PSI, its ability to show change over time appears reasonable. Despite the small number of somewhat disparate items within the respiratory symptoms, this domain appears to work well as a scale, and this is also supported by the congruent SEM analyses.

Overall, the correlation between the FluiiQ™ and the other measures included in the validation study were moderate and therefore demonstrating reasonable convergent validity with symptoms and measures of well-being. Weaker correlations were observed between the patient reported FluiiQ™ and the Physician GSQ, which is a common finding [26, 27]. Across all measures, a lower correlation was observed at day 1 compared to later assessments, which may relate to patients considering their general health to be good, regardless of their influenza episode.

As shown in Figure 3, the absence of significant paths from respiratory symptoms to daily activity and impact on emotions demonstrates that the majority of the effect on these two outcomes is from the systemic symptoms. The sharp discrimination between the impact of respiratory and systemic symptoms on the self-reported impact outcomes also suggests that there is an absence of a “general response factor” influencing the results (e.g., acquiescence, social desirability, or similar biases). In other words, patients appear to consider each question carefully and respond to its content in a way that is independent from other items in the domain. Our consultations with patients during the development work also supported this because we found the respiratory symptoms to be ranked somewhat lower on the bothersomeness scale. The congruence between SEM and the more direct questions about bothersomeness provide further support for the content and construct validity of the domains.

The robustness of a five-factor measurement model across the confirmed influenza/ILI groups provides support that data collected from these two samples can be pooled for analyses. This is also congruent with physician’s reports that it is difficult to distinguish true influenza from ILI in the clinical setting. Numerous studies in the literature that have found that when in-
fluenza is circulating in a community, ILI, particularly cough and fever, is predictive of true influenza infection [25,28,29]. Because our questionnaire development and validation protocols required site confirmation of circulating laboratory-confirmed influenza in their geographic region before recruitment and enrollment, our findings are not surprising.

An important finding from the perspective of the validity of the domains is the very strong paths between systemic symptoms and daily activities and impact on emotions compared to the absence of significant paths from respiratory symptoms to these latent variables. This clearly suggests that the distinction between respiratory symptoms and systemic symptoms has causal validity as well as factorial validity (convergent and discriminant validity), thus strengthening the construct validity of the distinction for evaluation purposes. It is plausible that viral infection in the upper respiratory areas (nose, throat) and the associated secretions result in the sore throat and nasal congestion. The secretions then travel down the respiratory tract, resulting in a dry reactive cough. The remaining (systemic) symptoms are a result of the immune system response to the viral load. Interferon and other mediators are released into the bloodstream and result in the widespread symptoms related to the systemic reactions to infectious illness. Given that there is a clear case for the distinction in the measurement of respiratory symptoms and systemic symptoms and a plausible biological causal path, a summed score of the 10 symptom items is not recommended. This is supported by the one-factor model, which does not fit and the correlation between the two symptom factors is not sufficiently high (r = 0.58) to justify adding up the two subdomains. Although investigators may choose a priori to combine the 10 symptoms as an outcome, our analyses suggest that separating the domains will provide more meaningful data.

The initial limitation of the FluIiQ™ is the respiratory domain because of its low reliability. This could possibly be a target for further development, although it seems unlikely that many more relevant and independent indicators of this construct can be generated. Measurement of test-retest reliability is a challenge in this area because symptoms of influenza change rapidly and it is not possible to select a test-retest period when a respondent has both stable symptoms and cannot easily recall their previous answers. We therefore used evidence about a scale’s ability to detect change as a way to compensate for lack of information about test-retest reliability. If a scale is highly unreliable, it would have a poor ability to detect change over time. Figure 4 and Table 4 provide evidence that the respiratory and systemic symptoms scales perform satisfactorily (detect change in a similar manner) as the magnitude of change and SD of change is similar. Although concept mapping and other qualitative and quantitative processes undertaken help to ensure that the impact of influenza is well covered, another potential limitation is our a priori exclusion of work and financial impacts. Although patient diaries cover these areas in detail, there may be influenza-specific impacts that we have missed. Future work should include a focus on these challenging areas. Finally, another arguable weakness of this study is the relatively small sample size for SEM. This is always a challenge when gathering data specifically for questionnaire development where there are few cases (rare disease) or the cost of gathering a large sample is prohibitive. A recent simulation study showed that WLSMV estimation of CFA models in Mplus performed equally well across a range of sample sizes between 250 and 1000. It was concluded [30] that “Overall, these results demonstrate that WLSMV estimation does not need the large sample sizes needed for WLS estimation.” Although this study was not extended to causal models, it did involve complex CFA models with moderate factor loadings, thus affirming support for the clear superiority of WLSMV estimation for CFA with categorical variables and relatively modest sample sizes.

An area for future work is defining the predictive validity of each FluIiQ™ domain. As mentioned earlier, it would be valuable to explore which specific FluIiQ™ symptoms or combination of symptoms predict true influenza or specific strains of influenza. A tool with high predictive validity could be a useful element in the surveillance of local outbreaks and even pandemics. Work needs to be done to explore whether the FluIiQ™ can be used to facilitate more accurate diagnosis of ILI in the busy clinical setting. It is likely that, with an increasing score of appropriately weighted items, there will be an increasing probability of true influenza being present in an individual. For example, our ILI definition was taken from clinical efficacy studies evaluating the Live Intranasal Influenza Vaccine because the ILI definition was much broader than the Centers for Disease Control and Prevention definition of fever (≥37.8°C, ≥100°F) plus cough, sore throat, or both in the absence of another known cause of illness, in part because we wanted to capture as many potential cases as possible. Further work on the FluIiQ™ could systematize the collection of the symptom profile to improve case definition accuracy. Furthermore, this may also provide more accurate data on prodromal aspects of influenza that lead to incapacitation and hospitalization. The FluIiQ™ is already being used in a variety of large multinational clinical trials and epidemiological studies, and secondary analysis from these studies may be used to understand the predictive validity and cross-cultural validity of the questionnaire. A copy of the questionnaire may be obtained from the authors.

The use of recommended patient and clinician engagement processes in the preparation of items and constructs, robust classical and modern psychometrics, and systematic evaluation of the items and domains in the target population has generated rigorous data to support the validity of the FluIiQ™ as a measure of symptom severity and the impact of influenza.

Source of financial support: Professor Osborne was funded in part by a National Health and Medical Research Council Population Health Career Development Award. This study was funded by Merck Sharpe & Dohme Corp.

REFERENCES

[8] Rosas SR, Camphausen LC. The use of concept mapping for scale
development and validation in evaluation. Eval Program Plann 2007;
30:125–35.
Questionnaire (heiQ): an outcomes and evaluation measure for patient
education and self-management interventions for people with chronic
[10] Trochim WM, Linton R. Conceptualization for planning and
approach to the development of complex interventions: the
Australian WorkHealth Program – Arthritis as a case study. Soc Sci
York: Guilford Press, 2006.
analysis: conventional criteria versus new alternatives. Struct Equat
[16] Yu C-Y. Evaluating cutoff criteria of model fit indices for latent
Available at: www.statmodel.com/download/Yudissertation.pdf.
analysis program employing Rasch unidimensional measurement
[18] Rasch G. Probabilistic models for some intelligence and attainment
2004.
[22] Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd
of life in critically ill patients: how to score and what is the clinical
[27] Sprangers MA, Aaronson NK. The role of health care providers and
significant others in evaluating the quality of life of patients with
epidemics with use of a clinical case definition. Clin Infect Dis 2000;31:
1166–9.
[29] Beauducel A, Herzberg FY. On the performance of maximum
likelihood versus means and variance adjusted weighted least squares