OBJECTIVE

To develop a sensitive, reliable, responsive, and easy-to-use instrument for assessing the severity and functional impact of the common cold.

STUDY DESIGN

We created an illness-specific health-related quality-of-life outcomes instrument using previous scales, expert opinion, and common knowledge. This original questionnaire was used in a 1999 randomized trial of echinacea for the common cold. In 2000 we employed cognitive interview and focus group qualitative methods to further develop the instrument. Semistructured interviews used open-ended questions to elicit symptoms, terminology, and perceived functional impact. Responses were used to improve the instrument.

POPULATION

The randomized trial watched 142 University of Wisconsin students for a total of 953 days of illness. The subsequent qualitative instrument development project recruited 74 adults with self-diagnosed colds for 56 in-person interviews and 3 focus groups.

OUTCOMES MEASURED

We measured specific symptoms, symptom clusters (dimensions), functional impact, and global severity.

RESULTS

The original questionnaire included 20 questions: a global severity indicator, 15 symptom-severity items using 9-point severity scales, and 4 yes/no functional assessments. Data from the trial provided evidence of 4 underlying dimensions: nasal, throat, cough, and fever and aches, with reliability coefficients of 0.663, 0.668, 0.794, and 0.753, respectively. Qualitative assessments from the interviews and focus groups led us to expand from 15 to 32 symptom-specific items and from 4 to 10 functional impairment items. The original 9-point severity scale was revised to 7 points. Two global severity questions bring the item count to 44. The instrument fits comfortably on the front and back of a single sheet of paper and takes 5 to 10 minutes to complete.

CONCLUSIONS

The Wisconsin Upper Respiratory Symptom Survey (WURSS) is now ready for formal validity testing or practical use in common cold research.

KEY WORDS

Common cold; quality of life; patient-oriented outcomes [non-MeSH]; qualitative methods [non-MeSH]; questionnaires; survey instrument [non-MeSH]; symptom measurement [non-MeSH]; upper respiratory infections. (J Fam Pract 2002; 51:265)

The common cold, usually caused by viral infection of the upper respiratory tract, is a very prevalent illness. On average, US adults suffer from 1 to 4 episodes per year. This high incidence, along with significant symptomatic and functional impairment, combine to make this syndrome an important health problem. Hundreds of trials have attempted to demonstrate effective treatments. Unfortunately, few efforts have been made to develop and validate instruments to measure the symptomatic and functional impact of the common cold.

The term “upper respiratory infection” (URI) is a nosologic category constructed by physicians and other health professionals to reflect an upper airway, mucus-producing, inflammatory reaction to infection, usually viral. It is a disease category. The terms rhinitis, rhinosinusitis, pharyngitis, and bronchitis are often used to indicate the anatomic area most affected. The term “common cold” is an illness term constructed and used by the general populace. This distinction between professional (disease) and popular (illness) conceptions provides the reasoning for participant-based, patient-oriented qualitative development of measurement tools. While many medical professionals may choose to measure URI disease by physical examination, viral culture, or laboratory analysis of
Using various modifications of the Jackson scale, researchers of the cold have characterized the frequency and severity of the 8 symptoms noted above in both natural colds and experimentally induced rhinovirus infections. Variability in symptom expression remains a hallmark of URI. Although specific pathogens are associated with the severity and distribution of symptoms at the population level, symptoms are poor predictors of etiology at the individual level. Infection itself is an imperfect predictor of symptom expression, as asymptomatic infections occur frequently, and as URI-like symptoms occur in people in whom it is not possible to demonstrate infections.25 Even among people with documented experimental infections of single strains of virus, variance outweighs central tendency in all symptom measurements.25,26

The search for objective disease measures with which to compare symptom scores has also progressed. To date, the following measures have been evaluated: detection of virus with culture or polymerase chain reaction,26,29 cytokine measurement,30-32 serologic markers,33 physical examination,34,35 radiologic imaging,36,37 rhinomanometry,38,39 mucus weight,40 mucus velocity, and number of tissues used.30 None have been shown to be superior to self-reported symptoms in terms of precision, reliability, or responsiveness or in their ability to predict functional impairment or subsequent illness. Perhaps more important, none have been shown to reflect the values of the people who experience colds. Although a number of quality-of-life instruments have been developed to assess allergic rhinitis,31-36 we have been unable to locate any specifically developed to assess URI.

We therefore decided to develop the Wisconsin Upper Respiratory Symptom Survey (WURSS) to provide a standardized measure for evaluating the negative consequences of the common cold. We were particularly interested in developing a health-related quality of life instrument that would represent the symptomatic and functional dimensions that are important to cold-sufferers.27-30 The instrument should be able to discriminate accurately between active intervention and placebo effects in randomized therapeutic trials and should balance brevity and ease of use with optimal precision, reliability, and responsiveness.41-45 It should be based on self-diagnosis and self-assessment because neither accepted criteria nor adequate tests are available to diagnose “upper respiratory infection” or “acute infectious rhinosinusitis”(“with or without “pharyngitis”) and because the vast majority of cold treatments will be taken without professional input after self-diagnosis.
METHODS

Phase 1: Initial Development During a Randomized Trial

The development of this study began in 1998 during the design of a randomized controlled trial (RCT) of echinacea as a cold treatment. We created our first instrument by showing successive drafts to friends and colleagues (mostly family physicians), stopping once we were satisfied that the questionnaire had reasonable face validity. This initial instrument rated global severity of illness (“How sick do you feel today?”) and 15 individual symptoms on a 9-point Likert-type scale. The 15 symptom-measuring items were complemented by 4 dichotomous (yes/no) functional outcome questions, adapted with permission from the validated Medical Outcomes Study 36-item Short-Form Health Survey (SF-36).

This initial instrument was used in the spring of 1999 in the echinacea RCT. This experience provided a good initial test of our instrument, as the participants were recruited within 36 hours of their first symptom and monitored each day until they had answered “No” to the question, “Do you think that you are still sick today?” for 2 days in a row. Each participant was asked to fill out the questionnaire both on paper and on a computerized data-collecting facsimile (available at http://www.fammed.wisc.edu/samplecold).

Phase 2: Further Instrument Development Using Qualitative Methods

After the echinacea RCT was completed, our primary concern was that we might be overlooking or underrepresenting important illness domains. We also suspected that wording, question order, response range, and other formatting concerns could be improved. To achieve these goals, we used qualitative instrument-development methodologies, involving the people we wanted to measure—cold-sufferers—in the development process.

After obtaining approval from the University of Wisconsin Medical School Human Subjects Committee, we began interviewing Madison-area adults who responded to community advertising asking for volunteers with colds. Inclusion criteria required answering “Yes” to the question, “Do you believe that you have a cold?” For an interview to be arranged, at least one cold symptom had to be present, and the research assistant had to be convinced that the caller was indeed suffering from a common cold. Prospective participants with itchy eyes, sneezing, or a history of allergy were excluded if either the participant or the interviewer felt that any current symptoms might have been caused by allergy. Interviews were held in a location of mutual convenience and with the aid of an interview guide developed by our research team. Interviewers were carefully trained in the research protocol and used interview guides for both the initial telephone screen and the in-person interviews. Interviewers included both clinicians and nonclinicians.

The semistructured interview guide used open-ended questions designed to elicit the participants’ own terminology for describing their colds (Table 1). We aimed for an understanding of how the experience of the cold influenced the lives of the participants. Participants were first asked to list all their
We used focus group methods in the final month of the study as an additional window into participants’ experiences. The focus groups used the same inclusion criteria as the long interviews and followed the same general format, first using open-ended questions to elicit symptoms and their impact, then administering the questionnaire and discussing item inclusion and formatting. However, we encouraged discussion rather than self-assessment, as the focus group methodology derives its strength from the interactive nature of conversation. For instance, a statement made by one participant would spark interest or recall in another, thereby generating a richer, fuller, and more representative description of symptoms and functional impact.

Individual interviews were held by 1 of 5 trained interviewers (B.B., L.L., R.M., E.S., J.S.). All 3 focus groups were run by the lead author, with at least 1 other research team member assisting. Interviews and focus groups were arranged as soon as possible after the initial telephone contact so that participants would still have cold symptoms while being interviewed. All interviews and focus groups were discussed in biweekly group meetings. Decisions on item inclusion, wording, and questionnaire format were made by research group consensus. Several versions of the questionnaire were brought back to cold-sufferers for further cognitive testing. The diversity of interviewers and respondents provided protection against personal bias in ascertaining and interpreting symptoms and impairments.

RESULTS

Phase 1

Of the 148 college students enrolled, 142 followed protocol and were included in the analysis. Of the 853 person-days documented, 546 (64%) were covered by both data systems; 287 (33.6%) came from paper surveys only; and 18 (2.1%) were filled out via computer only. Because only 2 (0.2%) questionnaires were missing any data, our data capture rate was 99.8%. Comparing data from the computerized and paper data sources provided evidence of consistency. Of the 546 days in which both paper and computer instruments provided data, 512 yielded identical responses (94% concordant) to the global severity of illness question. Of the 34 (6%) discrepancies, 29 were off by 1 point on the 9-point Likert-type scale and 5 discrepancies were off by 2 points. Comparing computer and paper responses with the 15 specific symptom questions also yielded high levels of concordance. Of 8190 item responses, 7777 (95%) were concordant, while 413 (5%) were classified as data discrepancies. Of these, 293 were off by

<table>
<thead>
<tr>
<th>Method of Data Collection</th>
<th>Number</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual interviews</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Focus groups</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>74*</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>49</td>
</tr>
<tr>
<td>Men</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native American</td>
<td>2</td>
</tr>
<tr>
<td>Black</td>
<td>12</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
</tr>
<tr>
<td>White</td>
<td>57</td>
</tr>
<tr>
<td>No response</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annual Income</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; $10,000</td>
<td>28</td>
</tr>
<tr>
<td>$10,000–19,999</td>
<td>17</td>
</tr>
<tr>
<td>$20,000–29,999</td>
<td>13</td>
</tr>
<tr>
<td>$30,000–49,999</td>
<td>9</td>
</tr>
<tr>
<td>$50,000–75,000</td>
<td>3</td>
</tr>
<tr>
<td>≤ $75,000</td>
<td>2</td>
</tr>
<tr>
<td>No response</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some high school</td>
<td>13</td>
</tr>
<tr>
<td>High school or equivalent</td>
<td>11</td>
</tr>
<tr>
<td>Some college</td>
<td>11</td>
</tr>
<tr>
<td>Associate or technical</td>
<td>6</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>21</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>6</td>
</tr>
<tr>
<td>Professional degree</td>
<td>4</td>
</tr>
<tr>
<td>No response</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tobacco Use</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>26</td>
</tr>
<tr>
<td>Past</td>
<td>19</td>
</tr>
<tr>
<td>Never smoker</td>
<td>28</td>
</tr>
<tr>
<td>No response</td>
<td>1</td>
</tr>
</tbody>
</table>

*Two participants were used in both data collection methods. NOTE: Age range was 19 to 71 years, mean = 35.9 years (standard deviation, 11.9).
and in the newspaper. Of these 108 callers, 27 were eligible but declined to participate; 7 did not meet inclusion criteria (were younger than 18 years of age, had current allergy symptoms, or did not have cold symptoms); and 74 met study criteria and elected to participate (Table 3). Those declining to participate usually did so because of inconvenience in arranging an immediate interview or because compensation ($10 for interview, $15 for focus group) was insufficient. Participants were met in person for semistructured individual interviews (n = 56) or focus groups (3 groups, 20 individuals total). Two people were interviewed both individually and in focus group.

Based on the information gained during interviews, the instrument-in-development underwent 6 revisions during 2000. Each modification was tested with additional interviews. A final version was created in December 2000. A few items from the initial instrument used in the echinacea trial were modified in response to participants’ descriptions and insights. Several other items were added to reflect symptoms and functional impairments described by participants in response to our open-ended questions. All items used wording provided by participants or tested during subsequent interviews and focus groups.

All symptoms spontaneously reported by at least 3 participants in either individual interview or focus group were included in the final version (Table 4). Figure 1 provides a frequency distribution of the symptoms described during the individual interviews.

On the basis of our participants’ comments, the distribution of severity data from the echinacea trial, and recommendations from published studies, we decided to decrease the response range from a 9-point to a 7-point Likert-type severity scale. The resulting severity range was marked at 1 (very mild), 3 (mild), 5 (moderate), and 7 (severe), following the majority opinion of our respondents. Unmarked (even-numbered) options were included, as most of the respondents felt they should have “in-between” choices. The functional outcome questions adapted from the SF-36 were replaced with participant-generated items, which were then scaled in a 7-point format similar to that used for the symptoms. A final question comparing today’s global severity with yesterday’s was added to provide a comparative measure of change over time (respondiveness). Figure 2 displays the final format of WURSS items.

The qualitative data provided by our informants improved our understanding of the symptomatic and functional impact of the common cold and assisted the development of the WURSS questionnaire. Although it is clear that people experience colds in
different ways, several common threads emerged. For instance, we found that our original instrument (like the Jackson criteria) had overrated individual symptoms and had underrated functional impact, interference with social relationships, and general malaise. Informants often told us that it was not necessarily the individual symptoms that bothered them, but the general feelings, described as “sick feeling,” “loss of energy,” “run down,” “tired,” “fatigue,” “malaise,” “lousy,” “lazy,” “spacey,” “blah,” “yucky,” “foggy,” “lightheaded,” “fuzzy brain,” “cloudy,” “disoriented,” “uncomfortable,” “distracted,” and “miserable.” Our informants also told us that they were bothered by the way their cold interfered with day-to-day activities and relationships. Colds affected physical activities such as breathing or walking, performance at work or in the home, and interactions with friends, family, and coworkers. Terms describing the most frequent and bothersome effects were incorporated into the final WURSS instrument.

**DISCUSSION**

Researchers of URIs and the common cold need a well-developed, standardized, validated outcomes instrument that reflects the experience and values of cold sufferers. While the Jackson scale and various modifications have been widely used, few data support the validity of these scales. Although correlations with external measures, such as physical examinations, mucus weight, and the ability to culture virus have been reported, the symptomatic and functional impact of colds has largely been neglected. Perhaps more important, the symptomatic measures used to date were apparently developed without significant input from the people whose illnesses were measured. Questionnaire development and cognitive testing methods have not been described, nor have adequate tests of psychometric properties been reported. Although the Jackson scale may demonstrate marginal face validity in terms of symptoms, it does not do so in terms of functional impact.

This article describes the first steps taken in the development and validation of a new illness-specific quality-of-life instrument for measuring the common cold. The WURSS instrument is more comprehensive than existing alternatives and better reflects cold-sufferers’ experiences and values. Therefore, it provides greater face validity. The length (44 items) reflects a compromise between ease of use and comprehensiveness. It is possible that a subset of the items will prove nearly as effective and that a short-form WURSS will eventually be available. Item reduction will need to be guided by both internal (factor analysis) and external (frequency and perceived value) considerations. The standardized 7-point severity scale used throughout the WURSS makes the instrument very easy to use. It also provides a severity range that our informants and previous researchers agree is optimal. The WURSS allows a cold-sufferer to swiftly and accurately assess his or her common cold. We hope that WURSS will prove worthy in terms of standard psychometric properties such as precision, reliability, and responsiveness. A large prospective study will be necessary for those assessments.
Limitations

The work described here has a number of limitations. The WURSS was developed in Madison, Wis., largely among people with self-diagnosed colds during the period from July to December 2000. The RCT occurred over several months in the spring of 1999 and was limited to college students. Our results are therefore limited by both population and etiologic agent, which in turn may influence the symptom and severity spectrums assessed. Although we aimed for and achieved a moderate degree of socioeconomic diversity (Table 3), our participants' responses may not be representative of the larger universe of cold-sufferers. The symptom distribution in Figure 1, for example, is unlikely to represent global cold symptom frequency accurately. Previous research with both natural and induced colds suggests that nasal symptoms and sore throat are usually more frequent than cough.\(^2\)\(^3\)\(^6\)\(^7\)\(^8\)\(^9\)\(^10\) The comprehensiveness of the instrument is more important for instrument development than are the specific item frequencies. Here, we feel that we succeeded in representing a sufficient range of items.

Another important limitation is the inherent variability and subjectivity of information generated from qualitative research. A similar instrument development effort carried out by different researchers would inevitably yield a somewhat different questionnaire. Eliciting and formatting terminology that reflects symptomatic and functional impact presents a number of challenges. Future research could employ a quantitative importance scale for participants to use in assessing the value of symptoms and functional impacts. Such value scales could be used alongside factor analysis models of item and dimension frequency and severity. These could in turn be compared with external criteria such as physician assessment, tissue counts, and nasal mucus weights, measurements of inflammatory cytokines, and quantitative viral cultures. Because no gold standard exists, single-criterion validity assessment will not be sufficient. Instead, the concept of construct validity will need to be invoked for future attempts at validation. Construct validity has been defined as “validity assessed by comparing the results of several contrasting tests of validity (including concurrent, convergent, and divergent validation studies) with predictions from a theoretical model.”\(^64\) Our work so far has only begun to scratch the surface of such rigorous validity assessment.

CONCLUSIONS

We have developed an instrument that measures patient-oriented outcomes identified as important by people with self-diagnosed common colds. We expect that the WURSS will do well with physician-diagnosed “bronchitis,” “sinusitis,” or “pharyngitis,” but as yet have no data with which to evaluate that supposition. We hope that the development of the WURSS stimulates other researchers to undergo similar efforts at aimed at patient-oriented outcome measurement and that the efforts can be compared. We have made the WURSS available for general use by placing a printable facsimile online at http://www.fammed.wisc.edu/wurss/. University-based health care researchers and other nonprofit entities may use the WURSS freely, but we do ask to be notified of such use. For-profit entities should contact us before using this copyrighted instrument.

The next step will be for WURSS to undergo large-scale psychometric testing with the goal of assessing its internal and external validity properties more accurately. We welcome comments, consultation, and collaboration and hope to involve other researchers as we move further in the direction of an accurate and reliable method for assessing the impact of the common cold.

Acknowledgments
The authors would like to acknowledge the participants who contributed their time and energy while they were sick. We also thank the many physicians and coworkers who contributed their knowledge and opinions, especially Nora Cate Scheffler, PhD; Mary Beth Plane, PhD; Jon Temme, MD, PhD; Donn D’Alessio, MD; and William Scheckler, MD, in Madison, Wis., and Jack Gwaltney, MD, in Charlottesville, Va. During most of this project, Dr Barrett had had support from the National Center for Complementary and Alternative Medicine at the National Institutes of Health, Grant #K23 AT00051-01.

REFERENCES