Patients and Medical Statistics

Interest, Confidence, and Ability

Steven Woloshin, MD, MS, 1,2 Lisa M. Schwartz, MD, MS, 1,2 H. Gilbert Welch, MD, MPH 1,2

1The VA Outcomes Group (111B), Department of Veterans Affairs Medical Center, White River Junction, VT, USA; 2The Center for the Evaluative Clinical Sciences, Dartmouth Medical School, Hanover, NH, USA.

BACKGROUND: People are increasingly presented with medical statistics. There are no existing measures to assess their level of interest or confidence in using medical statistics.

OBJECTIVE: To develop 2 new measures, the STAT-interest and STAT-confidence scales, and assess their reliability and validity.

DESIGN: Survey with retest after approximately 2 weeks.

SUBJECTS: Two hundred and twenty-four people were recruited from advertisements in local newspapers, an outpatient clinic waiting area, and a hospital open house.

MEASURES: We developed and revised 5 items on interest in medical statistics and 3 on confidence understanding statistics.

RESULTS: Study participants were mostly college graduates (52%); 25% had a high school education or less. The mean age was 53 (range 20 to 84) years. Most paid attention to medical statistics (6%) paid no attention. The mean (SD) STAT-interest score was 68 (17) and ranged from 15 to 100. Confidence in using statistics was also high: the mean (SD) STAT-confidence score was 65 (19) and ranged from 11 to 100. STAT-interest and STAT-confidence scores were moderately correlated (r = .36, P < .001). Both scales demonstrated good test-retest repeatability (r = .60, .62, respectively), internal consistency reliability (Cronbach’s α = .70 and .78), and usability (individual item nonresponse ranged from 0% to 1.3%). Scale scores correlated only weakly with scores on a medical data interpretation test (r = .15 and .26, respectively).

CONCLUSION: The STAT-interest and STAT-confidence scales are usable and reliable. Interest and confidence were only weakly related to the ability to actually use data.

KEY WORDS: decision making; patient education; statistic.

DOI: 10.1111/j.1525-1497.2005.0179.x

METHODS

Scale Development

The STAT-interest and STAT-confidence scales were developed as part of a larger project to teach people how to interpret medical statistics. We developed the interest and confidence measures based on our own experience, medical literature reviews, and advice from experts in statistics, cognitive psychology, and education at Dartmouth College and Carnegie Mellon University. We revised items based on this feedback. All items underwent extensive pilot testing with patients and members of the general public (here the focus was on wording and understandability). Five items assessed interest; 3 items assessed confidence. According to the Flesch–Kincaid scoring algorithm (Microsoft Word X-Mac), the items are written at the eighth grade level.

The authors have no conflicts of interest to declare for this article or this research.

Dr. Schwartz is supported by Veterans Affairs Career Development Awards in Health Services Research and Development. Drs. Schwartz and Woloshin are supported by Robert Wood Johnson Generalist Faculty Scholar Awards and National Cancer Institute grants CA91052-01 and CA104721. All the authors are supported by a Research Enhancement Award from the Department of Veterans Affairs. The views expressed herein do not necessarily represent the views of the Department of Veterans Affairs or the United States Government.

Address correspondence and requests for reprints to Dr. Woloshin: VA Outcomes Group (111B), Department of Veterans Affairs Medical Center, White River Junction, VT 05009 (e-mail: steven.woloshin@dartmouth.edu).

See editorial by Montori, p. 1071.
Scale Evaluation

Subjects. We recruited 224 English-speaking people from advertisements in local newspapers (n=104), the White River Junction VA outpatient clinic waiting area (n=76), and a hospital open house at Dartmouth Hitchcock Medical Center (n=44). The study was approved by the Dartmouth College IRB. Patient recruitment materials asked adults age 20 years and older to help us learn how to give people the information they need to make good medical decisions. We asked people to call a research assistant (who confirmed the subject’s age). Participants were mailed a survey with a stamped return envelope. About 7 to 10 days after receipt of their completed survey, we mailed out a retest survey. Two hundred and twenty individuals (98%) completed the retest survey. Participants were paid $25.

Item and Scale Scores. For each item, we calculated the proportion of responses left blank to measure question usability. To create scores for each scale, we summed the points assigned to each response category. Most questions offered 5-level responses ranging from “strongly disagree” (assigned 0 points) to “strongly agree” (4 points) with a neutral middle category. We reversed the response order for negatively worded questions. Two questions used different response sets (3 and 4 levels). We revised the scoring for these questions so that all questions would contribute equally to the aggregate scores. We assigned the lowest response category “0” points, the highest category “4” points, and calculated intermediate values for the middle categories using a simple linear transformation (i.e., 3-level values were 0, 2, 4, and 4-level values were 0, 1.33, 2.66, 4). Aggregate scores were transformed onto a 0 to 100 point scale, with higher scores meaning more of the attribute (i.e., greater interest, greater confidence).

Test–Retest Repeatability. We calculated Pearson correlation coefficients of the test and retest scores for the 220 people who completed the retest survey 2 weeks later.

Reliability. We measured the internal consistency reliability with Cronbach’s α, a measure of how responses to each question correlate with responses to all other questions in the scale.

Content Validity. We used feedback of experts in education, statistics, and cognitive psychology (locally and at Carnegie Mellon University) to assess the content validity of the measures, the extent to which the items capture the concept being measured without missing important aspects of the concept.

Construct Validity. The extent to which a measure actually captures the intended concept or construct was evaluated by analyzing responses to 2 additional questions in our survey meant to capture interest in statistics. The exact questions were:

“Next time I have to make a decision about my health, I plan to …
  a. Ask my doctor for statistics (yes/no)
  b. Track down medical statistics (yes/no)

We hypothesized that respondents with greater interest in statistics would be more likely to actively try to find statistics when making medical decisions.

We also sought to learn whether people with greater interest in medical statistics or great confidence in interpreting them were better able to interpret medical statistics. Our measure of ability was the medical data interpretation test. The 18-item data interpretation test measures the ability to make sense of and compare medical statistics about disease risk and risk reduction. Scores from this test range from 0 to 100 where higher scores represent higher abilities. The data interpretation test was modified early in the study; analyses relating attitudes to ability include the 175 participants who received the final version of the test.

Analyses

Means, standard deviations, and frequencies were used to describe item and scale scores. All comparisons were 2 sided and were considered statistically significant at P<.05. We used Pearson correlation to assess the association between interest, confidence, and medical data interpretation skills, analysis of variance to compare scores in different categories of educational attainment, and χ² for the construct validity analysis of interest score quartiles. We used STATA 8.0 (College Station, Tex).

RESULTS

The 224 participants represented a range of ages, incomes, and formal education although 52% had a college degree (Table 1). The mean age was 53 (range 20 to 84) years, 48% were female, and most were white. Tables 2 and 3 show the basic attributes of the individual items and the aggregate scale scores. Item nonresponse was low (ranging from 0% to 1.3%) for the interest and confidence questions.

Interest in medical statistics was fairly high: 94% paid some attention to medical statistics (36% paid considerable attention, and 58% a little) and 80% agreed that “to make wise decisions about my health, it is important to know how to interpret statistics.” The STAT-interest score was normally distributed with a slight left skew; the mean score was 68 (SD 17), the median was 70, and scores ranged from 15 to 100. Interest in medical statistics was high at all education levels: mean scores were 60 for those with less than a high school education, 67 for high school graduates, 71 for college graduates, and 70 for postgraduates (P=.09).

Confidence in using statistics was also high: 72% thought statistics were easy to understand, and 79% felt “comforted that I can make sense of medical statistics.” The STAT-confidence scores were normally distributed with a slight left skew:

<table>
<thead>
<tr>
<th>Table 1. Characteristics of the Study Sample (n=224)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (range)</td>
</tr>
<tr>
<td>Sex (% women)</td>
</tr>
<tr>
<td>Race (% white)</td>
</tr>
<tr>
<td>Household income (%)</td>
</tr>
<tr>
<td>&lt; $10,000</td>
</tr>
<tr>
<td>$10,000 to 24,999</td>
</tr>
<tr>
<td>$25,000 to 49,999</td>
</tr>
<tr>
<td>$50,000 to 99,999</td>
</tr>
<tr>
<td>≥ $100,000</td>
</tr>
<tr>
<td>Highest level of education (%)</td>
</tr>
<tr>
<td>&lt; High school graduate</td>
</tr>
<tr>
<td>High school degree</td>
</tr>
<tr>
<td>Some college</td>
</tr>
<tr>
<td>College degree</td>
</tr>
<tr>
<td>Postgraduate degree</td>
</tr>
</tbody>
</table>

*Item nonresponse was: 10 (income), 3 (education).