Reliability of an Arabic Version of the RAND-36 Health Survey and its Equivalence to the US-English Version

STEPHEN JOEL COONS, PHD,* SAUD ABDULAZIZ ALABDULMOHSIN, MS,† JOEL R. DRAUGALIS, PHD,* AND RON D. HAYS, PHD‡

OBJECTIVES. The objectives of this research were to: (1) evaluate the reliability and equivalence of the Arabic and English versions of the RAND-36 Health Survey (RAND-36) in a sample of Saudi Arabian citizens; and (2) assess the health status of a sample of Saudi Arabian citizens using both the Arabic and English versions.

METHODS. Both the Arabic and English versions of the survey were administered to a convenience sample of bilingual (English and Arabic) Saudi citizens (n = 415) at Saudi ARAMCO Company, Dhahran, Saudi Arabia. Internal consistency, equivalent-forms, and test-retest reliability were estimated for the eight multi-item scales in the Arabic and English versions. Mean scale scores were calculated for each version and compared with the general US population.

RESULTS. The median Cronbach’s alphas for the Arabic RAND-36 in multiple subgroups exceeded 0.70 for every scale except one. Two of the English RAND-36 scales had median Cronbach’s alphas that exceeded 0.70; the remainder exceeded 0.50. Two-week test-retest correlations were all statistically significant for both versions. Product-moment correlations to test the equivalence of the corresponding Arabic and English versions of the RAND-36 ranged from 0.73 to 0.92. Saudi citizens reported significantly higher vitality scores, but significantly lower physical functioning, social functioning, and general health perception scores than the general US population.

CONCLUSIONS. The results provide support for the reliability and equivalence of the Arabic and English versions of the RAND-36. Additional studies need to be conducted in a representative sample of the general Saudi population to further assess the psychometric properties of the Arabic version.

Key words: RAND-36, cross-cultural translation, quality of life, health status assessment, Saudi Arabia. (Med Care 1998;36:428–432)

The RAND-36 Health Survey is a generic health-related quality-of-life measure. It is comprised of the same items included in the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36). Both measures emanated from work begun at RAND in 1984 as part of the Medical Outcomes Study (MOS) and include multi-item scales that assess eight health concepts: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to emotional problems, mental health (emotional well-being), social functioning, vitality (energy/fatigue), and general health. The only difference between the measures is the scoring algorithm applied when calculating two of the scale scores (ie, general health, bodily

*From the Center for Pharmaceutical Economics and Department of Pharmacy Practice and Science, College of Pharmacy, The University of Arizona, Tucson.
†From the Pharmacy Services Division, Saudi ARAMCO Medical Services Organization, Saudi ARAMCO Company, Dhahran, Saudi Arabia.
‡From the Social Policy Department, RAND, Santa Monica, California and the Department of Medicine, School of Medicine, University of California, Los Angeles.

At the time this study was conducted, Mr. Alabdulmoisin was a graduate student in the Department of Pharmacy Practice at The University of Arizona College of Pharmacy.

Address correspondence to: S.J. Coons, PhD, Center for Pharmaceutical Economics, College of Pharmacy, PO Box 210207, The University of Arizona, Tucson, AZ 85721-0207; e-mail: coons@Pharmacy.Arizona.edu.
pain), which is independent of the measurement process. Hence, for the purpose of comparability with published general population norms, SF-36 scoring was used.

There is a growing body of evidence that the SF-36 can be successfully translated, validated, and normalized for use in other countries. This article describes the administration and evaluation of a new Arabic translation of the RAND-36 in a sample of Saudi Arabians. The translation process that preceded this work was conducted in accordance with a commonly accepted cross-cultural adaptation methodology. A description of the translation process, including a copy of the Arabic version, is available. In this article, the US-English version (other country-specific English versions of the SF-36 are now available) will be referred to as the English version for brevity.

Methods

Sample

This study was conducted at Saudi ARAMCO Company facilities in the eastern province of the kingdom of Saudi Arabia. A convenience sample was selected from several departments at Saudi ARAMCO Company. This sample was selected by asking each of the departments to generate a list of bilingual (Arabic and English) Saudi employees who were willing to participate in the study. Persons were selected from each department to work as coordinators with the second author.

After obtaining complete lists of participants from the departments, each person was assigned randomly to one of six groups. Eighty-nine subjects were assigned to each of Groups 1, 2, and 3, and 88 subjects were assigned to each of Groups 4, 5, and 6.

Data Collection

Data were collected by distributing the English and/or Arabic versions of the RAND-36 to the participants through the coordinators. Cover letters explaining the purpose of the study were provided. Additional questions were included to gather demographic information (ie, gender, age, and education) and the questionnaires were coded so that nonrespondents could be contacted. Only the second author had access to the list that linked the questionnaire codes to specific individuals. These questionnaires were adminis-

tered twice (time 1 = initial; time 2 = retest) as follows:

Group 1. Participants completed the English version immediately followed by the Arabic version. The assigned coordinator distributed a sealed envelope containing both versions to the participants. The participants were instructed to complete both questionnaires as follows: (a) complete the English questionnaire, insert it in the envelope provided and seal it, (b) complete the Arabic questionnaire, insert it in the other envelope and seal it, and (c) return both envelopes to the assigned coordinator. Aside from the order of administration, instructions for Groups 2 through 6 were as discussed for Group 1.

Group 2. Participants completed the Arabic version first, immediately followed by the English version.

Group 3. Participants completed the English version followed by the Arabic version after a 2-week interval.

Group 4. Participants completed the Arabic version followed by the English version after a 2-week interval.

Group 5. Participants completed the Arabic version twice (2 weeks apart).

Group 6. Participants completed the English version twice (2 weeks apart).

Scoring

All scales were transformed linearly to a 0 to 100 possible range of scores, with 0 and 100 representing the least and most favorable health state, respectively. All scores reflect the percent of the total possible score for that scale. For comparison of scores from this sample to that of the US general population, scoring of the RAND-36 followed the procedures used for the published norms for the SF-36. For other analyses, scoring of the pain and general health perception scales followed the RAND-36 recommendations.

Data Analysis

Statistical analyses were conducted using SPSS for VAX/VMS, Version 4.1.

Descriptive Statistics. The mean ages, percent male and female, and education level of the respondents in each group were calculated.
Internal Consistency Reliability. Cronbach's coefficient alpha was used to estimate internal consistency reliability for each of the administrations of Arabic and English versions.\textsuperscript{8}

Test–Retest Reliability. To assess test–retest reliability during a 2-week interval, Pearson product–moment correlations were computed between initial and retest administrations for both Group 5 and Group 6.

Equivalent-Forms Reliability. Pearson product–moment correlation coefficients were computed to evaluate equivalent forms reliability between the Arabic and English versions administered on the same day (Groups 1 and 2).

Scale Score Means. Scale score means were computed for the Arabic and English versions to determine if central tendency varied by version. In addition, paired t tests were computed comparing initial and retest administrations.

Results

Subjects

Of the 531 subjects randomly assigned to the six study groups, 446 (84\%) individuals returned their questionnaires. Of these, 31 individuals were excluded from the study because they had missing data for one or more items. Thus, the analytic sample size was 415.

The mean age of the subjects was 34.9 ± 7.0 years, with a range of 19 to 59 years. Seventy-eight percent of the subjects were male, and 52\% were college graduates. A one-way analysis of variance comparing the age of the six groups revealed no statistically significant difference [F(5, 358) = 0.929, P > 0.05]. Chi-square analysis revealed no significant differences in the proportion of men and women in the six study groups.

Internal Consistency Reliability: Arabic Version

Initial Administration. Alpha reliability coefficients for the initial administration of the Arabic version (Groups 2, 4, and 5, n = 212) ranged from 0.60 (general health) to 0.87 (physical functioning). Except for the general health scale, reliability coefficients exceeded 0.70, satisfying Nunnally's standard of acceptable reliability for group-level studies.\textsuperscript{9}

Retest Administration. Alpha coefficients for retest administrations (Groups 1, 3, and 5, n = 181) ranged from 0.57 (general health) to 0.88 (physical functioning), and were 0.70 or higher for all scales except for general health.

Internal Consistency Reliability: English Version

Initial Administration. Alpha coefficients for the initial administrations of the English version (Groups 1, 3 and 6, n = 181) ranged from 0.56 (role limitations—physical, social functioning) to 0.89 (physical functioning). Alpha coefficients for four other scales were less than 0.70 (ie, vitality (0.58); general health (0.61); role limitations—emotional (0.66); mental health (0.68)).

Retest Administration. Alpha coefficients for the retest administrations (Groups 2, 4 and 6, n = 179) ranged from 0.60 (social functioning) to 0.90 (physical functioning). Alpha coefficients for two other scales were less than 0.70 (ie, vitality (0.61); general health (0.68)).

Test–Retest Reliability

Two-week test–retest product–moment correlations were all statistically significant and ranged from 0.29 (bodily pain) to 0.80 (mental health) for the Arabic version (n = 63) and from 0.46 (role limitations—physical) to 0.77 (general health) for the English version (n = 52).

Equivalent-Forms Reliability

Equivalent-forms product–moment correlations between corresponding scales for those administered the English version immediately followed by the Arabic version (n = 53) ranged from 0.78 (social functioning) to 0.91 (mental health). Correlations for those administered the Arabic version immediately followed by the English version (n = 65) ranged from 0.73 (vitality) to 0.92 (mental health).

Mean Scores for the Arabic and English Versions

The mean scale scores for the initial administration of the Arabic version were computed for participants in Groups 2, 4, and 5 (n = 170). The mean scale scores for the initial administration of the English version were computed for participants in Groups 1, 3, and 6 (n = 177). For this analysis, the bodily pain and general health perception scales were calculated using the SF-36
scoring system used to create the general US adult population norms (see Table 1). The mean scale scores for a sample of adults in the general US population matched for age and gender to the Saudi Arabian samples also are provided.7

**Differences Between the Mean Scale Scores**

Paired t tests were computed comparing mean scale scores for the Arabic and English versions. There were no significant differences for those administered the English version followed immediately by the Arabic version (n = 53), supporting the equivalence of the two versions. For those administered the Arabic version followed immediately by the English version (n = 65), no significant differences were found for six of the eight scales; however, means for the vitality and social functioning scales were significantly different (average differences were only approximately 3 points on the 0 to 100 score range).

**Discussion**

The median internal consistency reliability coefficients for all administrations (Groups 1, 3, and 5) of the Arabic version exceeded 0.70 for every scale except for general health (median alpha = 0.59). The median internal consistency reliability coefficients for all administrations (Groups 2, 4, and 6) of the English version exceeded 0.50. Therefore, the results of this study provide support for the reliability of the Arabic version and are consistent with previous reliability estimates reported for the English version.

Both the Arabic and English versions tended to have internal consistency reliability coefficients equal to or above the acceptable standards for group comparisons;9,10 however, the Arabic version had higher median values than the English version. Although all participants were bilingual, their mother language was Arabic. As a result, it is likely that the participants had a better understanding of the Arabic version than the English version, leading to more internally consistent responses. In addition, the Arabic version was adapted to the Saudi culture, whereas the English version was developed for the dominant US culture.

Test–retest correlations during a 2-week time interval were similar for the English and Arabic versions. The size of the correlations indicate a noteworthy degree of variation between initial

---

**Table 1. Mean RAND-36 Scale Scores for the Initial Administrations of the Arabic and English Versions of the RAND-36 Compared With the General US Population**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Group 2, 4, and 5</th>
<th>Group 1, 3, and 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Items</td>
<td>Initial Arabic Administrations (n = 170)</td>
</tr>
<tr>
<td>Physical Functioning</td>
<td>10</td>
<td>87.3 (17.1)†</td>
</tr>
<tr>
<td>Role Limitations–Physical</td>
<td>4</td>
<td>84.3 (30.3)‡</td>
</tr>
<tr>
<td>Role Limitations–Emotional</td>
<td>3</td>
<td>76.3 (35.2)‡</td>
</tr>
<tr>
<td>Vitality</td>
<td>4</td>
<td>67.6 (16.8)‡</td>
</tr>
<tr>
<td>Mental Health</td>
<td>5</td>
<td>75.2 (16.1)</td>
</tr>
<tr>
<td>Social Functioning</td>
<td>2</td>
<td>82.0 (19.3)‡</td>
</tr>
<tr>
<td>Bodily Pain§</td>
<td>2</td>
<td>80.3 (21.9)</td>
</tr>
<tr>
<td>General Health†</td>
<td>5</td>
<td>71.1 (16.1)§</td>
</tr>
</tbody>
</table>

*Mean scale scores for the general US population were matched for age and gender to the individual samples. Significance of differences between the mean RAND-36 scale scores for the sample and the general US population:
†P < 0.01.
‡P < 0.05.
§P < 0.0001.
¶P < 0.001.

Scale scores were calculated using the SF-36 scoring system.
and retest administrations. The timing of the study may have contributed to the fluctuation in scale scores. Many Saudi citizens suffer from colds, influenza, and related symptoms during the time of year the study was conducted. Because of possible real changes, test–retest reliability estimates need to be evaluated with caution, especially when studying a dynamic process such as health status.11

The equivalence of the corresponding Arabic and English versions of the RAND-36 scales was assessed using correlations and paired t tests. Product-moment correlations ranged from 0.73 to 0.92 between corresponding scales. Only two out of 16 paired t tests were statistically significant. These results provide strong support for the equivalence of the Saudi and English versions.

When the mean scale scores for the Arabic version were compared with those for the general US population, a number of significant differences were found (Table 1). The Saudi citizens (n = 170) had significantly lower mean scale scores for five of the eight scales when compared with the US sample. In addition, one scale score mean (ie, vitality) was significantly higher in the Saudi sample than in the US sample.

When the mean scale scores for the English version were compared with those for the general US population, four of the eight scale score means were significantly different (Table 1). Again, only the mean scale score for vitality was higher in the Saudi citizens (n = 177) than in the US sample. The US sample had higher mean scores than the Saudi sample for the other scales in which significant differences were found (ie, physical functioning, social functioning, and general health).

The results of this study provide support for the reliability of an Arabic version of the RAND-36 and its equivalence with the English version. This sample, however, is not necessarily representative of the Saudi population. Additional studies need to be conducted to test the reliability and validity of the Arabic version using a well-designed sampling frame from the general Saudi population. Studies also are needed to evaluate the sensitivity of the Arabic version to variation in disease severity as well as its responsiveness to the effects of medical treatments. In addition, further research should examine to what extent this new Arabic version is applicable in other Arab cultures.

Acknowledgments

The authors thank Saudi ARAMCO Company, The University of Arizona’s Center for Pharmaceutical Economics, and RAND for their assistance.

References


