Abstract

Objective To adapt the Nottingham Health Profile (NHP) for Japanese and to describe the results of the assessment of its psychometric properties.

Methods Assessments included test-retest reliability over approximately a 2-week interval, internal consistency and construct validity in 133 patients with COPD.

Results The distribution of scores indicated that most of the NHP sections exhibited a floor effect, although this is greatly reduced with the NHP-Distress scale. The test-retest reliability was above 0.8 for all sections when patients reporting any change in their health status rating were excluded. Cronbach’s alpha coefficients reflected the number of items contained in each section. The internal consistency of the emotional reactions section at one timepoint and the physical mobility section at both timepoints were lower than expected to be higher. All sections except the pain section could be used to distinguish patients who reported their health status to be good or fair from those who rated it to be poor or very poor.

Conclusion The adaptation of the NHP for Japanese was successful. Most sections showed reasonable test-retest reliability, indicating that they produced acceptable levels of random measurement error. The internal consistency of the sections was confirmed, although the alpha values of the emotional reactions and physical mobility sections were lower than might be expected for scales of their length. Different sections of the Japanese NHP were shown to have known group validity.

Key words: Nottingham Health Profile, health status, quality of life, health-related quality of life, perceived distress

Introduction

To evaluate generic health status or health-related quality of life (HRQoL), many “instruments” such as the Sickness Impact Profile (SIP) (1), the Medical Outcome Study (MOS) short-form (SF)-36 (2), the Nottingham Health Profile (NHP) (3, 4), and the Quality of Well Being (QWB) (5) are used in the United States and in Europe. Since these instruments were created in the United States and Europe, work is being done to establish Japanese versions of some of the instruments in order that they be used in Japan. When the purpose of the evaluation is to compare the health status of patients with different diseases, to study epidemiology, or to show data related to medical policy issues, evaluation using generic instruments is necessary. Development of such instruments that can be used in Japan is highly desired. For the SF-36 or NHP used in the United States and Europe, data representing a normal range based on age and sex from surveys of a general population sample are available. Fukuhara and colleagues have been developing a Japanese version of the SF-36 in an international project (6–8). This enables a comparison of health status in areas of different languages.

In general, various methods are used to make Japanese versions of surveys; multiple translations between Japanese and English, or panel discussions with patients are the most common methods (6, 9–15). The Japanese version must have
the same characteristics of the original English survey. If the answer to a certain question differs among patients with a similar background, one can assume that the translation was not successful, or there are cultural differences. After revision or by changing some of the questions, the characteristics of the original questionnaire can be maintained.

The NHP is a generic, self-administered questionnaire originally designed to measure perceived health problems (3, 4). However, the NHP has been used by many clinical researchers to examine the HRQoL in the literature (16–20). Although other generic instruments such as the SIP and SF-36 were developed in the USA, the NHP was established in the UK (21, 22).

The purpose of the present work was to develop a Japanese version of the NHP based on international collaboration, and to describe the psychometric testing of the Japanese adaptation of the NHP. Assessments included test-retest reliability over approximately a 2-week interval, internal consistency and construct validity.

Methods

The Nottingham Health Profile

The NHP contains 38 items grouped into six sections relating to patient distress. These sections are energy level (3 items), pain (8 items), emotional reactions (9 items), sleep (5 items), social isolation (5 items), and physical mobility (8 items). All items have a yes/no answer format. Each section is scored from 0 to 100; the higher the score, the greater the perceived health problems.

In addition to these six sections there is also an index embedded within the NHP; the NHP-Distress (NHPD) index. This scale has 24 items. Scores range from 0 to 24, with a higher score indicating greater distress.

Translating the NHP into Japanese

A first translation panel was composed of six bilingual Japanese persons (3 males and 3 females) of different ages. Health care providers and chronically ill individuals were not included. This panel was headed by Japanese researchers (K.N. and T.H.), and an English research scientist (S.P.M.) also participated, in order to explain the precise meaning of the items. This panel considered the most appropriate translations of the items, instructions and responses. If the panel could not agree on a single translation for a cord or phrase, alternative translations were passed to the second translation panel for consideration.

The second panel was headed by the same Japanese researchers (K.N. and T.H.). The lay panel consisted of six Japanese persons (3 males and 3 females) of different ages who were free from chronic illness, had average or below average education, and were not able to speak English. Participants considered the items, instructions and responses, chose between alternative translations presented or made their own suggestions, providing that the original meaning was maintained. Modifications were subsequently made and a second translation was produced.

Field-testing the draft questionnaire involved interviewing 15 individuals with chronic obstructive pulmonary disease (COPD). The interviews were conducted at Kyoto University. The interviewer (T.H.) explained the purpose of the study and the nature of the instrument and its method of development. When the respondent had completed the NHP, they were asked if they found the questions relevant, easy to understand, and properly worded. Suggestions for changes in wording were noted. The interviewer then questioned the respondents in more detail about difficult or missed items and the cause of the difficulty. A reconciled final version was forwarded to a survey to determine reliability and construct validity.

Study design

The test-retest reliability of a measure is an estimate of its reproducibility over time when no change in condition has taken place. This was examined at approximately 2-week intervals in consecutive COPD patients. The following information was collected: demographics (gender, age and marital status), health status (good/fair/poor/very poor) at times 1 and 2, and the NHP at times 1 and 2. Patients who had been attending the outpatient clinics at Kyoto University and Ayabe Municipal Hospital over a period of six months were eligible for the study. They were surveyed at a time when they were clinically stable. Entry criteria included a diagnosis of COPD based on the definition provided by the American Thoracic Society; chronic airflow limitation, a smoking history of over twenty pack-years, no exacerbation in the preceding four weeks, and no changes in treatment regimen over the preceding four weeks. Chronic airflow limitation was defined by a forced expiratory volume in one second (FEV1)/forced vital capacity (FVC) ratio of less than 0.7 for all previous measurements.

Statistical analysis

The test-retest reliability was assessed by correlating scores on the instrument collected on two different occasions. A high correlation shows that the instrument produces little random measurement error. As the NHP produces ordinal scores, Spearman rank correlation coefficients were employed to assess the level of association. The coefficient was calculated for the whole sample and for the subset who had no change in their health status among surveys between time points.

Internal consistency was assessed using Cronbach’s alpha coefficients (23). Alpha measures the extent to which the items in a scale are interrelated. The reason for the low alpha value for some sections can be found in the complete item statistics resulting from the item reliability analysis. These statistics include the corrected item-total correlations (CITC), which indicate the degree to which individual items are associated with their respective scales.

Known group validity can be assessed by testing the ability of the measures to distinguish between groups of patients.
that differ according to some known factor. The factor used for the present investigation was patient-perceived health status. Mann-Whitney U test (for two groups) or Kruskal-Wallis one-way analysis of variance (for three or more groups) were employed to test for differences between groups.

**Results**

**Patient demographic and health status information**

Of the 133 patients in the sample, 130 (97.7%) were males and 3 (2.3%) were females. Their age ranged between 50 and 86 years with a mean age of 71.1 years (SD 6.4) and a median age of 71.0 years (interquartile range 67.0 to 75.5). One-hundred-and-sixteen patients (87.2%) were married, 14 (10.5%) were widowed, 2 (1.5%) were divorced or separated and 1 (0.8%) was single. The patients’ rating of their health status is shown in Table 1.

**Score distribution of the NHP and external factors**

Table 2 shows the score distribution of the NHP at each administration. The percentage of respondents scoring the minimum and maximum possible values shows that most of the NHP sections have a floor effect (where patients score 0), although this effect is greatly reduced with the NHP-Distress scale. Ceiling effects (where patients score 100) are minimal for all sections.

Comparison of NHP section scores by gender was not possible due to the low number of females in the sample. Very few differences by age or marital status were found. The only observed differences were with patients aged over 70 years who scored higher on the physical mobility section than those aged 70 or below (p<0.05 at both time points) and unmarried patients who scored higher than married patients on the pain section (p<0.05 at both time points).

**Test-retest reliability**

Spearman rank correlation coefficients were used to evaluate the test-retest reliability of the different sections of the NHP. Table 3 shows that the test-retest reliability was above 0.85 for the social isolation section and the NHP-Distress index only. However, although the values for the other sections are less than desirable, they are above 0.7. Furthermore, when those patients who reported any change in their health status rating were excluded, the reliability coefficients for all sections were above 0.8. The reliability obtained compares favorably with that found for other language versions of the NHP.

Table 4 shows mean values with standard deviations and the median values with interquartile ranges at each admini-
It can be seen that the median values were similar at both administrations, with scores not significantly different. This confirms the stability of the scores.

**Internal consistency**

Table 5 shows Cronbach’s alpha coefficients for the six sections of the Japanese version of the NHP. The values obtained reflect the number of items contained in each section. The exceptions to this are for the emotional reaction section at time 1 and the physical mobility section at both time points; the alpha values for both these scales would be expected to be higher, given their number of items (nine and eight respectively).

*including only those patients whose health status did not change, from reference #32.

*from “unpublished study of 279 UK patients with psoriatic arthritis”, from reference #33.
The CITC for item 16 (“The days seem to drag.”) in the emotional reaction scale was low. The removal of this item would raise the overall alpha to 0.82, a value more in line with that obtained at time 2. The low association of this item with the others is also evident in the NHP-Distress scale. However, the problem is not as evident at time 2, either in the emotional reaction section or in the NHP-Distress scale. A single problematic item can also be identified in the physical mobility section (item 14: “I’m unable to walk at all.”). However, unlike with the emotional reaction section, the poor performance of one item is not the only reason for the low alpha values; the removal of this item, or any other, would not raise the alpha to an acceptable level.

The CITC’s obtained from the other sections confirm the findings of the overall alpha; the items are adequately inter-related. The minor exception to this is the pain section, where the CITC for item 4 (“I have unbearable pain.”) is low at time 1. Overall, however, the problem is not considered great.

**Construct validity**

Evidence of the validity of the NHP was gained by examining the measure’s ability to distinguish between groups of patients that differed according to their self-rated health status (good, fair, poor or very poor). Due to the low number of observations in the extreme categories of health status, only two groups were compared in the following analyses; good/fair and poor/very poor. The results of the comparisons are reported in Table 6. All sections were able to distinguish patients who reported their health status to be good/fair from those who rated it to be poor/very poor, with the latter group obtaining higher scores (greater distress) at both time points (Table 6). The exception to this was the pain section, which showed no difference in scores at time 1.

**Discussion**

The results of the present study show that, in general, the adaptation of the NHP into Japanese has been successful. We found the Japanese version of the NHP to be acceptable and easy to administer to male patients with COPD. The results presented are consistent with previous reports on the original NHP instrument, and suggest that the Japanese version is similarly reliable and valid.

Most sections showed reasonable test-retest reliability, indicating that they produce little random measurement error.

<p>| Table 6. Ability of the NHP to Distinguish Different Self-rated Health Status Groups |
|-----------------------------------------------|---------------|---------------|---------------|
| Time 1 | Time 2 |</p>
<table>
<thead>
<tr>
<th>n</th>
<th>Median (IQR)</th>
<th>p-value*</th>
<th>n</th>
<th>Median (IQR)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy level</td>
<td>74</td>
<td>33.3 (0.0–33.3)</td>
<td>&lt;0.001</td>
<td>69</td>
<td>0.0 (0.0–33.3)</td>
</tr>
<tr>
<td>Good/Fair</td>
<td>51</td>
<td>66.7 (33.3–100.0)</td>
<td>ns</td>
<td>46</td>
<td>66.7 (33.3–100.0)</td>
</tr>
<tr>
<td>Poor/Very poor</td>
<td>72</td>
<td>0.0 (0.0–12.5)</td>
<td>&lt;0.001</td>
<td>71</td>
<td>0.0 (0.0–12.5)</td>
</tr>
<tr>
<td>Pain</td>
<td>50</td>
<td>12.5 (0.0–37.5)</td>
<td>&lt;0.001</td>
<td>45</td>
<td>11.1 (0.0–55.6)</td>
</tr>
<tr>
<td>Good/Fair</td>
<td>73</td>
<td>0.0 (0.0–11.0)</td>
<td>&lt;0.001</td>
<td>71</td>
<td>0.0 (0.0–11.1)</td>
</tr>
<tr>
<td>Poor/Very poor</td>
<td>50</td>
<td>22.2 (0.0–47.2)</td>
<td>&lt;0.001</td>
<td>45</td>
<td>11.1 (0.0–55.6)</td>
</tr>
<tr>
<td>Emotional reactions</td>
<td>75</td>
<td>20.0 (0.0–20.0)</td>
<td>&lt;0.001</td>
<td>70</td>
<td>0.0 (0.0–20.0)</td>
</tr>
<tr>
<td>Good/Fair</td>
<td>49</td>
<td>20.0 (0.0–40.0)</td>
<td>&lt;0.005</td>
<td>47</td>
<td>40.0 (20.0–60.0)</td>
</tr>
<tr>
<td>Poor/Very poor</td>
<td>72</td>
<td>0.0 (0.0–20.0)</td>
<td>&lt;0.001</td>
<td>70</td>
<td>0.0 (0.0–40.0)</td>
</tr>
<tr>
<td>Social isolation</td>
<td>49</td>
<td>20.0 (20.0–80.0)</td>
<td>&lt;0.001</td>
<td>46</td>
<td>40.0 (20.0–60.0)</td>
</tr>
<tr>
<td>Good/Fair</td>
<td>72</td>
<td>12.5 (0.0–25.0)</td>
<td>&lt;0.001</td>
<td>71</td>
<td>12.5 (0.0–25.0)</td>
</tr>
<tr>
<td>Poor/Very poor</td>
<td>49</td>
<td>25.0 (12.5–25.0)</td>
<td>&lt;0.001</td>
<td>46</td>
<td>25.0 (12.5–37.5)</td>
</tr>
<tr>
<td>Physical mobility</td>
<td>69</td>
<td>2.0 (0.0–4.0)</td>
<td>&lt;0.001</td>
<td>68</td>
<td>1.0 (0.0–3.8)</td>
</tr>
<tr>
<td>Good/Fair</td>
<td>49</td>
<td>7.0 (4.0–11.0)</td>
<td>&lt;0.001</td>
<td>43</td>
<td>6.0 (3.0–12.0)</td>
</tr>
</tbody>
</table>

*Mann-Whitney U Test.
The internal consistency of the sections was confirmed, although the alpha values of the emotional reactions and physical mobility sections were lower than might be expected for scales of such length.

In the case of the emotional reactions scale, this resulted from the poor performance of a single item. For both scales, problems are less evident at the second time point. It is recommended that future data sets are used to examine further the properties of these two sections. Most sections were able to distinguish patients who differed according to how they rated their health status, with those rating it as good or fair reporting lower levels of distress than respondents reporting poor or very poor health status. The only exception to this was the pain section, which failed to show differences at time 1.

In the present study assessment of convergent and divergent validity was not performed due to the unavailability of an appropriate comparator measure. The study had some limitations. The patient sample selected for the validation study was restricted to mainly male patients with COPD, as this was the sample for which the instrument was required in the first instance. This limits the conclusions that can be drawn about the validity of the measure for use with other patient groups. However, it is common practice to use generic health status measures such as the Nottingham Health Profile and SF-36 in specific patient groups without pre-validation. Such failure to establish the suitability of an instrument for use with specific populations could explain inconsistent results (24–26).

Relatively high proportions of patients in the study scored zero on individual sections of the NHP, which would limit the instrument’s ability to detect changes in health status. This finding is compatible with previous results reported for patients with COPD (27). However, scores on the NHP commonly exhibit such end effects (28). Indeed even larger end effects have been reported for other generic health status measures (28, 30). Such large end effects are a major cause of the insensitivity of generic measures and result from their need to cover the full range of health statuses (31).

Despite these limitations the results of this study indicate that the Japanese NHP has reasonable internal consistency and reliability and encouraging known group validity. Such a measure will provide an additional means of including patients in Japan.

Acknowledgments: The authors are grateful to Mitsubishi Pharma Corporation for sponsoring this work. Requests for the Japanese version of the NHP should be addressed to Koichi Nishimura, M.D.

References

22) Prieto L, Alonso J, Ferrer M, Anto JM. Are results of the SF-36 health survey and the Nottingham Health Profile similar? A comparison in...


