available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

# Translation and validation of the SF-36 Health Survey for use among Turkish and Moroccan ethnic minority cancer patients in The Netherlands

*Rianne Hoopman*[a,b], *Caroline B. Terwee*[b], *Martin J. Muller*[a], *Neil K. Aaronson*[a,*]

[a]*Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*

[b]*EMGO Institute, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands*

ARTICLE INFO

*Article history:*
Received 31 July 2006
Accepted 7 August 2006
Available online 2 October 2006

*Keywords:*
Quality of life assessment
SF-36
Cancer
Turkish
Moroccan
Translation
Ethnic minorities
Validation

ABSTRACT

In this study, the SF-36 Health Survey was translated into two oral Moroccan languages and the existing Turkish version was culturally adapted for use in The Netherlands, and was tested among 79 Moroccan and 90 Turkish cancer patients. There were normal levels of missing item responses but a higher administration time. With minor exceptions, the scale structure of the SF-36 was confirmed and the reliability of the scales met the 0.70 criterion for group comparisons. The questionnaire distinguished clearly between subgroups formed on the basis of performance status and was responsive to change in performance status over time. Some evidence of differential item function (DIF) was found in both ethnic groups. These results support the use of the SF-36 among Turkish and Moroccan cancer patients in The Netherlands. Additional studies are needed to confirm the psychometrics of the questionnaire when used among these ethnic minority groups in other Western European countries.



## 1. Introduction

There is an increasing need for questionnaires to assess the health status and health-related quality of life (HRQOL) of Turkish and Moroccan residents in Europe. There are currently more than 4 million Turks and Moroccans living in Western Europe.[1] An increasing number of first generation Turkish and Moroccan immigrants are now reaching the age at which chronic diseases, including cancer, are emerging as serious health problems. However, (first generation) ethnic minorities are typically excluded or seriously underrepresented in HRQOL studies because they do not speak the language of their host country, and the available HRQOL questionnaires are not available and have not been validated in their mother tongues.

One of the most widely used generic health status questionnaires is the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36). The SF-36 was developed in the United States in the late 1980s for a longitudinal investigation of the self-reported health status of patients with a range of chronic conditions.[2] It has been translated and validated in more than 50 languages (www.sf-36.org). However, the SF-36 has not been translated into Moroccan languages, and the Turkish version,[3] is yet to be investigated among Turkish immigrants living in Western Europe. This is important in that the majority of Turkish immigrants come from rural areas of Turkey,

* *Corresponding author:* Tel.: +31 20 512 2481; fax: +31 20 512 2322.
E-mail address: n.aaronson@nki.nl (N.K. Aaronson).
0959-8049/$ - see front matter 
doi:10.1016/j.ejca.2006.08.011

where educational and literacy levels are lower than those in the country as a whole.

In this paper, we report on the translation and psychometric evaluation of the SF-36 (version 1.0) in two oral Moroccan languages (Moroccan-Arabic and Tarifit), and the cultural adaptation of the Turkish version for use among ethnic minority cancer patients living in The Netherlands. We also report on the equivalence of these translations to the Dutch version of the SF-36, based on preliminary differential item function analysis.

## 2. Methods

### 2.1. Translation and cultural adaptation of the SF-36

The SF-36 is composed of 36 questions with standardised response choices, organised into eight multi-item scales. We used the 'acute' version with a 1-week time frame.[2,4,5] We followed the SF-36 guidelines for the forward–backward translation of the English language version of the SF-36 into two Moroccan languages.[6,7]

As only well-educated Moroccans have a good command of the official language in Morocco, Standard-Arabic, we translated the questionnaire into two oral languages commonly spoken among Moroccans in The Netherlands: Moroccan-Arabic and Tarifit. The Moroccan-Arabic translation was generated in phonetic Arabic script. The Tarifit language (spoken by Rifberbers in northern Morocco) has an original, ancient script. However, because it is not well known and rarely used, we generated the translation in the more commonly used Latin script. The Moroccan language versions were developed for oral administration. For both Moroccan versions we produced a male and a female version, as the grammar in these languages varies, in part, as a function of the sex of the respondent. Audiotaped versions of the two Moroccan translations were produced for purposes of interviewer training. In the Moroccan-Arabic language no equivalents could be found for some words, in which case loanwords from Standard-Arabic and, in one instance, a Dutch word (for 'vacuum cleaner') were used. Due to a lack of familiarity among Moroccan respondents with endorsing statements,[8,9] the general health perception items of the SF-36 (GH2-GH5; see Table 2) were converted to questions.

The Turkish version of the SF-36[3] was linguistically and culturally adapted for use among Turkish immigrants living in The Netherlands. This involved minor changes in wording that reflect the language as used currently in The Netherlands.

### 2.2. Sample and recruitment of patients

This study was part of a larger investigation of four HRQOL questionnaires for use among Turkish and Moroccan cancer patients in The Netherlands. Patients were recruited from seven outpatient oncology clinics in four cities in The Netherlands. Consecutive patients were eligible if they were at least 18 years old, had a life expectancy greater than 6 months, were diagnosed with cancer after 1985 and were still under medical supervision. Patients or one of their parents had to have been born in Turkey or Morocco. Finally, they had to be proficient in Moroccan-Arabic, Tarifit or Turkish, irrespective of their proficiency level in Dutch.

### 2.3. Instruments and procedures

Eligible patients were invited to participate by a bilingual letter followed by a personal invitation (by phone or in the waiting room) by one of the bilingual, female research assistants. The study was approved by the local ethical committees of the seven participating hospitals.

Patients completed the SF-36 together with three other HRQOL questionnaires (the EORTC QLQ-C30, the COOP/WONCA Charts and the Rotterdam Symptom Checklist), in random order. For Turkish patients, the questionnaires were either self-administered or administered in the form of an interview, depending on the preference of the patient. For Moroccan patients the questionnaires were administered orally. The SF-36 was administered twice with an interval of 3 months.

Data on diagnosis, stage of disease and treatment were retrieved from the hospital medical records. Demographic data and information on comorbidity, literacy levels and patient's judgement of their proficiency in Dutch were obtained from the patients. Performance status was assessed by the research assistants using the Karnofsky Performance Status scale (KPS).[10–12]

### 2.4. Statistical analyses

All SF-36 scale scores were transformed linearly to a scale from 0 to 100, with 0 and 100 representing the least and most favourable health outcomes, respectively.

*Descriptive statistics* were generated to evaluate the time required to complete the questionnaire, missing items, items for which explanations were provided, score distributions, and floor and ceiling effects.

*Multitrait scaling analyses* were employed to examine item-convergent validity (item-scale correlations should be >0.40) and item-discriminant validity (items should correlate significantly higher, i.e., 2 standard errors or greater, with their own scale than with other scales). Because the magnitude of the standard error is heavily influenced by sample size, and the sample size of the current study was relatively small, one standard error was also used as a more liberal criterion for evaluating item-discriminant validity.[13]

*Internal consistency reliability* of the multi-item scales was assessed by Cronbach's coefficient $\alpha$. A value of 0.70 or greater was considered as adequate for group comparisons.[14]

*Interscale correlations* were calculated to determine if the correlations between scales were lower than the internal consistency estimates of the scales, indicating that each scale was assessing a unique concept.[13]

*Known groups validity* was evaluated by comparing subgroups of patients known to differ on clinical variables. It was hypothesised that patients with a higher performance status, no comorbid conditions, local/locoregional disease and in follow-up would report better functioning than patients with a lower performance status, comorbidity, metastatic disease and under active treatment.

*Responsiveness* was evaluated by comparing changes over time in SF-36 scores in subgroups of patients formed on the

**Table 1 – Patients' sociodemographic and clinical characteristics**

| | Turkish (n = 90) | Moroccan (n = 79) |
|---|---|---|
| | N (%) | N (%) |
| *Gender* | | |
| Female | 47 (52) | 31 (39) |
| Male | 43 (48) | 48 (61) |
| *Age*[a] | | |
| 22–49 | 44 (49) | 34 (43) |
| 50–74 | 46 (51) | 45 (57) |
| *Education* | | |
| No education | 15 (17) | 36 (45) |
| Low | 40 (44) | 13 (16) |
| Middle | 24 (27) | 18 (23) |
| High | 11 (12) | 10 (13) |
| Missing | | 2 (3) |
| *Literate* | | |
| Yes | 77 (86) | 54 (68) |
| No | 12 (13) | 23 (29) |
| Missing | 1 (1) | 2 (3) |
| *Proficiency (speaking) Dutch* | | |
| Not/weak/poor | 66 (73) | 50 (63) |
| Good/excellent | 22 (25) | 27 (34) |
| Missing | 2 (2) | 2 (3) |
| *Mode of administration*[b] | | |
| Interview | 64 (71) | 69 (87) |
| Self-administered | 25 (28) | 10[f] (13) |
| Combination | 1 (1) | |
| *Primary cancer diagnosis* | | |
| Breast | 21 (23) | 18 (23) |
| Head and neck | 21 (23) | 14 (18) |
| Colorectal | 7 ( 8) | 8 (10) |
| Urogenital | 7 ( 8) | 9 (11) |
| Gynaecological | 9 (10) | 3 ( 4) |
| Lung | 8 ( 9) | 5 ( 6) |
| Other | 17 (19) | 22 (28) |
| *Stage at first assessment* | | |
| Local/locoregional | 70 (78) | 66 (83) |
| Metastatic | 20 (22) | 11 (14) |
| Missing | | 2 (3) |
| *In follow-up or active treatment*[c] | | |
| In follow-up | 65 (72) | 64 (81) |
| Active treatment | 25 (28) | 15 (19) |
| *Time since primary diagnosis* | | |
| 0–1 year | 43 (48) | 30 (38) |
| 2–5 year | 26 (29) | 29 (37) |
| 6–17 year | 21 (23) | 20 (25) |
| *Comorbidity*[d] | | |
| None | 28 (31) | 26 (33) |
| 1 | 24 (27) | 10 (13) |
| 2 or more | 38 (42) | 43 (54) |
| *KPS score*[e] | | |
| 30–70 | 41 (46) | 37 (47) |
| 80–100 | 48 (53) | 38 (48) |
| Missing | 1 (1) | 4 (5) |

a Mean age Turkish: 49.5 (SD = 12.0) years; Mean age Moroccans: 50.4 (SD = 13.3) years.
b Including the 6 questionnaires that could not be fully administered; all six were intended to be orally administered.
c Active treatment, under active treatment with chemotherapy, radiotherapy or completed <2 months ago. In follow-up, no current treatment.
d Comorbidity, one or more of the self reported conditions: diabetes, kidney disease, cardiovascular disease, high blood pressure, COPD, arthritis, back pain.
e KPS, Karnofsky Performance Status; Mean KPS score Turkish: 71.6 (range 40–100); median score = 80; Mean KPS score Moroccans (73.5 (range 30–100); median score = 80).
f The translated QLQ-C30 version for the Moroccan group was intended to be orally administered, but 10 highly educated patients were able to and preferred to complete the Moroccan-Arabic version in written form.

basis of changes in performance status (KPS score improved, stable or deteriorated). It was hypothesised that patients whose KPS scores improved or remained stable over time would report better HRQOL than patients whose KPS scores had worsened.

Finally, *differential item functioning* (DIF) was evaluated to test the equivalence of the Turkish and Moroccan translations to the Dutch version of the SF-36. We compared the data of the Turkish and Moroccan samples with those from a previous study of 376 Dutch cancer patients.[15] We tested uniform and non-uniform DIF for all items from the eight scales using ordinal regression analysis.[16] We first tested for non-uniform DIF by modelling the item response as a logit-linear function of the translation (Dutch *versus* Turkish, or Dutch *versus* Moroccan), the scale score and the interaction between translation and scale score. The interaction term represents the possible non-uniform DIF. Non-uniform DIF (indicating that the magnitude and direction of cultural/language differences in item scores varies as a function of the overall scale score) was considered to be present when the interaction term was significant with a *P*-value less than 0.001. For items without non-uniform DIF, uniform DIF was tested by modelling the item response as a logit-linear function of the translation and the scale score, with the translation term representing possible uniform DIF. Uniform DIF (testing the direction and magnitude of cultural/language differences in item scores) was considered to be present if the odds ratio of the translation term was outside the interval 0.53–1.89 (log odds ratio, $\beta$, numerically larger than 0.64). All analyses were corrected for sex, age and stage of the disease.

For most of the psychometric analyses the two Moroccan language samples were analysed together, due to the limited sample sizes. The exception was the internal consistency reliability analyses and the interscale correlation analyses, which were performed separately for these two language groups.

## 3. Results

### 3.1. *Sample response, sociodemographic and clinical characteristics*

A bilingual letter of invitation was sent to 140 Turkish and 175 Moroccan patients. Of these letters, 21 addressed to the Turkish patients and 34 to the Moroccan patients were not deliverable and up-to-date addresses or telephone numbers

could not be obtained. Two Turkish and three Moroccan patients were excluded because they did not speak the language under study. Of the remaining patients who were traceable and eligible, 90 Turkish and 79 Moroccan patients (48 of whom spoke Moroccan-Arabic and 31 Tarifit) participated, representing a 77% and 57% response rate, respectively.

Reasons for non-participation included lack of interest, feeling too ill and not being allowed to participate by the family. Table 1 shows the sociodemographic and clinical characteristics of the study sample.

At the time of the second assessment, 82 Turks and 71 Moroccans were still potentially available for continued study participation (8 Turks and 8 Moroccans were excluded as they were terminally ill, had returned to Turkey or Morocco, had failed to successfully complete the questionnaire at the first assessment or had died). Of these remaining patients, 58 Turks and 46 Moroccans completed the second assessment,

**Table 2 – Missing values, items of the SF-36 requiring explanation**

| Item | Description | Turkish (N = 89) | | | Moroccan (N = 74) | | |
|---|---|---|---|---|---|---|---|
| | | Missing (%) | Interviews (N = 62) Not missing but explained (%)[a] | Random answer (N = 62)[b] | Missing (%) | Interviews (N = 62) Not missing but explained (%)[a] | Random answer (N = 62)[b] |
| HT | Health compared to one week ago | – | – | – | **3 (4.1%)** | – | – |
| PF1 | Vigorous activities | 1 (1.1%) | – | – | **3 (4.1%)** | **13 (21.0%)*** | 1 (1.6%) |
| PF2 | Moderate activities | 3 (3.4%) | – | – | – | 2 (3.2%) | – |
| PF3 | Lifting, carry groceries | 3 (3.4%) | 1 (1.6%) | – | **9 (12.2%)*** | – | – |
| PF4 | Climb several flights | – | – | – | – | **4 (6.5%)** | 1 (1.6%) |
| PF5 | Climb one flight | – | – | – | 1 (1.4%) | 2 (3.2%) | 1 (1.6%) |
| PF6 | Bend, kneel | – | – | – | 2 (2.7%) | **3 (4.8%)** | 1 (1.6%) |
| PF7 | Walk mile (kilometer) | 3 (3.4%) | 1 (1.6%) | – | 2 (2.7%) | – | **5 (8.1%)*** |
| PF8 | Walk several blocks (few 100 m) | 2 (2.2%) | 1 (1.6%) | – | **4 (5.4%)** | – | **3 (4.8%)** |
| PF9 | Walk one block (100 m) | 3 (3.4%) | – | 1 (1.6%) | **3 (4.1%)** | 1 (1.6%) | **3 (4.8%)** |
| PF10 | Bathe, dress | – | – | – | 1 (1.4%) | 1 (1.6%) | – |
| RP1 | Cut down time – physcial | **8 (9.0%)*** | 1 (1.6%) | – | **5 (6.8%)** | 1 (1.6%) | 1 (1.6%) |
| RP2 | Accomplished less – physical | 3 (3.4%) | 1 (1.6%) | 1 (1.6%) | **3 (4.1%)** | 1 (1.6%) | 1 (1.6%) |
| RP3 | Limited in kind – physical | 2 (2.2%) | 2 (3.2%) | – | **8 (10.8%)*** | 1 (1.6%) | 1 (1.6%) |
| RP4 | Had difficulty – physical | 2 (2.2%) | – | – | **4 (5.4%)** | 2 (3.2%) | 1 (1.6%) |
| BP1 | Pain | – | – | – | 1 (1.4%) | – | – |
| BP2 | Pain interfere | 1 (1.1%) | – | – | 2 (2.7%) | – | 1 (1.6%) |
| GH1 | General health | **4 (4.5%)** | – | 1 (1.6%) | 1 (1.4%) | 1 (1.6%) | – |
| GH2 | Sick easier | **5 (5.6%)** | – | 1 (1.6%) | **4 (5.4%)** | **4 (6.5%)** | **4 (6.5%)** |
| GH3 | As healthy | **5 (5.6%)** | 1 (1.6%) | 1 (1.6%) | 2 (2.7%) | 2 (3.2%) | **3 (4.8%)** |
| GH4 | Health to get worse | **6 (6.7%)** | – | 2 (3.2%) | **3 (4.1%)** | – | 2 (3.2%) |
| GH5 | Health excellent | **6 (6.7%)** | – | – | **5 (6.8%)** | – | **3 (4.8%)** |
| VT1 | Full of pep | **4 (4.5%)** | – | 1 (1.6%) | 2 (2.7%) | **3 (4.8%)** | – |
| VT2 | Energy | **5 (5.6%)** | 1 (1.6%) | 1 (1.6%) | **5 (6.8%)** | – | – |
| VT3 | Worn out | 2 (2.2%) | 1 (1.6%) | 1 (1.6%) | **3 (4.1%)** | – | – |
| VT4 | Tired | 2 (2.2%) | – | – | **3 (4.1%)** | – | – |
| SF1 | Social extent | **4 (4.5%)** | – | – | **4 (5.4%)** | – | – |
| SF2 | Social time | **6 (6.7%)** | – | – | **5 (6.8%)** | – | 1 (1.6%) |
| RE1 | Cut down time - emotional | 2 (2.2%) | – | – | **4 (5.4%)** | **3 (4.8%)** | - |
| RE2 | Accomplished less - emotional | 2 (2.2%) | – | – | **3 (4.1%)** | 1 (1.6%) | 2 (3.2%) |
| RE3 | Not careful - emotional | **6 (6.7%)** | – | 1 (1.6%) | **6 (8.1%)*** | 2 (3.2%) | 2 (3.2%) |
| MH1 | Nervous | 1 (1.1%) | – | – | 1 (1.4%) | **4 (6.5%)** | – |
| MH2 | Down in dumps | 3 (3.4%) | – | – | 2 (2.7%) | 1 (1.6%) | 1 (1.6%) |
| MH3 | Calm and peaceful | 1 (1.1%) | – | – | **4 (5.4%)** | 1 (1.6%) | – |
| MH4 | Blue/sad | 2 (2.2%) | – | – | **4 (5.4%)** | 2 (3.2%) | – |
| MH5 | Happy | **4 (4.5%)** | – | – | **5 (6.8%)** | – | **3 (4.8%)** |

*Scales*: PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

Bold, more than 4%.

*, more than 8%.

a Refers to the interviews where research assistants made notes of items that needed to be explained (or where other comments of patients were made that are not included in this table). In the column only the items were included where research assistants explained an item and the item is a non-missing.

b Random answer = research assistant had the impression that the patient did not fully understand the item and gave a wrong/randomly response option.

representing a 71% and 65% response rate, respectively. Reasons for not participating were similar to those for the first assessment.

### 3.2. *Descriptive statistics and qualitative results (feasibility)*

The majority of both Turkish (71%) and Moroccan (87%) patients completed the SF-36 in an interview format (Table 1). The average time required to complete the questionnaire was 19.8 min and 17.6 min for the Moroccan and Turkish samples, respectively (range 5–55 min). One Turkish and five Moroccan patients did not complete the questionnaire (i.e. more than half of the items) because they did not understand the response categories (4), felt too ill (1) or were not motivated (1). These patients were all older than 55 years and illiterate. These cases were excluded from further analysis.

On average, 3.2% (range 0–9.0%) of the individual questionnaire items were missing in the Turkish sample and 4.4% (range 0–12.2%) in the Moroccan sample (Table 2). In the Moroccan sample, a higher number of items required some explanation by the research assistant. The items that had more than 8% missing or required explanation were RP1 in the Turkish group and PF1, PF3, RP3 and RE3 in the Moroccan group. The single Standard-Arabic loanword 'limit' (used throughout the Moroccan-Arabic version, but first met in PF1) often required explanation. The research assistants also observed that some (illiterate) patients had difficulties in discerning distances (PF7–PF9). The full range of scores was observed for the 8 scales with the exception of the GH and MH scales in both groups. Relatively high ceiling or floor effects were found for the RP and RE scales in both samples (Table 3).

### 3.3. *Internal consistency reliability*

Internal consistency reliability estimates (Cronbach's $\alpha$) for the eight SF-36 scales were above 0.70, with the exception of the VT scale in the Moroccan sample, the SF scale in the Turkish sample, and the GH scale in the Turkish and Moroccan samples (Table 3). For the two Moroccan language groups, most scales met the 0.70 criterion, with the exception of the GH scale in Moroccan-Arabic language group and the VT scale in the Tarifit language group.

### 3.4. *Multitrait scaling analyses*

Almost all items exceeded the 0.4 criterion for convergent validity on all scales. The exceptions were items MH3, GH2 and GH4 in the Turkish sample and items VT1, GH2, GH3 and GH4 in the Moroccan sample (Table 4). Using the one standard error criterion, the large majority of scales had more than 90% scaling success in both samples, with the exception of the GH scale (71.4%) and VT scale (85.7%) in the Moroccan sample. When employing the more stringent two standard criterion, item discriminant validity was 100% successful only for the RE scale in the Moroccan group, and for none of the scales in the Turkish group. In general, more problems with item-discriminant validity were observed in the Moroccan group than in the Turkish group.

**Table 3 – Median, mean, standard deviation, percentage floor and ceiling and Cronbach's $\alpha$ of the SF-36 scales**

| | Turkish (N = 89) | | | | | | Moroccan (N = 74) | | | | | | M-Arabic (N = 47) | Tarifit (N = 27) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | Median | Mean | SD | % Floor | % Ceiling | Cronbach's $\alpha$ (N) | N | Median | Mean | SD | % Floor | % Ceiling | Cronbach's $\alpha$ (N) | Cronbach's $\alpha$ (N) | Cronbach's $\alpha$ (N) |
| PF | 89 | 50.0 | 51.9 | 26.2 | 1.1 | 10.1 | 0.89 (80) | 73 | 60.0 | 58.5 | 30.1 | 1.4 | 11.0 | 0.94 (60) | 0.94 (41) | 0.92 (19) |
| RP | 88 | 0 | 31.4 | 40.1 | 54.5 | 18.2 | 0.88 (79) | 72 | 50.0 | 49.7 | 39.2 | 26.4 | 27.8 | 0.80 (63) | 0.81 (40) | 0.78 (23) |
| BP | 89 | 51.0 | 50.0 | 27.7 | 3.4 | 10.1 | 0.88 (88) | 73 | 51.0 | 54.5 | 31.6 | 8.2 | 21.9 | 0.91 (72) | 0.93 (45) | 0.86 (27) |
| GH | 86 | 48.5 | 49.7 | 20.8 | 0 | 1.2 | 0.67 (75) | 73 | 45.0 | 44.8 | 21.1 | 1.4 | 0 | 0.67 (64) | 0.65 (41) | 0.72 (23) |
| VT | 88 | 48.3 | 43.1 | 23.8 | 6.8 | 2.3 | 0.75 (80) | 72 | 50.0 | 49.0 | 24.2 | 1.4 | 6.9 | 0.66 (67) | 0.81 (42) | 0.40 (25) |
| SF | 87 | 62.5 | 57.9 | 28.9 | 2.3 | 18.4 | 0.67 (81) | 72 | 75.0 | 67.5 | 31.7 | 6.9 | 30.6 | 0.81 (67) | 0.75 (41) | 0.91 (26) |
| RE | 86 | 33.3 | 44.8 | 42.0 | 40.7 | 26.7 | 0.80 (82) | 71 | 66.7 | 55.9 | 45.6 | 35.2 | 46.5 | 0.91 (66) | 0.91 (40) | 0.90 (26) |
| MH | 88 | 52.0 | 55.0 | 21.5 | 0 | 2.3 | 0.78 (82) | 72 | 60.0 | 58.5 | 25.2 | 0 | 6.9 | 0.83 (66) | 0.86 (42) | 0.79 (24) |

*Scales:* PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

**Table 4 – Multitrait scaling analyses of SF-36 in Turkish (N = 81[a]) and Moroccan (N = 67[a]) cancer patients**

| Scale | | Item convergent validity | Item convergent validity scaling succes[b] | Item discriminant validity[c] | Item discriminant validity scaling succes[d] |
|---|---|---|---|---|---|
| PF | Turkish | 0.49–0.73 | 10/10 | 81.4 (57/70) | 100 (70/70) |
| PF | Moroccan | 0.67–0.80 | 10/10 | 77.1 (54/70) | 100 (70/70) |
| RP | Turkish | 0.71–0.84 | 4/4 | 82.1 (23/28) | 100 (28/28) |
| RP | Moroccan | 0.53–0.65 | 4/4 | 25.0 (7/28) | 100 (28/28) |
| BP | Turkish | 0.74–0.74 | 2/2 | 71.4 (10/14) | 100 (14/14) |
| BP | Moroccan | 0.75–0.75 | 2/2 | 35.7 (5/14) | 100 (14/14) |
| GH | Turkish | 0.29–0.57 | 3/5 | 45.7 (16/35) | 94.3 (33/35) |
| GH | Moroccan | 0.18–0.54 | 2/5 | 14.3 (5/35) | 71.4 (25/35) |
| VT | Turkish | 0.46–0.62 | 4/4 | 53.6 (15/28) | 100 (28/28) |
| VT | Moroccan | 0.34–0.55 | 3/4 | 7.1 (2/28) | 85.7 (24/28) |
| SF | Turkish | 0.53–0.53 | 2/2 | 28.6 (4/14) | 92.9 (13/14) |
| SF | Moroccan | 0.69–0.69 | 2/2 | 50.0 (7/14) | 100 (14/14) |
| RE | Turkish | 0.53–0.77 | 3/3 | 76.2 (16/21) | 100 (21/21) |
| RE | Moroccan | 0.80–0.82 | 3/3 | 100 (21/21) | 100 (21/21) |
| MH | Turkish | 0.35–0.67 | 4/5 | 65.7 (23/35) | 94.3 (33/35) |
| MH | Moroccan | 0.47–0.69 | 5/5 | 54.3 (19/35) | 91.4 (32/35) |

a N = number of patients that completed all scales (=more than half of the items for each scale).
b Item convergent validity scaling success = number of item-scale correlations greater than 0.40/total number of item-scale correlations (corrected for overlap).
c Item discriminant validity scaling success = percentage of items with scaling success (number of correlations of items with own scales significantly higher (⩾2SD) than correlations with other scales/total number of correlations).
d Item discriminant validity scaling success = percentage of items with scaling success (number of correlations of items with own scales significantly higher (⩾1SD) than correlations with other scales/total number of correlations).

### 3.5. *Interscale correlations*

All interscale correlations were lower than the internal consistency estimates of their own scales, indicating that each scale was assessing a relatively unique concept (not tabled).

### 3.6. *Known groups validity*

The strongest and most consistent evidence of known groups validity was found when using KPS (in both the Turkish and Moroccan samples) and comorbidity (in the Moroccan sample only) as grouping variables (Table 5). Contrary to expectations, few statistically significant differences were observed in SF-36 scores as a function of disease stage or treatment status.

### 3.7. *Responsiveness*

Statistically significant differences in the expected direction were observed as a function of changes in KPS scores (worsened, stable or improved) for all but the MH scale in the

**Table 5 – Known groups validity of SF-36 scales**

| | KPS[a] | | Comorbidity[b] | | Stage of disease[c] | | Status of treatment[d] | |
|---|---|---|---|---|---|---|---|---|
| | T | M | T | M | T | M | T | M |
| PF | ● | ● | ◒ | ● | ○ | ○ | ◒ | ○ |
| RP | ● | ● | ○ | ● | ● | ○ | ○ | ◒ |
| BP | ● | ● | ○ | ● | ◒ | ○ | ○ | ○ |
| GH | ● | ● | ◒ | ● | ○ | ○ | ○ | ○ |
| VT | ● | ● | ◒ | ● | ○ | ○ | ○ | ○ |
| SF | ● | ● | ○ | ◒ | ● | ○ | ○ | ○ |
| RE | ● | ● | ○ | ○ | ○ | ○ | ○ | ● |
| MH | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ |

T, Turkish; M, Moroccan
●, Statistically significant difference (<0.05) between the known groups.
◒, *P*-value between 0.05 and 0.10.
○, No statistically significant difference between known groups.
a KPS groups: score 30–70 (T: *n* = 41; M: *n* = 33) *versus* score 80–100 (T: *n* = 47 and M: *n* = 38).
b Comorbidity groups: no comorbidity (T: *n* = 28; M: *n* = 23) *versus* 1 or more comorbidities (T: *n* = 61; M: *n* = 51).
c Stage of disease: local/locoregional (T: *n* = 69; M: *n* = 60) *versus* metastatic (T: *n* = 20; M: *n* = 11).
d Status of treatment: under control (T: *n* = 63; M: *n* = 58) *versus* under treatment (T: *n* = 25; M = 13).

Turkish sample. In the Moroccan sample, the only statistically significant differences in the expected direction were found for changes over time in the RP and SF scales (Table 6).

### 3.8. *DIF analyses*

In the Turkish sample, the interaction term between translation and scale score was significant ($P < 0.001$) for two items (MH3 and SF2), indicating non-uniform DIF (Table 7). Uniform DIF was observed for 14 items in 7 scales in the Turkish sample and 21 items in 8 scales in the Moroccan sample. The GH, RE and SF scales appeared to be most susceptible to DIF in both samples, and the MH scale in the Turkish sample. Very high odds ratios were found in the Turkish sample for items PF1, RE3 and MH2. In the Moroccan sample, high odds ratios were found for items PF1, PF2 and VT4.

## 4. Discussion

In this paper, we have reported the results of a psychometric study of the SF-36 when employed among Turkish and Moroccan cancer patients in The Netherlands. Approximately three-quarters of the Turkish patients and two-thirds of the Moroccan patients did not speak Dutch or indicated that they had low levels of proficiency, thus supporting the need for translated versions of the SF-36. The average time required to complete the SF-36 was about twice that reported in other studies of this questionnaire.[3,17,18] In part, this reflects the fact that the questionnaire was administered orally to respondents who were often illiterate. However, the time required for administration was significantly longer than for the EORTC QLQ-C30,[19] probably reflecting the fact that the SF-36 has longer item stems and varies the response categories used throughout the questionnaire. The number of missing responses for individual items was comparable with other studies using paper-and-pencil administration [3,15,17,18,20–24], but higher compared to studies using interview or telephone administration.[21]

Overall, the scaling structure, reliability and validity of both the Turkish and Moroccan versions of the SF-36 were satisfactory. The GH and VT scales exhibited the most problems. Problems with the GH scale were also observed in a previous study of the SF-36 among Dutch cancer patients , where it was suggested that such general health perception questions may be less appropriate for use among cancer patients.[15] In the current study, some patients found it difficult to respond to questions about future health, indicating instead that 'only God knows'. This attitude has also been reported in a study of the Lebanese version of the SF-36.[18] The very low reliability of the VT scale in the Tarifit subgroup needs further investigation.

The differences in SF-36 scores as a function of performance status and, to a lesser extent, comorbidity, supported the known groups validity of the questionnaire. The fact that SF-36 scores did not vary significantly as a function of disease stage or treatment status may be attributed, at least in part, to

**Table 6 – Responsiveness analysis with ANOVA comparing groups with better, stable or worse performance status between 1st and 2nd assessment as measured with KPS score**

| Scale | KPS change | Turkish | | | | Moroccan | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | N | Mean T2-T1 | SD T2-T1 | P-value | N | Mean T2-T1 | SD T2-T1 | P-value |
| PF | Worse | 17 | –11.7 | 19.9 | 0.000 | 12 | –10.0 | 17.7 | 0.103 |
| | Stable | 23 | –10.7 | 17.5 | | 19 | 4.0 | 17.3 | |
| | Better | 17 | 13.5 | 17.9 | | 11 | 2.5 | 19.5 | |
| RP | Worse | 17 | –29.4 | 50.2 | 0.003 | 12 | –30.6 | 31.8 | 0.001 |
| | Stable | 23 | –4.0 | 44.4 | | 19 | 3.5 | 37.0 | |
| | Better | 17 | 26.5 | 41.9 | | 12 | 27.1 | 30.0 | |
| BP | Worse | 17 | –13.5 | 36.8 | 0.000 | 12 | 1.7 | 29.5 | 0.295 |
| | Stable | 23 | –15.6 | 21.7 | | 19 | 5.2 | 22.4 | |
| | Better | 17 | 19.8 | 23.2 | | 12 | 17.0 | 24.7 | |
| GH | Worse | 17 | –10.8 | 15.9 | 0.018 | 12 | –0.8 | 16.4 | 0.214 |
| | Stable | 22 | –5.0 | 12.4 | | 19 | –5.2 | 12.5 | |
| | Better | 16 | 7.0 | 24.4 | | 12 | 5.5 | 20.6 | |
| VT | Worse | 17 | .0 | 26.1 | 0.009 | 12 | –7.9 | 15.6 | 0.074 |
| | Stable | 22 | –5.5 | 18.5 | | 19 | –5.8 | 20.6 | |
| | Better | 16 | 17.1 | 21.3 | | 12 | 7.4 | 13.7 | |
| SF | Worse | 17 | –11.8 | 44.3 | 0.048 | 12 | –7.3 | 18.8 | 0.022 |
| | Stable | 22 | –2.8 | 23.8 | | 19 | –2.0 | 17.3 | |
| | Better | 17 | 18.4 | 39.1 | | 12 | 14.6 | 23.1 | |
| RE | Worse | 17 | –41.2 | 46.4 | 0.006 | 12 | .0 | 37.6 | 0.739 |
| | Stable | 22 | 1.5 | 31.7 | | 19 | 10.5 | 43.1 | |
| | Better | 15 | 6.7 | 59.4 | | 12 | 1.4 | 41.7 | |
| MH | Worse | 17 | –4.9 | 19.3 | 0.093 | 12 | –6.7 | 11.1 | 0.054 |
| | Stable | 22 | –.2 | 17.2 | | 19 | –0.8 | 15.7 | |
| | Better | 16 | 9.3 | 19.7 | | 12 | 9.3 | 19.8 | |

**Table 7 – Results of uniform and non-uniform DIF by ordinal regression analysis: odds, (and confidence interval) and P-values of the Dutch cancer sample versus the Turkish or Moroccan immigrant sample (corrected for age, sex and stage of disease)**

| | Turkish | | Moroccan | |
|---|---|---|---|---|
| | Odds ratio | *P*-value | Odds ratio | *P*-value |
| *PF scale* | | | | |
| PF1 | **10.22 (4.83–21.59)** | <0.001 | **9.56 (4.57–19.99)** | <0.001 |
| PF2 | **1.93** (1.00–3.72) | 0.05 | **6.95 (3.31–14.60)** | <0.001 |
| PF3 | **0.41** (0.22–0.77) | <0.01 | **0.51** (0.26–1.02) | 0.06 |
| PF4 | 0.92 (0.49–1.72) | 0.80 | 0.75 (0.39–1.44) | 0.39 |
| PF5 | 0.94 (0.48–1.85) | 0.86 | 1.15 (0.52–2.51) | 0.73 |
| PF6 | 0.82 (0.45–1.49) | 0.52 | **0.33** (0.17–0.65) | 0.001 |
| PF7 | 1.77 (0.91–3.43) | 0.09 | 0.94 (0.46–1.91) | 0.86 |
| PF8 | 1.04 (0.51–2.09) | 0.92 | 0.76 (0.34–1.69) | 0.50 |
| PF9 | 1.02 (0.48–2.16) | 0.96 | **0.31** (0.23–1.23) | 0.14 |
| PF10 | **0.30** (0.14–0.63) | <0.01 | **0.31** (0.13–0.75) | <0.01 |
| *RP scale* | | | | |
| RP1 | 1.24 (0.46–3.31) | 0.67 | 0.53 (0.19–1.46) | 0.22 |
| RP2 | **0.46** (0.15–1.46) | 0.19 | **0.46** (0.17–1.24) | 0.12 |
| RP3 | 1.30 (0.40–4.23) | 0.66 | **2.40** (0.81–7.09) | 0.11 |
| RP4 | 0.62 (0.20–1.90) | 0.40 | **2.09** (0.72–6.05) | 0.18 |
| *BP scale* | | | | |
| BP1 | 1.28 (0.64-2.58) | 0.49 | 0.69 (0.32-1.51) | 0.35 |
| BP2 | 1.24 (0.68-2.27) | 0.48 | **2.51** (1.26-4.98) | <0.01 |
| *GH scale* | | | | |
| GH1 | 0.64 (0.38–1.07) | 0.09 | **0.48** (0.27–0.85) | 0.01 |
| GH2 | **0.36** (0.23–0.59) | <0.001 | **0.50** (0.30–0.83) | <0.01 |
| GH3 | **1.96** (1.21–3.18) | <0.01 | **2.68** (1.58–4.53) | <0.001 |
| GH4 | 1.50 (0.92–2.45) | 0.11 | 0.62 (0.37–1.05) | 0.08 |
| GH5 | **2.97** (1.79–4.92) | <0.001 | **2.61** (1.48–4.61) | 0.001 |
| *VT scale* | | | | |
| VT1 | 0.85 (0.52–1.39) | 0.51 | **0.41** (0.24–0.69) | 0.001 |
| VT2 | **2.42** (1.44–4.08) | 0.001 | 0.63 (0.36–1.11) | 0.119 |
| VT3 | 0.61 (0.37–1.03) | 0.06 | 1.82 (1.04–3.20) | 0.04 |
| VT4 | 1.03 (0.61–1.72) | 0.92 | **4.69 (2.63–8.38)** | <0.001 |
| *SF scale* | | | | |
| SF1 | **0.49** (0.27-0.89) | 0.02 | **0.40** (0.20-0.81) | 0.01 |
| SF2 | *0.95 (0.93–0.97)* | <0.001* | **2.76** (1.33–5.74) | <0.01 |
| *RE scale* | | | | |
| RE1 | **6.75** (1.64–27.69) | <0.01 | 0.62 (0.12–3.23) | 0.57 |
| RE2 | 1.15 (0.32–4.10) | 0.83 | **12.41** (1.88–81.75) | <0.01 |
| RE3 | **0.10 (0.03–0.28)** | <0.001 | **0.13** (0.03–0.60) | <0.01 |
| *MH scale* | | | | |
| MH1 | 0.68 (0.41–1.11) | 0.12 | 1.24 (0.71–2.17) | 0.44 |
| MH2 | **0.28 (0.16–0.50)** | <0.001 | 1.60 (0.86–2.99) | 0.14 |
| MH3 | *0.96 (0.94–0.98)* | <0.001* | 1.19 (0.68–2.09) | 0.54 |
| MH4 | 0.60 (0.36–1.01) | 0.06 | **1.98** (1.10–3.56) | 0.02 |
| MH5 | **2.76** (1.64–4.63) | <0.001 | 0.67 (0.38–1.16) | 0.15 |

Uniform DIF for language was considered to be present if the odds ratio is outside the interval of 0.53–1.89 and is presented in bold. Non-uniform DIF was considered to be present if the interaction of language with the total score was found to be statistically significant ($P < 0.001$) and is presented in the table with an * and in italic.

the fact that the study sample included relatively few patients with metastatic disease or patients under active treatment.

The SF-36 demonstrated better responsiveness to change over time in the Turkish sample than the Moroccan sample. This might be due to the relatively higher percentage of patients with an unchanged performance status in the latter sample.

The SF-36 exhibited DIF (uniform and non-uniform) in all scales in the Moroccan group and in almost all scales in the Turkish group, and for 40% of the items in the Turkish group and 60% of the items in the Moroccan group. Clear examples of DIF were found, for example, for item PF1 (vigorous activities) and item PF2 (moderate activities), where Dutch patients tended to score lower (report more limitations) than Turkish and Moroccan patients. This may be due to cultural differences in the appropriateness of the examples used to describe activity levels. Activities such as participating in strenuous sports, pushing a vacuum and cycling are probably more appropriate for Dutch than for Turkish or Moroccan respondents. Therefore, Turks and Moroccans may be less likely to indicate problems with these activities simply because they tend not to perform them anyway.

In conclusion, the two Moroccan (Tarifit and Moroccan-Arabic) and the adapted Turkish versions of the SF-36 exhibit generally satisfactory psychometric properties and thus are promising for use among these ethnic minority cancer patient populations in The Netherlands. We would recommend additional studies with larger samples of patients under active treatment and with longer follow-up in order to better understand the responsiveness of these versions of the SF-36 to changes in clinical status over time. We would also encourage additional larger quantitative investigations of DIF, as well as smaller qualitative studies using cognitive debriefing techniques to better understand potential differences in responses to the SF-36 as a function of language and culture. Finally, additional studies are needed to confirm the psychometrics of the questionnaire when used among these ethnic minority groups residing in other Western European countries.

## Conflict of interest statement

None declared.

## Acknowledgements

This study was supported by Grant No. NKI 99-1724 from the Dutch Cancer Society. The authors thank the following individuals for their help in recruiting patients into the study: J.H. Schornagel, Department of Internal Medicine, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam; G. van Andel Department of Urology, Onze Lieve Vrouwe Gasthuis, Amsterdam; C.E.E. Koning, Department of Radiotherapy, Medical Center Haaglanden, Den Haag; M.J.H. Schweitzer, Department of Internal Medicine, Sint Lucas Andreas Hospital, Amsterdam; G. Brutel de la Rivière, Department of Pathology, Sint Lucas Andreas Hospital, Amsterdam; J.A. Langendijk, Department of Radiotherapy, Vrije Universiteit Medical Center, Amsterdam; P.C. Levendag, Department of Radiotherapy, Erasmus Medical Center, Rotterdam; G. Dahmen, Department of Patient Registration, Erasmus Medical Center, Rotterdam; A.J. Gelderblom, Department of Clinical Oncology, Leiden University Medical Center, Leiden. We thank

Dr. D.L. Knol, EMGO Institute/Department of Clinical Epidemiology and Biostatistics of VU University Medical Centre, for assistance on the DIF analyses. Special thanks to the research assistants, Ş. Çelik, N. Gündogan, F. El Haddouchi, M. Idrissi and L. Tahiri, for carrying out the data collection, and all of the patients for their willingness to participate in the study.

## REFERENCES

1. Salt J. European International Migration: Evaluations of the Current Situation. Strasbourg, Council of Europe; 2002.
2. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36) I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
3. Pinar R. Reliability and construct validity of the SF-36 in Turkish cancer patients. *Qual Life Res* 2005;**14**:259–64.
4. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey manual and interpretation guide*. Boston (MA): New England Medical Center, The Health Institute; 1993.
5. Ware JE, Kosinski M, Keller SD. *SF-36 physical and mental health summary scales – a users' manual*. Boston (MA): New England Medical Center, the Health Institute; 1994.
6. Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. *J Clin Epidemiol* 1998;**51**:913–23.
7. Ware JE, Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. *Int J Technol Assess Health Care* 1995;**11**:525–51.
8. Meloen JD, Veenman J. Het is maar de vraag.. Onderzoek naar responseeffecten bij minderhedensurveys. Lelystad, The Netherlands, Koninklijke Vermande BV; 1990.
9. Curvers J. Met ongeletterde ogen. Kennis van taal en schrift van analfabeten. Amsterdam, Aksant; 2002.
10. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod A, editor. *Evaluation of Chemotherapeutic Agents*. New York: Colombia University; 1949. p. 199–205.
11. Schaafsma J, Osoba D. The Karnofsky Performance Status Scale re-examined: a cross-validation with the EORTC-C30. *Qual Life Res* 1994;**3**:413–24.
12. Schag CC, Heinrich RL, Ganz PA. Karnofsky performance status revisited: reliability, validity, and guidelines. *J Clin Oncol* 1984;**2**:187–93.
13. Ware JE, Harris WJ, Gandek B, Rogers BW, Reese PR. *MAP-R for Windows: multitrait multi-item analysis program-revised user's guide*. Boston (MA): Health Assessment Lab; 1997.
14. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;**16**:297–334.
15. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998;**51**:1055–68.
16. Petersen MA, Groenvold M, Bjorner JB, et al. Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Qual Life Res* 2003;**12**:373–85.
17. Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalisation. *J Epidemiol Community Health* 2003;**57**:259–63.
18. Sabbah I, Drouby N, Sabbah S, Retel-Rude N, Mercier M. Quality of Life in rural and urban populations in Lebanon using SF-36 Health Survey. *Health Qual Life Outcomes* 2003;**6**:30.
19. Hoopman R, Muller MJ, Terwee CB, Aaronson NK. Translation and validation of the EORTC QLQ-C30 for use among Turkish and Moroccan ethnic minority cancer patients in The Netherlands. Eur J Cancer 2006 Jun 5 (epub ahead of print).
20. Bjorner JB, Damsgaard MT, Watt T, Groenvold M. Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. *J Clin Epidemiol* 1998;**51**:1001–11.
21. Gandek B, Ware Jr JE, Aaronson NK, et al. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998;**51**:1149–58.
22. McHorney CA, Ware Jr JE, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;**32**:40–66.
23. Sullivan M, Karlsson J, Ware Jr JE. The Swedish SF-36 Health Survey–I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Soc Sci Med* 1995;**41**:1349–58.
24. Wagner AK, Wyss K, Gandek B, et al. A Kiswahili version of the SF-36 Health Survey for use in Tanzania: translation and tests of scaling assumptions. *Qual Life Res* 1999;**8**:101–10.