UvA-DARE (Digital Academic Repository)

Evaluation of the psychometric properties of the SF-36 health survey for use among Turkish and Moroccan ethnic minority populations in the Netherlands

Hoopman, R.; Terwee, C.B.; Devillé, W.; Knol, D.L.; Aaronson, N.K.

Link to publication

Citation for published version (APA):

Evaluation of the psychometric properties of the SF-36 health survey for use among Turkish and Moroccan ethnic minority populations in the Netherlands

Rianne Hoopman · Caroline B. Terwee · Walter Devillé · Dirk L. Knol · Neil K. Aaronson

Abstract

Purpose To evaluate the psychometrics of the SF-36 Health Survey among Turkish and Moroccan ethnic minority populations in the Netherlands and to compare the results to those based on the indigenous Dutch population. *Methods* Data were derived from the Second Dutch National Survey of General Practice. In total, SF-36 data were available for 409 Turkish, 377 Moroccan, and 9,628 Dutch respondents. Language subgroup analyses were performed for the Turkish ($n = 162$) and Moroccan-Arabic ($n = 52$) versions of the SF-36. Psychometric properties were evaluated by descriptive statistics, factor analysis, internal consistency estimates, known-group comparisons, and differential item function (DIF) analysis. *Results* Low levels of missing SF-36 data were observed across all groups. Confirmatory factor analysis supported the hypothesized SF-36 scale structure for the Dutch and Turkish group, but high correlations between the MH and VT factors were observed in the Moroccan group. All Cronbach's alpha coefficients were above 0.70, except for the Vitality scale in the Moroccan group and the Vitality and Social Functioning scales in the Moroccan-Arabic language subgroup. Known-groups validity was found across samples using age, sex, education, and comorbidity as grouping variables, but not marital status. Some evidence for DIF was found in both ethnic group samples. *Conclusions* The results generally support the use of the SF-36 for general population research among Turkish and Moroccan ethnic minorities in the Netherlands. Additional studies are needed to confirm the psychometrics of the questionnaire when used among these populations in other Western European countries.

Keywords SF-36 Health survey · Health status assessment · Cross-cultural · Moroccan · Turkish · Ethnic minority · Psychometrics · Factor analysis · DIF

R. Hoopman · N. K. Aaronson (✉)
Division of Psychosocial Research and Epidemiology,
The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam,
The Netherlands
e-mail: n.aaronson@nki.nl

R. Hoopman · C. B. Terwee · D. L. Knol
Department of Epidemiology and Biostatistics and the EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands

W. Devillé
Netherlands Institute for Health Services Research (NIVEL), Utrecht, The Netherlands

Introduction

The Medical Outcomes Study SF-36-Item Short Form Health Survey (SF-36) is one of the most widely used health status questionnaires worldwide [1]. The SF-36 was originally developed in the United States in the late 1980s for a longitudinal investigation into the self-reported health status of general populations and patients with a wide range of chronic conditions [2]. Interest in the SF-36 has grown steadily, and it has been translated and psychometrically evaluated in more than 50 languages (www.sf-36.org), including non-western cultures and ethnic minorities within countries.

There are currently approximately 2.8 million Turkish and 1.2 million Moroccan immigrants living in Western Europe [3]. In the Netherlands, approximately 10% of the

total population (1.8 million) is of non-western immigrant origin, of which Turkish and Moroccan immigrants constitute (together with Suriname and Antillean/Aruban immigrants) the major groups. In 2008, 2.3% of the total population was of Turkish descent (i.e., at least one parent born in Turkey) and 2.0% of Moroccan descent. Of these groups, 52 and 55%, respectively, were not born in the Netherlands and are first generation immigrants. It is expected that the number of non-western immigrants to the Netherlands will continue to increase reaching 2.7 million in 2050 [4].

Both Turkish and Moroccan immigrant groups, while clearly having some distinct cultural characteristics, come from an Islamic background, tend to be poorly educated, and come from primarily rural areas of their home country. They are of particular interest from the perspective of cultural adaptation and translation of health-related quality of life (HRQL) questionnaires in that, in comparison with other immigrant groups from former Dutch colonies (Suriname and the Antilles), they typically have no background in or exposure to the Dutch culture or language prior to their arrival in the Netherlands.

Interest in assessing the self-reported health status and HRQOL of Turkish and Moroccan immigrants is increasing. Not only are these ethnic minority populations growing, but also the first generation of immigrants is now reaching the age at which the incidence of chronic disease rises sharply. In 2003, there were 26,164 Turkish and 22,953 Moroccans over the age of 55 living in the Netherlands. These numbers are expected to rise to about 46,017 and 40,666, respectively, by 2015 [5].

The inclusion of Turkish and Moroccan immigrants in HRQOL research, and particularly those of the first generation, is hampered by two primary factors. First, the majority of first generation Turkish and Moroccan immigrants do not speak the language of their host countries. As a consequence, they are not able to complete HRQOL questionnaires in the available Western European translations. In the Netherlands, 60% of the Turkish and Moroccan immigrants above 55 years of age have difficulties speaking Dutch [6]. Second, there is a high illiteracy rate among these first generation cohorts. In the Netherlands, 70% of the Turkish and 90% of the Moroccan immigrants above 55 years of age have had no formal education [6]. These factors need be taken into consideration when choosing the language (i.e., oral vs. official, written) and mode of administration (i.e., oral vs. written) of HRQOL measures to be used in these populations.

In 2006, the Turkish version of the SF-36 and two translated versions in oral Moroccan languages (Moroccan-Arabic and Tarifit, in phonetic Arabic and Latin script respectively) were culturally adapted and psychometrically evaluated for the use among Turkish and Moroccan ethnic minority patients with cancer in the Netherlands [7]. In this study, the original scale structure of the SF-36 and the internal consistency reliability were confirmed. 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 between the Turkish and Moroccan translations, and the Dutch version of the SF-36. The study supported the use of the SF-36 in both ethnic groups of patients with cancer. However, these translated versions have not yet been assessed and evaluated among the general (healthy) Turkish and Moroccan population in the Netherlands.

The aim of this study was to psychometrically evaluate the Dutch, Moroccan-Arabic, and Turkish versions of the SF-36 [7, 8] for use among general Turkish and Moroccan ethnic minority populations in the Netherlands. The psychometric results based on these ethnic minority groups were also compared with those based on data from the indigenous Dutch general population.

Methods

The SF-36

The SF-36 (version 1.0) is composed of 36 questions with standardized response options. The questions are organized into eight multi-item scales: physical functioning (PF), role limitations due to physical problems (RP), bodily pain (BP), general health perceptions (GH), vitality (VT), social functioning (SF), role limitations due to emotional problems (RE), and general mental health (MH). One additional item assesses health transition (HT). Response categories are of the Likert-type, ranging from two to six categories. Both a 'standard' version (employing a 4 week time frame) and 'acute' version (employing a 1 week time frame) are available [1, 2, 9]. In this study, the 'standard' version was used.

The Dutch and Moroccan-Arabic versions of the SF-36 were translated from the original English version following standard SF-36 guidelines [7, 8, 10, 11]. The Turkish version, originally developed in Turkey [12], was linguistically and culturally adapted for use among Turkish immigrants living in the Netherlands [7]. The translation into Moroccan-Arabic, an oral language, was done in phonetic Arabic script to make it appropriate for interviewer administration [7].[1]

---

[1] The oral Moroccan language *Tarifit* [7] was not used in this study.

Data collection

The data were collected in 2001 as part of the Second Dutch National Survey of General Practice (DNSGP2), which is part of a longitudinal, nationwide health monitoring system [13]. The DNSGP2 study was carried out in 104 general practices (stratified by region, level of urbanization and practice type) throughout the Netherlands. All 385,461 patients registered in these practices were sent a one-page questionnaire in four languages that included sociodemographic information (age, sex, education, country of birth, etc.) and an informed consent form. A total of 294,999 patients returned the questionnaire (response rate = 76.5%). From this respondent pool, cohorts were formed from the indigenous Dutch general population and the four ethnic minority groups (Turkish, Moroccan, Surinamese, and Antillean) for the purpose of a 90-min health interview.

The self-reported, computer-assisted, home-based health interview was planned for an approximate 5% random sample of the indigenous Dutch cohort. In total, 12,699 of the 19,685 Dutch individuals who were invited to participate in the interview did so (response rate = 64.5%). Of these 12,699 individuals, 9,628 were 18 years of age or older and included in the current analysis.

From those who returned the initial study form (294,999), 11,540 were from one of the four ethnic minority groups, of whom 7,355 were 18 years of age or older. From both the Turkish and Moroccan groups, a sample of 1,200 individuals was drawn at random. Individuals from these random samples were invited to participate in the study until approximately 400 respondents per group had been recruited. In total, 862 Turkish immigrants and 848 Moroccan were invited, of whom 409 (response rate 47%) and 377 (response rate 44%) individuals, respectively, participated (Fig. 1).

These respondents could choose to undergo the health interview in a language for which a SF-36 translation was available: Dutch, Turkish, Moroccan-Arabic, or English. If the respondent was not proficient in one of these languages, but spoke another language such as Berber (for Moroccans) or Kurdish (for Turks), the interviewer translated the questionnaire in real time. All questionnaire data collection among the Turkish and Moroccan respondents was interview-based and was computer-administered by the interviewer.

In the DNSGP2 survey, to facilitate efficient administration of the SF-36, a "routing" algorithm was employed for the PF and the BP scales. Specifically, for the PF scale, if respondents reported no limitations for items PF1 (vigorous activities), PF4 (climbing several flights of stairs), PF7 (walking more than a kilometer), and PF8 (walking a few hundred meters), the items PF2 (moderate activities),

PF5 (climbing one flight of stairs), and PF9 (walking 100 m) were respectively skipped. In the analysis, these skipped items were coded as having the response 'no, not limited at all'. For the BP scale, respondents skipped the item BP2 (Did pain interfere with your normal social activities) if they had answered 'none' to item BP1 (how much bodily pain did you have in the past 4 weeks?). In these cases, for the purposes of analysis, the BP2 item was coded as 'not at all'. Finally, in the original SF-36, the SF2 item (physical or emotional problems interfering with social activities) has five response options. In the DNSGP2 survey, the response category 'most of the time' was split into two response options: 'often' and 'most of the time.' For the analysis, these two response categories were combined to parallel the original five-level response scale.

Statistical analyses

Group and subgroup analysis

The indigenous Dutch sample (n = 9,628) was analyzed as one group. For the ethnic minority groups, the analysis was performed at two levels. First, the groups were formed based on culture. The analyses were conducted for the total Turkish (n = 409) and Moroccan (n = 377) samples, regardless of the language in which the SF-36 was administered (Dutch, English or in one of the mother tongue languages). These results represent the validation of the SF-36 in these cultural groups and can be considered as representative for how SF-36 data will be collected in future studies of these groups in the Netherlands. Second, subgroup analyses were performed at the language level. The analyses were performed for the Turkish and Moroccan groups that had completed the Turkish (n = 162) or Moroccan-Arabic (n = 52) language versions of the SF-36. These results represent the validation of the Turkish and Moroccan language versions of the SF-36.

As the language of administration was not systematically recorded in the DNSGP2 study (2001), we formed the language subgroups (Turkish and Moroccan-Arabic) on the basis of two assumptions: (1) we assumed that language proficiency in Dutch might have increased over time, but not decreased. The language used in a follow-up survey, performed 4 years later, in 2005, was recorded. All respondents who chose the Turkish or Moroccan language versions in 2005 were included in the language subgroups, but respondents who chose the Dutch SF-36 language version in 2005 were not included in the language subgroups, even though they might not have been proficient in Dutch in 2001. Based on this decision rule, 86 Turkish and 15 Moroccan respondents could be identified, who were assumed to have completed the SF-36 in Turkish or

Fig. 1 Overview of sampling and response of the Turkish and Moroccan groups
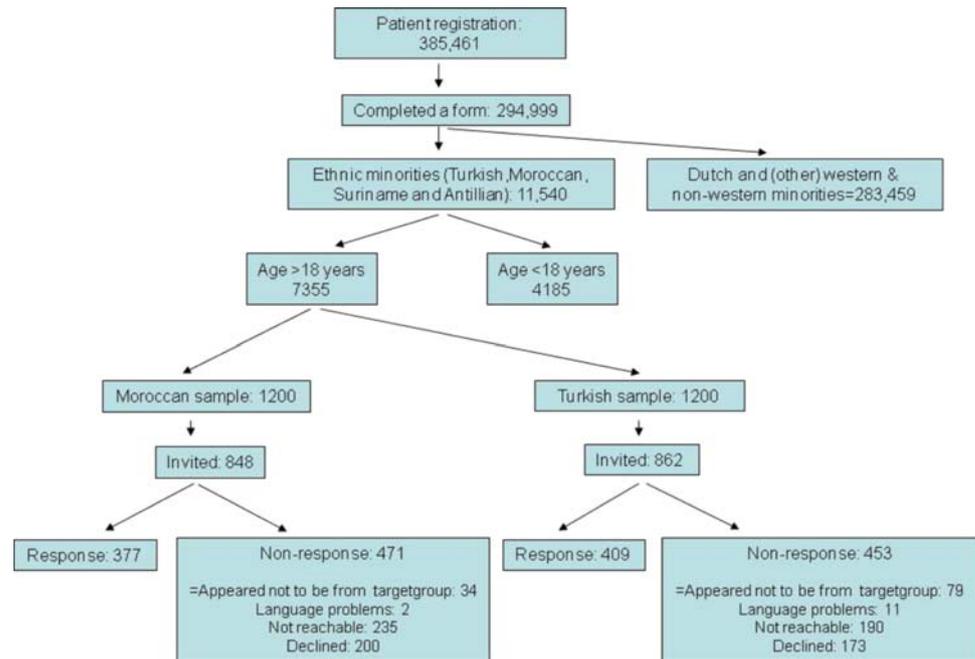
Moroccan-Arabic during the first survey; (2) respondents were assumed to have chosen to be interviewed in Turkish or Moroccan-Arabic if they indicated in the background questionnaire (in 2001) that they had no or only very limited proficiency in the Dutch language, that Turkish or Moroccan-Arabic was their native language and that they were proficient in that language.

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 favorable health outcomes respectively.

Descriptive statistics were generated to evaluate the missing questionnaires, missing items, score distributions, and floor and ceiling effects. Missing were accounted for according to the standard scoring algorithm of the SF-36. Scale scores were calculated on the basis of the mean scores of the available items. If 50% or more of the items were missing, the scale score was not calculated.

Confirmatory factor analyses for ordered categorical variables were performed in Mplus using the method of weighted least squares with mean and variance adjustment (WLSMV). The comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) were used as measures for model fit. A CFI and TLI of >0.95 and a RMSEA of <0.05 were considered as adequate fit.

Internal consistency reliability of the multi-item scales was assessed by Cronbach's coefficient α. A value of 0.70 or greater was considered as adequate for the purposes of group comparisons.

Known-groups validity was evaluated by comparing subgroups of patients known to differ on relevant background variables. Mean differences (i.e., the mean for the group where higher scores were expected minus the mean for the group where lower scores were expected) and effect sizes (i.e., mean difference divided by the standard deviation of the group where higher mean scores were expected) were calculated for all scales. In all groups, it was hypothesized that those respondents who were male, younger, married (or having a registered partnership), who had a higher level of education, and who had no or only one comorbid condition would have significantly higher SF-36 scores than those who were female, older, not married, lower educated and with two or more comorbid conditions. It was hypothesized that the observed differences in mean scores between groups formed on the basis of age and comorbidity would be larger for the physical health scales (PF, RP, BP, and GH scale) than for the mental scales (VT, SF, RE and MH) and that the mean differences between groups formed on the basis of marital status would be larger for the mental health scales. We also hypothesized that the largest mean differences would be observed on the basis of comorbidity grouping, and the smallest differences on the basis of marital status grouping. Finally, we hypothesized that the indigenous Dutch sample would generally score higher on the SF-36 than either the Turkish or the Moroccan samples.

Differential item functioning (DIF) was evaluated to test the equivalence of the outcome of the Turkish and Moroccan cultural (language) samples to that of the indigenous Dutch sample. We tested for both uniform and nonuniform DIF for all items from the eight SF-36 scales

using ordinal regression analysis [14]. We first tested for nonuniform DIF by modelling the item response as a logit-linear function of ethnic group (Dutch vs. Turkish, or Dutch vs. Moroccan), the scale score, and the interaction between ethnic group and scale score. The interaction term represents the possible nonuniform DIF. Nonuniform DIF (indicating that the magnitude and direction of ethnic group 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. Uniform DIF was tested by modeling the item response as a logit-linear function of the ethnic group and the scale score, with the translation term representing possible uniform DIF. Uniform DIF (testing the direction and magnitude of ethnic group differences in item scores) was considered to be present if the odds ratio of the ethnic group term was outside the interval 0.53–1.89 (log odds ratio β numerically larger than 0.64) [15]. All analyses were corrected for sex, age, and comorbidity.

Results

Sample background characteristics

Compared to the Dutch respondents, the total Turkish and Moroccan respondent groups were generally younger, had less formal education, and were more often married (Table 1). The Turkish and Moroccan-Arabic language subgroups consisted of relatively more women and elderly, and had lower education and more comorbid conditions compared to the total Turkish and Moroccan groups.

Descriptive statistics

Complete SF-36 data were available for all Turkish and Moroccan respondents. In the Dutch sample, 8 questionnaires could not be analyzed due to a large number of missing items. At the item level, on average, 0.22% (range 0–2.0%) of the individual questionnaire items was missing in the Turkish sample, 0.22% (range 0–3.7%) in the Moroccan sample, and 0.13% (range 0–0.21%) in the Dutch sample. The full range of scores was observed for the 8 SF-36 scales, with the exception of the GH and VT scales in the Moroccan sample (Table 2). Relatively high ceiling or floor effects were observed for the RE and RP scales in all three samples (Table 2).

Factor analysis

In the Dutch sample, the factor structure had an adequate fit (CFI 0.963, TLI 0.992, RMSEA 0.058). In the Turkish

sample, the model had a borderline fit (CFI 0.961, TLI 0.989, RMSEA 0.080). In the Moroccan sample, we first found a nonadmissible solution, i.e., nonpositive definiteness of the factor correlation matrix. There was also a negative residual variance for item RE1 and a high correlation of the factor vitality with mental health (0.98). After some restrictions (correlation between factor VT and MH fixed at 0.95 and error variance of item RE1 fixed at 0.05), the model fit was borderline (CFI 0.965, TLI 0.983., RMSEA 0.083). The only item that did not fit well was item VT1 [(estimated) loading 0.155] (Table 3). This item correlated with VT2, but not with the other items.

Internal consistency reliability

Cronbach's alpha coefficients for the eight SF-36 scales were above 0.70 for all samples, with the exception of the VT scale in the total Moroccan sample ($\alpha = 0.61$) and the VT ($\alpha = 0.54$) and the SF ($\alpha = 0.63$) scales in the Moroccan-Arabic language subgroup (Table 2).

Known-groups validity

A summary of the results of the known-groups validity analyses are presented in Table 4. The hypothesis that respondents who were male, younger, married, more highly educated, and having no or only one comorbid condition would report a better health status as assessed by the SF-36 (in total, 40 comparisons) was confirmed for 80% in the Turkish sample, 60% in the Moroccan sample, and 95% in the Dutch sample. The hypothesis that married respondents would have significantly higher SF-36 scores than unmarried respondents was not confirmed for any of the SF-36 scales in the Turkish or Moroccan samples.

The hypothesis that the mean differences between age and comorbidity subgroups would be larger for the physical health scales than for the mental health scales of the SF-36 (16 comparisons each) was confirmed in 75–100% of the cases in the three study samples. The hypothesis that the mean differences between the marital status groups would be larger for the mental health scales than for the physical health scales could not be confirmed in the Moroccan group.

The hypothesis that the largest observed mean differences would be between comorbidity subgroups and the smallest between marital status subgroups (32 comparisons) was confirmed in the large majority of cases (84–100%) in all three study samples.

Finally, as hypothesized, the Dutch sample reported better health outcomes than the total Turkish group (for all scales) and the total Moroccan group (except for the PF and the RP scales, where the differences were not statistically significant, data not presented in tabular form).

Table 1 Respondent background characteristics

	Turkish (N = 409)	Moroccan (N = 377)	Turkish language subgroup (N = 162)[a]	Moroccan-Arabic language subgroup (N = 52)[b]	Dutch (N = 9,628)
Gender					
Female	215 (53%)	196 (52%)	95 (59%)	34 (65%)	5,334 (55%)
Male	194 (47%)	181 (48%)	67 (41%)	18 (35%)	4,294 (45%)
Age mean (sd)	36.8 (12.1)	36.6 (12.6)	40.5 (12.9)	41.4 (12.0)	48.9 (17.0)
Age range	18–76	18–79	21–69	20–67	18–79
Age					
18–49	331 (81%)	316 (84%)	114 (70%)	39 (75%)	5,149 (54%)
50–79	77 (19%)	60 (16%)	47 (29%)	12 (23%)	4,478 (46%)
Missing	1 (0.2%)	1 (0.3%)	1 (0.6%)	1 (2%)	1 (0%)
Education					
Primary school or less	177 (43%)	166 (44%)	106 (65%)	41 (79%)	1,427 (15%)
Basic high school	209 (51%)	180 (48%)	53 (33%)	10 (19%)	6,152 (64%)
Advanced high school	23 (6%)	29 (8%)	3 (2%)	1 (2%)	2,035 (21%)
Missing		2 (0.5%)			14 (0.1%)
Marital status					
Not married	47 (11%)	60 (16%)	4 (2%)	3 (6%)	2,023 (21%)
Married/registered partnership	318 (78%)	289 (77%)	139 (86%)	41 (79%)	6,359 (66%)
Divorced	27 (7%)	16 (4%)	7 (4%)	4 (8%)	496 (5%)
Widowed	15 (4%)	7 (2%)	11 (7%)	2 (4%)	740 (8%)
Missing	2 (0.5%)	5 (1%)	1 (0.6%)	2 (4%)	10 (0.1%)
Comorbidity					
None	158 (39%)	165 (44%)	47 (29%)	18 (35%)	3,358 (35%)
One	83 (20%)	103 (27%)	36 (22%)	18 (35%)	2,669 (28%)
Two or more	168 (41%)	109 (29%)	79 (49%)	16 (31%)	3,601 (37%)
Country of birth					
The Netherlands	57 (14%	40 (11%)	7 (4%)	1 (2%)	
Turkey/Morocco/else	352 (86%)	337 (89%)	155 (96%	51 (98%)	
Reason of immigration					
Work/reunion with spouse	192 (47%)	180 (48%)	110 (68%)	38 (73%)	
Children came with parents	132 (32%)	117 (31%)	36 (22%)	10 (19%)	
Other	28 (7%)	38 (10%)	9 (6%)	3 (6%)	
Missing	57 (14%)	42 (11%)	7 (4%)	1 (2%)	
Mother language					
Turkish	344 (84%)		152 (94%)		
Kurdish	8 (2%)		1 (1%)		
Other Turkish	9 (2%)		2 (1%)		
Dutch	48 (12%)	62 (16%)	7 (4%)	1 (2%)	
Moroccan-Arabic		195 (52%)		43 (83%)	
Tarifit		40 (11%)		4 (8%)	
Tamazight		67 (18%)		4 (8%)	
Tachelit		6 (2%)		–	
Other Moroccan		7 (2%)		–	
Proficiency in speaking Dutch					
None	24 (6%)	18 (5%)	20 (12%)	10 (19%)	
Low	120 (29%)	88 (23%)	98 (61%)	34 (65%)	
Moderate	105 (26%)	114 (30%)	26 (16%)	4 (8%)	

Table 1 continued

	Turkish (N = 409)	Moroccan (N = 377)	Turkish language subgroup (N = 162)[a]	Moroccan-Arabic language subgroup (N = 52)[b]	Dutch (N = 9,628)
High	160 (39%)	156 (41%)	18 (11%)	4 (8%)	
Missing		1 (0.3%)			

Comorbidity = one or more of the following self-reported conditions: diabetes, stroke, myocardial infarction, myocardial disorder, cancer, migraine/severe headache, high blood pressure, vascular disorder, asthma/COPD, psoriasis, chronic eczema, dizziness/fall, intestine disorder, incontinence, back disorder, osteoarthritis (hip/knee), rheumatoid arthritis, neck/shoulder disorder, arm/hand disorder, and other chronic condition

[a] Turkish respondents who completed the SF-36 in Turkish

[b] Moroccan respondents who completed the SF-36 in Moroccan-Arabic

Table 2 Median, mean, standard deviation, percentage floor and ceiling, and Cronbach's α for the SF-36 scales

Turkish (N = 409) | | | | | | | Turkish subgroup (N = 162)

	N	Median	Mean	SD	% Floor	% Ceiling	Cronbach's α (N)	Cronbach's α (N)
PF	408	90.0	80.4	24.9	1.0	38.5	0.93 (397)	0.92 (159)
RP	408	100.0	70.8	41.0	19.6	61.0	0.93 (404)	0.92 (160)
BP	409	74.0	70.9	31.9	4.4	45.0	0.95 (403)	0.95 (159)
GH	408	65.0	60.2	23.0	1.0	2.2	0.81 (405)	0.82 (160)
VT	408	55.0	53.8	22.2	1.5	1.0	0.76 (404)	0.73 (158)
SF	409	87.5	75.6	26.7	2.4	38.1	0.77 (401)	0.78 (162)
RE	406	100.0	75.0	39.1	17.2	67.2	0.89 (403)	0.90 (160)
MH	408	68.0	64.5	21.2	1.2	2.2	0.81 (397)	0.80 (157)

Moroccan (N = 377) | | | | | | | Moroccan subgroup (N = 52)

	N	Median	Mean	SD	% Floor	% Ceiling	Cronbach's α (N)	Cronbach's α (N)
PF	377	95.0	86.9	19.6	0.5	47.7	0.91 (354)	0.84 (50)[b]
RP	376	100.0	77.1	37.0	13.3	66.5	0.91 (373)	0.92 (52)
BP	376	100.0	75.8	29.6	1.3	51.1	0.94 (359)	0.93 (51)
GH	377	62.0	61.1	22.8	0	2.7	0.83 (374)	0.79 (51)
VT	377	55.0	56.9	18.5	0	1.9	0.61 (374)	0.54 (52)[b]
SF	377	87.5	79.8	23.3	0.3	40.8	0.74 (373)	0.63 (51)[b]
RE	376	100.0	82.0	35.5	13.0	77.1	0.91 (375)	0.88 (52)
MH	377	72.0	69.2	20.0	0.3	3.2	0.82 (374)	0.85 (52)

Dutch (N = 9,620[a])

	N	Median	Mean	SD	% Floor	% Ceiling	Cronbach's α (N)
PF	9,620	95.0	85.8	21.9	0.5	42.4	0.93 (9603)
RP	9,620	100.0	79.3	35.8	12.1	70.6	0.91 (9610)
BP	9,619	84.0	78.4	24.7	0.6	45.5	0.92 (9617)
GH	9,620	72.0	70.3	20.5	0.1	3.8	0.77 (9599)
VT	9,620	75.0	69.5	18.9	0.1	3.4	0.80 (9597)
SF	9,618	100.0	85.5	21.5	0.7	55.0	0.78 (9612)
RE	9,618	100.0	89.6	26.6	5.6	84.2	0.84 (9611)
MH	9,620	84.0	80.1	15.9	0.1	6.6	0.83 (9605)

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

[a] 8 Dutch respondents were excluded due to large numbers of missing items

[b] Deviant (0.07 or more) from total Turkish or Moroccan population

Table 3 Results of confirmatory factor analysis

SF-36 item		Loading		
		Turkish ($n = 409$)	Moroccan ($n = 377$)[a]	Dutch ($n = 9,620$)
PF by	PF1 Vigorous activities	0.892	0.864	0.880
	PF2 Moderate activities	0.906	0.874	0.934
	PF3 Lifting/carrying	0.912	0.881	0.904
	PF4 Climbing several stairs	0.914	0.936	0.956
	PF5 Climbing 1 flight stairs	0.896	0.929	0.970
	PF6 Bending/kneeling	0.893	0.868	0.861
	PF7 Walking a km	0.919	0.916	0.920
	PF8 Walking few 100 m	0.964	0.940	0.966
	PF9 Walking 100 m	0.918	0.891	0.987
	PF10 Bathing/dressing	0.823	0.787	0.849
RP by	RP1 Cut down time	0.957	0.945	0.928
	PR2 Accomplished less	0.967	0.923	0.945
	RP3 Limited in kind	0.990	0.974	0.977
	PP4 Difficulty performing	0.971	0.981	0.978
BP by	BP1 Pain	0.949	0.971	0.933
	BP2 Pain interfere	0.995	0.960	0.993
GH by	GH1 General health	0.850	0.898	0.846
	GH2 Easier sick	0.615	0.696	0.634
	GH3 As healthy	0.633	0.635	0.596
	GH4 Health get worse	0.615	0.581	0.501
	GH5 Health excellent	0.884	0.863	0.925
VT by	VT1 Full of pep	0.485	0.155	0.682
	VT2 A lot of energy	0.795	0.631	0.811
	VT3 Worn out	0.777	0.794	0.743
	VT4 Tired	0.747	0.738	0.759
SF by	SF1 Social-extent	0.886	0.849	0.889
	SF2 Social-time	0.834	0.861	0.873
RE by	RE1 Cut down time	0.975	0.975	0.959
	RE2 Accomplished less	0.978	0.972	0.993
	RE3 Not careful	0.914	0.967	0.855
MH by	MH1 Nervous	0.617	0.717	0.663
	MH2 Down in dumps	0.840	0.913	0.845
	MH3 Calm and peaceful	0.660	0.554	0.727
	MH4 Downhearted/blue	0.861	0.902	0.848
	MH5 Happy	0.601	0.629	0.704

[a] After some restrictions (correlation between factor VT and MH fixed at 0.95 and error variance of item RE1 fixed at 0.05)

DIF analyses

The results of the DIF analyses are summarized in Table 5. In the total Turkish sample, nonuniform DIF was found for 6 items (PF1, PF2, GH1, VT1, MH3, and MH5). In the total Moroccan sample, non-uniform DIF was found for 3 items (PF1, GH1, and MH3).

Uniform DIF was observed for 7 items in the total Turkish and Moroccan samples. The PF1, VT4, and RE1 showed uniform DIF in both ethnic groups. High odds ratios (odds ratio and confidence interval outside the range 0.53–1.89) were found in the Turkish sample for the items PF1 and PF7. In the Moroccan sample, high odds ratios were found for the items VT1 and VT3.

In the language subgroups, similar patterns (OR above or below 1) of uniform DIF were found (for 86% of the items in the Turkish subgroup and 80% of the items in the Moroccan subgroup). Nonuniform DIF was found in 7 items (5 similar items) in the Turkish subgroup and 1 (similar item) in the Moroccan-Arabic subgroup.

Table 4 Summary of known-group comparisons (Student's t test and ANOVA)

	Sex		Age		Marital status		Education		Comorbidity	
	Δ (BI)	ES	Δ (BI)	ES	Δ (BI)	ES	Δ (BI)	ES	Δ (BI)	ES
Turkish sample (n = 409)										
PF	9.5 (4.7–14.3)*	0.42	32.4 (27.0–37.7)*	1.70	0.8 (−5.1 to 6.7)	0.03	16.7 (12.0–21.3)*	0.86	25.4 (21.2–29.7)*	1.60
RP	12.0 (4.1–19.9)*	0.33	23.9 (14.0–33.9)*	0.61	1.2 (−8.5 to 10.8)	0.03	16.3 (8.4–24.3)*	0.43	35.8 (28.5–43.1)*	1.14
BP	16.1 (10.1–22.2)*	0.56	22.1 (14.4–29.7)*	0.75	1.4 (−6.1 to 8.9)	0.05	14.9 (8.8–21.0)*	0.53	36.3 (31.0–41.5)*	1.67
GH	9.6 (5.2–14.0)*	0.43	22.4 (17.1–27.7)*	1.06	0.5 (−5.0 to 5.9)	0.02	14.9 (10.6–19.2)*	0.73	26.7 (23.0–30.5)*	1.55
VT	12.3 (8.1–16.4)*	0.61	13.0 (7.6–18.4)*	0.60	1.1 (−4.1 to 6.3)	0.05	12.1 (7.9–16.3)*	0.60	19.9 (15.9–23.8)*	1.05
SF	11.2 (6.1–16.3)*	0.46	20.4 (14.0–26.7)*	0.83	4.2 (−2.1 to 10.5)	0.16	11.7 (6.5–16.8)*	0.47	22.8 (18.0–27.5)*	1.15
RE	13.3 (5.8–20.9)*	0.39	24.2 (14.7–33.6)*	0.68	4.8 (−4.4 to 14.1)	0.12	14.0 (6.5–21.6)*	0.41	30.2 (23.0–37.4)*	1.02
MH	10.1 (6.1–14.1)*	0.53	7.2 (1.9–12.4)*	0.34	2.3 (−2.7 to 7.3)	0.11	7.1 (3.0–11.2)*	0.36	17.6 (13.8–21.4)*	1.05
Moroccan sample (n = 377)										
PF	4.7 (0.7–8.6)*	0.28	13.6 (8.4–18.9)*	0.72	−3.6 (−8.4 to 1.2)	−0.18	8.7 (4.8–12.6)*	0.51	16.6 (12.5–20.6)*	1.17
RP	8.5 (1.1–16.0)*	0.25	6.5 (−3.7–16.8)	0.18	−5.7 (−14.8 to 3.4)	−0.15	2.5 (−5.1–10.1)	0.07	33.3 (25.7–40.8)*	1.14
BP	8.1 (2.2–14.1)*	0.29	12.9 (4.8–21.0)*	0.44	−4.4 (−11.7 to 2.9)	−0.15	11.5 (5.5–17.4)*	0.42	31.4 (25.6–37.2)*	1.32
GH	4.1 (−0.5–8.7)	0.18	16.7 (10.6–22.8)*	0.75	−6.9 (−12.4 to −1.3)*!	−0.31	13.1 (8.6–17.5)*	0.61	22.7 (18.1–27.2)*	1.12
VT	7.4 (3.7–11.1)*	0.40	6.3 (1.2–11.4)*	0.34	−5.5 (−10.0 to −1.0)*!	−0.31	6.5 (2.7–10.2)*	0.35	15.6 (11.8–19.5)*	0.88
SF	6.7 (2.0–11.4)*	0.31	10.9 (4.5–17.2)*	0.47	−0.5 (−6.2 to 5.2)	−0.02	6.0 (1.3–10.7)*	0.26	18.5 (13.7–23.4)*	0.94
RE	7.0 (−0.2–14.2)	0.22	−0.2 (−10.1–9.6)	−0.01	6.5 (−2.2 to 15.1)	0.19	2.4 (−4.8–9.7)	0.07	25.0 (17.5–32.6)*	0.89
MH	5.5 (1.5–9.5)*	0.28	6.4 (0.9–11.9)*	0.32	2.5 (−2.4 to 7.3)	0.13	4.0 (−0.1–8.1)	0.20	15.7 (11.5–19.9)*	0.89
Dutch sample (n = 9,620)										
PF	4.8 (4.0–5.7)*	0.24	13.5 (12.7–14.4)*	0.89	2.7 (1.7–3.6)*	0.13	16.9 (15.7–18.1)*	0.88	20.1 (19.3–21.0)*	1.47
RP	7.6 (6.2–9.0)*	0.23	3.4 (2.0–4.8)*	0.10	2.4 (0.9–3.9)*	0.07	8.4 (6.4–10.4)*	0.24	23.3 (21.9–24.7)*	0.83
BP	7.1 (6.1–8.1)*	0.31	5.0 (4.0–5.9)*	0.21	1.6 (0.6–2.7)*	0.07	7.4 (6.0–8.8) *	0.31	20.6 (19.6–21.5)*	1.05
GH	2.2 (1.4–3.0)*	0.11	10.0 (9.2–10.8)*	0.53	1.5 (0.6–2.3)*	0.07	11.1 (10.0–12.2)*	0.56	19.1 (18.3–19.8)*	1.20
VT	6.2 (5.4–6.9)*	0.34	0.2 (−0.6–0.9)	0.01	3.8 (3.1–4.6)*	0.20	6.0 (5.0–7.1) *	0.33	12.3 (11.5–13.0)*	0.75
SF	5.8 (4.9–6.6)*	0.30	2.0 (1.2–2.9)*	0.10	4.1 (3.2–5.0)*	0.20	5.8 (4.6–7.0)*	0.28	12.6 (11.8–13.5)*	0.75
RE	4.2 (3.1–5.3)*	0.18	−0.8 (−1.8–0.3)	−0.03	3.7 (2.6–4.9)*	0.15	1.8 (0.3–3.3)*	0.07	8.9 (7.8–10.0)*	0.41
MH	5.3 (4.7–6.0)*	0.37	1.3 (0.7–1.9)*	0.09	5.4 (4.7–6.1)*	0.37	6.3 (5.4–7.2)*	0.42	7.5 (6.9–8.2)*	0.55

Sex = male minus female; Age = young minus old; Martial status = married/registered partnership minus not married/no registered partnership; Education = high school minus primary school; Comorbidity = none or one condition minus 2 or more conditions; ES = effect size

* $P < 0.05$

*! Statistically significant in the opposite direction to what had been hypothesized

Table 5 Results of uniform and nonuniform DIF by ordinal regression analysis: odds ratio (β), confidence interval, and *P*-values of the Dutch sample ($N = 9,620$) vs. the Turkish or Moroccan sample

	Turkish $N = 409$		Moroccan $N = 377$	
	OR	*P*-value	OR	*P*-value
PF scale				
PF1	**2.56** (**1.92–3.41**)**	0.000	**1.98** (1.49–2.61)	0.000**
PF2	1.07 (0.78–1.48)**	0.669	0.64 (0.46–0.89)	0.008
PF3	0.80 (0.60–1.07)	0.132	0.89 (0.66–1.22)	0.478
PF4	0.74 (0.54–1.02)	0.066	0.79 (0.56–1.13)	0.194
PF5	1.01 (0.68–1.50)	0.950	0.82 (0.51–1.30)	0.395
PF6	1.70 (1.25–2.32)	0.001	1.02 (0.75–1.40)	0.894
PF7	**0.38** (**0.28–0.51**)	0.000	0.55 (0.39–0.79)	0.001
PF8	**0.50** (0.34–0.73)	0.000	0.55 (0.35–0.88)	0.013
PF9	0.68 (0.43–1.08)	0.101	**0.46** (0.25–0.83)	0.010
PF10	0.62 (0.39–0.99)	0.045	0.74 (0.38–1.43)	0.364
RP scale				
RP1	**2.96** (1.74–5.03)	0.000	1.47 (0.82–2.64)	0.197
RP2	**0.38** (0.22–0.64)	0.000	0.62 (0.35–1.09)	0.095
RP3	1.29 (0.65–2.56)	0.468	0.94 (0.47–1.90)	0.871
RP4	0.58 (0.34–0.99)	0.045	0.75 (0.42–1.34)	0.326
BP scale				
BP1	1.17 (0.83–1.67)	0.372	1.28 (0.88–1.85)	0.193
BP2	1.09 (0.80–1.50)	0.581	1.16 (0.83–1.63)	0.391
GH scale				
GH1	0.81 (0.65–1.00)**	0.050	0.89 (0.71–1.11)	0.295**
GH2	0.56 (0.46–0.68)	0.000	0.67 (0.54–0.82)	0.000
GH3	1.85 (1.52–2.26)	0.000	1.83 (1.49–2.25)	0.000
GH4	0.93 (0.77–1.12)	0.448	0.85 (0.70–1.03)	0.103
GH5	1.22 (0.98–1.53)	0.079	1.31 (1.04–1.66)	0.023
VT scale				
VT1	0.64 (0.53–0.77)**	0.000	**0.38** (**0.31–0.47**)	0.000
VT2	1.22 (1.00–1.50)	0.051	0.93 (0.75–1.14)	0.478
VT3	1.05 (0.86–1.28)	0.623	**2.54** (**2.06–3.14**)	0.000
VT4	**2.04** (1.67–2.50)	0.000	**2.12** (1.72–2.61)	0.000
SF scale				
SF1	1.28 (0.97–1.70)	0.079	1.76 (1.30–2.38)	0.000
SF2	0.78 (0.60–1.02)	0.074	0.57 (0.43–0.74)	0.000
RE scale				
RE1	**2.07** (1.00–4.27)	0.050	**4.32** (1.75–10.67)	0.002
RE2	0.86 (0.44–1.71)	0.674	0.70 (0.31–1.56)	0.381
RE3	0.63 (0.37–1.06)	0.084	0.78 (0.42–1.48)	0.453
MH scale				
MH1	1.46 (1.19–1.78)	0.000	**1.96** (1.59–2.42)	0.000
MH2	0.56 (0.45–0.69)	0.000	0.71 (0.56–0.89)	0.003
MH3	1.34 (1.09–1.64)**	0.005	0.70 (0.57–0.86)	0.001**
MH4	0.70 (0.57–0.86)	0.001	1.42 (1.14–1.77)	0.002
MH5	1.69 (1.39–2.07)**	0.000	0.98 (0.80–1.20)	0.864

All analyses were corrected for age, sex, and comorbidity

Log odds ratio = $e^{\text{Estimate} \pm 1.96 \cdot SE}$

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

Nonuniform DIF was considered to be present if the interaction of language with the total score was found to be statistically significant ($P \leq 0.001$) and is presented in the table with an **

Discussion

In this paper we have reported the results of a study of the psychometric properties of the SF-36 Health Survey when employed among Turkish and Moroccan ethnic minority groups in the Netherlands, and compared to individuals from the indigenous Dutch population.

Descriptive statistics

In the Turkish and Moroccan samples, the number of missing items was very low (mean of less than 1%) and was comparable to rates found in other studies in which the SF-36 was administered as a (telephonic) interview in non-Western samples. [16–20]. Higher rates of missing data (3–4.5%) were observed in the study of Hoopman et al. [7], in which the SF-36 was administered (primarily in interview form) to Turkish and Moroccan patients diagnosed with cancer in the Netherlands. The higher (but still low) rates of missing data observed in the latter study may be due to the fact that, in that study, interviewers were explicitly instructed to be restraint in explaining any items to the respondents or to assist in any way in completing the questionnaire. Such stringent administrative rules are not always applied in survey research.

Factor analysis

The scale structure of the SF-36 fit well in the Dutch and Turkish groups. In the Moroccan group, the expected structure did not fit well, as the MH and VT scale were highly correlated, indicating that they are not discerning factors. After the correction (where this correlation was "allowed"), the model fit was improved.

The only item that did not fit well into the scale was item VT1. This corresponds with the finding of uniform DIF for this item in this study and in our previous study on patients with cancer [7]. Item VT1 also showed low item convergent and item discriminant validity (using multitrait scaling analysis) in the Moroccan group in the previous study of Hoopman et al. [7].

Internal consistency reliability

The internal consistency reliability of the SF-36 scales was satisfactory for all 3 samples, with the exception of the VT scale in the total Moroccan group and the VT and SF scales in the Moroccan-Arabic language subgroup. The less-than-satisfactory results for the VT scale were also reported in a study of Moroccan ethnic minority patients with cancer [7].

Known groups validity

The majority of the know groups comparisons yielded results consistent with what had been hypothesized. In the Turkish and Moroccan samples, no significant differences were observed in SF-36 scores as a function of marital status. This lack of association between marital status and health status has also been reported in studies in Lebanon [19], Greece [21], and Morocco [18].

DIF analyses

Considering the items that showed uniform DIF either in the current study or in the previous study of Hoopman et al. [7], we found in the Turkish group that 83% (15 of 18 items) and in the Moroccan group 67% (16 of 24 items) of these items had an odds ratio in the same direction (below or above 1.0, although not all significant). This indicates that there were relatively consistent findings across the two studies. In contrast, there were almost no consistent findings between the current study and that of Hoopman et al. with regard to non-uniform DIF (only item MH3 in the Turkish group showed non-uniform DIF in both studies).

Items that showed uniform DIF in both studies in the Turkish group were PF1, RP2 and RE1 and in the Moroccan group PF1, PF9, VT1 and VT4. In both studies in both ethnic groups, clear DIF was found in the PF1 item (participating in strenuous sports and running), with Dutch respondents tending to report more limitations than Turkish or Moroccan respondents. Turks and Moroccans may be less likely to indicate problems with these activities, simply because they tend not to perform them [7]. There is no obvious explanation for the finding that, in both studies, in the Moroccan group, items VT1 and VT4 showed clear DIF, with Dutch respondents tending to report less energy and more tiredness.

The fact that similar results were obtained in the current study as in the previous study among patients diagnosed with cancer [7], is important, in that it suggests that: (1) the findings are not specific to cancer patient populations only; and (2) the findings are not specific to a certain language version of the SF-36. Rather the results suggest that culture is responsible for the fact that Turkish and Moroccan respondents respond differently to some of the SF-36 items compared to Dutch respondents. In future research, investigators should be aware of possible cultural differences in responses to and interpretation of some SF-36 items. However, this is primarily of importance when comparing results across cultural groups, rather than investigating within group issues.

This study had several limitations that should be noted. First, recruitment into the study began with a one-page postal questionnaire that required self-completion. This could possibly have had a negative effect on the response rate, particularly among illiterate individuals. This, in turn, may have led to an under-representation of certain sub-groups (e.g., elderly and women) whose literacy rates tend to be lower in these ethnic minority populations. Nevertheless, the distribution of age and sex of the respondents per ethnic group did not differ systematically from the national figures [22, 23].

Second, the failure to systematically record the language in which the questionnaire was administered in the DNSGP2 survey required post-hoc classification into language categories based on the best available information and certain assumptions. Although we are fairly confident that the criteria used to classify language of administration were appropriate, we cannot rule out some degree of misclassification.

Third, the routing algorithms used in administering the PF and the BP items may have had some influence on the scale score distributions, factor analysis results, internal consistency estimates, and DIF results. To examine the possible influence of routing in the PF scale, we reanalyzed the data from the study of Hoopman et al. [7] to compare the results based on the standard administration procedures with those based on the routing algorithm used in the current study. The scale score distributions and the internal consistency reliability estimates were nearly identical. Similarly, the mean PF scale score and the Cronbach's α coefficients derived from the current study for the Dutch sample were very similar to those reported by Aaronson et al. in 1996 in a study employing the original administrative algorithm [8]. For these reasons, we believe that routing did not have a significant influence on the current study results.

Finally, to ensure that collapsing the 6-level response scale used for the SF2 item in the DNSGP2 survey to 5 levels as in the original SF2 item did not effect the psychometric results, we repeated all statistical analyses using the 6-level response scale. The results were very nearly identical to those obtained when using the collapsed 5 item response scale, with only very much deviations that were not statistically significant (data not presented but available upon request).

In conclusion, the findings from this study indicate that the SF-36 has generally satisfactory psychometric properties when employed among Turkish and Moroccan ethnic minority populations in the Netherlands and that the psychometric results are similar to those observed in the indigenous Dutch population. Additional studies are needed to confirm the psychometrics of the SF-36 when used among these ethnic minority groups residing in other Western European countries, including Belgium, France, Germany, and Spain.

References

1. Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 health survey manual and interpretation guide*. Boston, MA: New England Medical Center, The Health Institute.
2. Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*, 473–483. doi:10.1097/00005650-199206000-00002.
3. Salt, J. (2002). *European international migration: Evaluations of the current situation*. Strasbourg: Council of Europe.
4. Centraal Bureau voor de Statistiek. (2008). *Jaarrapport Integratie 2008*. Den Haag/Heerlen: CBS.
5. Sociaal en Cultureel Planbureau. (2004). *Cijferrapport Allochtone Ouderen. Werkdocument 105*. Den Haag: SCP.
6. Sociaal en Cultureel Planbureau. (2005). *Jaarrapport Integratie 2005*. Den Haag: SCP/WPDC/CBS.
7. Hoopman, R., Terwee, C. B., Muller, M. J., & Aaronson, N. K. (2006). Translation and validation of the SF-36 health survey for use among Turkish and Moroccan ethnic minority cancer patients in The Netherlands. *European Journal of Cancer, 42*, 2982–2990. doi:10.1016/j.ejca.2006.08.011.
8. Aaronson, N. K., Muller, M., Cohen, P. D., Essink-Bot, M. L., Fekkes, M., Sanderman, R., et al. (1998). Translation, validation, and norming of the Dutch language version of the SF-36 health survey in community and chronic disease populations. *Journal of Clinical Epidemiology, 51*, 1055–1068. doi:10.1016/S0895-4356(98)00097-3.
9. Ware, J. E., Kosinski, M., & Keller, S. D. (1994). *SF-36 physical and mental health summary scales—a users' manual*. Boston, MA: New England Medical Center, the Health Institute.
10. Bullinger, M., Alonso, J., Apolone, G., Leplege, A., Sullivan, M., Wood-Dauphinee, S., et al. (1998). Translating health status questionnaires and evaluating their quality: The IQOLA project approach. International quality of life assessment. *Journal of Clinical Epidemiology, 51*, 913–923. doi:10.1016/S0895-4356(98)00082-1.
11. Ware, J. E., Keller, S. D., Gandek, B., Brazier, J. E., & Sullivan, M. (1995). Evaluating translations of health status questionnaires. Methods from the IQOLA project. International quality of life assessment. *International Journal of Technology Assessment in Health Care, 11*, 525–551. doi:10.1017/S0266462300008710.
12. Pinar, R. (2005). Reliability and construct validity of the SF-36 in Turkish cancer patients. *Quality of Life Research, 14*, 259–264. doi:10.1007/s11136-004-2393-3.
13. Westert, G. P., Schellevis, F. G., De Bakker, D. H., Groenewegen, P. P., Bensing, J. M., & van der, Z. J. (2005). Monitoring health inequalities through general practice: The Second Dutch National Survey of general practice. *European Journal of Public Health, 15*, 59–65. doi:10.1093/eurpub/cki116.
14. Petersen, M. A., Groenvold, M., Bjorner, J. B., Aaronson, N., Conroy, T., Cull, A., et al. (2003). Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Quality of Life Research, 12*, 373–385. doi:10.1023/A:1023488915557.
15. Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.
16. Demiral, Y., Ergor, G., Unal, B., Semin, S., Akvardar, Y., Kivircik, B., et al. (2006). Normative data and discriminative properties of short form 36 (SF-36) in Turkish urban population. *BMC Public Health, 6*, 247. doi:10.1186/1471-2458-6-247.
17. Gandek, B., Ware, J. E, Jr, Aaronson, N. K., Alonso, J., Apolone, G., Bjorner, J., et al. (1998). 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. *Journal of Clinical Epidemiology, 51*, 1149–1158. doi:10.1016/S0895-4356(98)00106-1.
18. Khoudri, I., Ali, Z. A., Abidi, K., Madani, N., & Abouqal, R. (2007). Measurement properties of the short form 36 and health-related quality of life after intensive care in Morocco. *Acta Anaesthesiologica Scandinavica, 51*, 189–197. doi:10.1111/j.1399-6576.2006.01225.x.
19. Sabbah, I., Drouby, N., Sabbah, S., Retel-Rude, N., & Mercier, M. (2003). Quality of Life in rural and urban populations in Lebanon using SF-36 health survey. *Health and Quality of Life Outcomes, 1*, 30. doi:10.1186/1477-7525-1-30.
20. Wagner, A. K., Wyss, K., Gandek, B., Kilima, P. M., Lorenz, S., & Whiting, D. (1999). A Kiswahili version of the SF-36 health survey for use in Tanzania: Translation and tests of scaling assumptions. *Quality of Life Research, 8*, 101–110. doi:10.1023/A:1026441415079.
21. Pappa, E., Kontodimopoulos, N., & Niakas, D. (2005). Validating and norming of the Greek SF-36 health survey. *Quality of Life Research, 14*, 1433–1438. doi:10.1007/s11136-004-6014-y.
22. Centraal Bureau voor de Statistiek. (2002). *Jaarboek Statistiek 2002*. Voorburg: CBS.
23. Uiters, E. (2007). *Primary health care use among ethnic minorities in the Netherlands. A comparative study*. Rotterdam: Erasmus Universiteit Rotterdam.