

US general population norms for telephone administration of the SF-36v2

Gregory A. Maglinte^{a,*}, Ron D. Hays^{a,b}, Robert M. Kaplan^c

^aDepartment of Health Services, UCLA School of Public Health, Los Angeles, CA 90095, USA

^bDepartment of Medicine, UCLA Division of General Internal Medicine & Health Services Research, Los Angeles, CA 90095, USA

^cOffice of Behavioral and Social Sciences Research, National Institutes of Health, Bethesda, MD 20892, USA

Accepted 4 September 2011; Published online 24 January 2012

Abstract

Objective: US general population norms for mail administration of the Medical Outcomes Study 36-Item Short Form Version 2 (SF-36v2) were established in 1998. This article reports SF-36v2 telephone-administered norms collected in 2005–2006 for adults aged 35–89 years.

Study Design and Setting: The SF-36v2 was administered to 3,844 adults in the National Health Measurement Study (NHMS), a random-digit dial telephone survey. Scale scores and physical and mental component summary (PCS and MCS) scores were computed.

Results: When compared with 1998 norms (mean = 50.00, standard deviation [SD] = 10.00), SF-36v2 scores for the 2005–2006 general population tended to be higher: physical functioning (mean = 50.68, SD = 14.48); role limitations due to physical health problems (mean = 49.47, SD = 14.71); bodily pain (mean = 50.66, SD = 16.28); general health perceptions (mean = 50.10, SD = 16.87); vitality (mean = 53.71, SD = 15.35); social functioning (mean = 51.37, SD = 13.93); role limitations due to emotional problems (mean = 51.44, SD = 13.93); mental health (mean = 54.27, SD = 13.28); PCS (mean = 49.22, SD = 15.13); MCS (mean = 53.78, SD = 13.14). PCS and MCS factor scoring coefficients were similar to those previously reported for the 1998 norms. SF-36v2 norms for telephone administration were created.

Conclusion: The higher scores for NHMS data are likely due to the effect of telephone administration. The 2005–2006 norms can be used as a reference to interpret scale and component summary scores for telephone-administered surveys with the SF-36v2. © 2012 Elsevier Inc. All rights reserved.

Keywords: SF-36; Health-related quality of life; General population norms; Population survey; Factor analysis; Factor scoring coefficients

1. Introduction

The Medical Outcomes Study 36-Item Short Form Version 2 (SF-36v2; QualityMetric, Lincoln, RI) is one of the most commonly used generic health-related quality of life (HRQOL) instruments. The SF-36v2 was derived from 40 concepts captured in the Medical Outcomes Study [1,2]. It is composed of eight multi-item scales: physical functioning, role limitations due to physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health. Additionally, the scores from the SF-36v2 scales can be combined into physical and mental component summary (PCS and MCS) scores [3].

The first version of the SF-36 was made available in 1990 [4]. Based on work in the International Quality of Life Assessment project [5] and the Veterans Health Survey [6], item wording and response categories were refined to create

the SF-36v2. The SF-36v2 uses norm-based scoring in which the scale and component summary scores have a mean of 50 and standard deviation of 10 in the US general population. Norms were derived from the 1998 National Survey of Functional Health Status sample of respondents that was designed, based on geographical region, market size, age, income, and household size, to be representative of the noninstitutionalized adult US population [5]. The 1998 SF-36v2 US general population norms have allowed for comparisons of scores from groups with different demographic characteristics or different health conditions.

Standard scoring of the PCS and MCS is based on factor score coefficients or weights from each scale of the SF-36v2 obtained from principal component analysis with orthogonal rotation (i.e., uncorrelated factors) [7]. Alternatively, physical and mental component summaries can be viewed as correlated. Physical and mental health summary scores derived from obliquely rotated factor solutions have also been derived for the SF-36v1 [8,9].

Norms for the SF-36v2 have not been updated since 1998. However, the median age and proportion of minorities have

* Corresponding author. Tel.: 818-732-0788.

E-mail address: greg.maglinte@ucla.edu (G.A. Maglinte).

What is new?

- This study provides Medical Outcomes Study 36-Item Short Form Version 2 (SF-36v2) telephone-administered norms for the 2005–2006 US general population aged 35–89 years.
- The 2005–2006 SF-36v2 scores are higher than the 1998 norms, which are likely due to telephone administration.
- The 2005–2006 norms can be used to help interpret SF-36v2 scale and component summary scores.

increased in the US population since then [10]. Furthermore, the prevalence of obesity has increased among men from 1999 to 2004 [11], whereas age-adjusted mortality from cardiovascular disease has declined [12]. In order for scale and component summary score comparisons to be more current and relevant, the SF-36v2 norms need updating.

Over a 13-month period in 2005–2006, a cross-sectional survey of 3,844 US adults was conducted in the National Health Measurement Study (NHMS). Several generic HRQOL were administered, including the SF-36v2. In this study, we report new US norms for the eight scales of the SF-36v2, and the PCS and MCS. We provide both orthogonal and oblique factor score coefficients.

2. Methods

2.1. Study design

The NHMS is a random-digit dial telephone survey of adults in the United States aged 35–89 years. The cross-sectional survey was administered using a computer-assisted telephone interview script by the University of Wisconsin Survey Center from June 2005 to August 2006. All subjects in the NHMS provided verbal informed consent, and the survey was approved by the University of Wisconsin Institutional Review Board (protocol #H-2004-0083).

2.2. Sample

Participants in the NHMS were household residents aged 35–89 years in the continental United States with working noncellular telephone numbers. African Americans and individuals 65 years or older were oversampled. The 40% of households for which we obtained street addresses from reverse directories initially received a letter describing the purpose of the survey and an incentive of \$2 cash to increase study participation. They were notified they would receive a \$25 check if they completed the interview. The

preliminary sample contained 47,933 telephone numbers (Fig. 1). After accounting for nonworking and nonresidential numbers, 14,394 households were eligible for screening, of which 11,656 screening interviews were completed. A total of 6,822 households were identified as having an eligible respondent, and 3,844 respondents completed the telephone survey. The analysis was conducted on this final sample.

The response rate was calculated using two different methods [13]. The simple estimate was calculated as the proportion of eligible respondents (6,822) who completed the interview (3,844), 56%. However, assuming that the proportion of eligible respondents among unscreened households was the same for screened households ($6,822/11,656 = 0.585$), there could have been 1,602 ($0.585 \times 2,738$) eligible respondents in the 2,738 households not eligible for screening. Thus, the alternative response rate was based on the number of eligible respondents among screened and unscreened households, and was estimated as 46% ($3,844/[6,822 + 1,602]$).

To evaluate the comparability of the NHMS to other general population studies, we considered responses to the most widely used general health item: “In general, would you say your health is excellent, very good, good, fair, or poor?” Responses in this sample were compared with the Patient-Reported Outcomes Measurement Information System (PROMIS), Medical Expenditure Panel Survey (MEPS), Behavioral Risk Factor

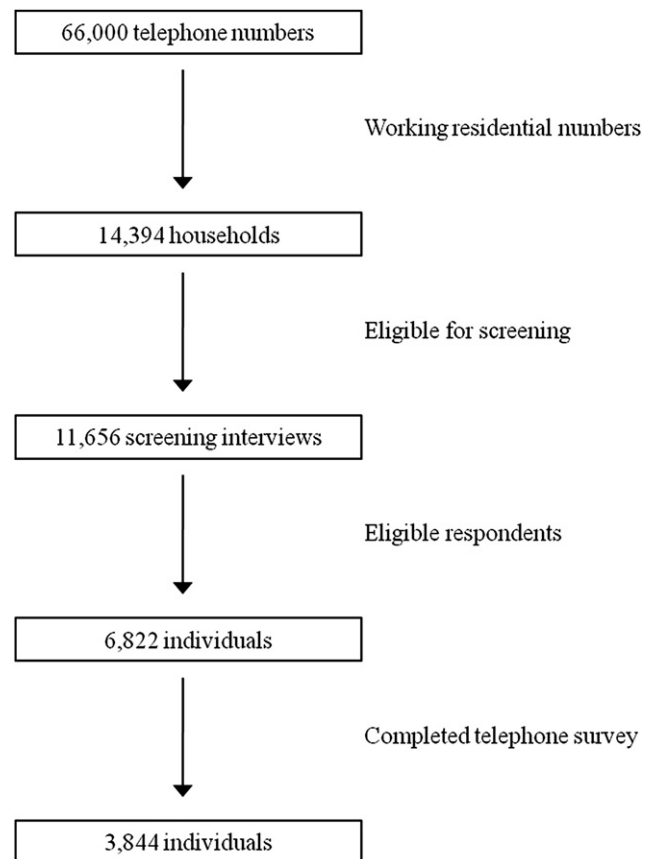


Fig. 1. National Health Measurement Study (NHMS) Sample Selection.

Surveillance System (BRFSS), and National Health and Nutrition Examination Survey (NHANES).

2.3. Weighted analysis

To account for the complex study design, including oversampling of African Americans and older individuals, and produce estimates from the NHMS that are representative of the US general population, sampling weights were implemented in the analysis and poststratified to the 2000 US Census population on age (35–44, 45–64, 65–89 years), race (African American, white, other), and gender (male, female). Furthermore, the poststratified sampling weights were trimmed (i.e., combined with raked weights) so that no one individual accounted for more than 95% of the total weight in a subgroup. Weighted estimates were calculated using PROC SURVEYMEANS with SAS software, version 9.1 (SAS Institute Inc., Cary, NC) [14].

2.4. Scale and component summary scores

We calculated norm-based (mean of 50, standard deviation [SD] of 10) scale T-scores. The calculation involved two steps, and is similar to the procedure used by Ware et al. [5]. In the first step, z-scores (standardized to the 1998 US population) were calculated by subtracting the 1998 US mean scale score from the observed 0–100 score and dividing the difference by the 1998 US standard deviation of the scale score. In the second step, z-scores were converted into norm-based scores by multiplying them by 10 and adding 50 to the product.

In addition, we standardized the SF-36v2 scores to the 2005–2006 US general population aged 35–89 years using the NHMS data. Weighted 0–100 scores can be transformed to a T-score metric with the mean fixed to 50 and SD to 10 in the NHMS sample.

Weighted mean PCS and MCS scores are also provided. The standard PCS and MCS scores were obtained from the public NHMS data set, and are based on orthogonal (uncorrelated) factors.

2.5. Factor analysis with orthogonal and oblique rotation

We estimate factor score coefficients using two methods: orthogonal (uncorrelated) rotation such as Ware et al. [5,7], and oblique (correlated) rotation such as Farivar et al. [9]. The analysis was conducted using PROC FACTOR with SAS software, version 9.1 [14], specifying the extraction of two factors [7]. Both sets of factor score coefficients were compared with the orthogonal values produced by Ware et al. [7] and the oblique values produced by Farivar et al. [9].

3. Results

3.1. Sample characteristics

The age range was 35–89 years and the weighted mean age of the estimated 2005–2006 US general population

was about 54 years (standard error [SE] = 0.35) (Table 1). A greater proportion of the population was white (82%), non-Hispanic (96%), and female (53%). The population was well educated, as more than 60% of individuals had some college education or had a college degree. A greater proportion of the population was in the higher income groups, with almost 60% earning an annual household income of \$50,000 or more. Approximately, 85% of the population was in “good” or “excellent” health (Table 2), with a mean general health score of 3.55 (SE = 0.03, SD = 1.77) (Table 3). The weighted mean score and response frequencies are comparable to other general population studies.

3.2. Comparisons of self-rated health to other national samples

The SF-36v2 general health item scores indicated that most of the 2005–2006 US general population had “good” to “excellent” health. The mean NHMS general health item score was not significantly different from previously reported national norms established in several online, mail,

Table 1. Weighted 2005–2006 US general population demographics ($N = 3,844$) in the NHMS

Continuous variable	Weighted mean (weighted SE)	Min, max
Age	54.31 (0.35)	35, 89
Categorical variable	Frequency ^a	Weighted percent
Gender		
Female	2,203	52.9
Male	1,641	47.1
Race		
White	2,562	81.8
African American	1,086	10.6
Asian	34	2.4
Other	144	5.3
Hispanic		
Yes	118	4.1
No	3,706	95.9
Education		
Less than HS	464	8.3
HS degree	1,159	28.3
Some college	856	22.2
College	819	25.2
Masters/doctorate	522	16.0
Household income		
<\$10,000	253	3.5
\$10,000–\$14,999	255	3.0
\$15,000–\$19,999	273	4.6
\$20,000–\$24,999	281	5.3
\$25,000–\$34,999	418	10.3
\$35,000–\$49,999	551	14.9
\$50,000–\$74,999	625	21.4
>\$75,000	861	37.1

Abbreviations: NHMS, National Health Measurement Study; SE, standard error; HS, high school.

^a Cumulative frequencies may not equal total sample (N) due to missing values.

Table 2. Weighted general health item response frequencies in three national studies

Response	NHMS ^a	MEPS ^{b,c}	BRFSS ^{a,b}
	(N = 3,841)	(N = 20,777)	(N = 3,52,036)
Excellent—5	19.61	19.09	20.72
Very good—4	34.58	35.89	32.16
Good—3	30.29	30.44	30.31
Fair—2	12.11	11.49	12.34
Poor—1	3.41	3.09	4.46

Abbreviations: NHMS, National Health Measurement Study; MEPS, Medical Expenditure Panel Survey; BRFSS, Behavioral Risk Factor Surveillance System.

^a Telephone survey.

^b Source: Liu et al. [15], epub ahead of print.

^c Mail survey.

and telephone surveys, PROMIS, MEPS, BRFSS, and NHANES (see Table 2) [15]. Furthermore, NHMS general health item response frequencies were similar to MEPS and BRFSS as reported by Liu et al. [15].

3.3. Scale and component summary scores

The mean standard norm-based scale scores ranged from 49.47 for role-physical to 54.27 for mental health (Table 4). With the exception of role-physical and the PCS, all mean scale scores and the MCS score are higher (more positive) for the 2005–2006 US general population compared with the 1998 US general population. Weighted age- and gender-stratified 2005–2006 US general population scale scores, and standard PCS and MCS scores are shown in Tables 1–16 in the Appendix, on the journal's Web site at www.jclinepi.com.

3.4. Orthogonal and oblique factors

Factor loadings and corresponding factor scoring coefficients for orthogonal and oblique rotations of the NHMS data are shown in Tables 5 and 6, respectively. When compared with the physical summary orthogonally rotated coefficients produced by Ware et al. [5,7] using the 1998 US

Table 3. Weighted general health item mean scores in five national studies

Study	N	Mean	SE	SD
NHMS ^a	3,841	3.55	0.029	1.80
2006–2007 PROMIS ^{b,c}	11,794	3.50	0.017	1.85
2004 MEPS ^{b,d}	20,777	3.56	0.012	1.73
2005 BRFSS ^{a,b}	352,036	3.52	0.004	2.37
2001–2002 NHANES ^{a,b}	6,873	3.50	0.017	1.41

Abbreviations: SE, standard error; SD, standard deviation; NHMS, National Health Measurement Study; PROMIS, Patient-Reported Outcomes Measurement Information System; MEPS, Medical Expenditure Panel Survey; BRFSS, Behavioral Risk Factor Surveillance System; NHANES, National Health and Nutrition Examination Survey.

^a Telephone survey.

^b Source: Liu et al. [15], epub ahead of print.

^c Online survey.

^d Mail survey.

Table 4. Weighted 2005–2006 US general population SF-36v2 scale and component summary scores in the NHMS

Scale	n	Mean	SE	SD
PF	3,842	50.68	0.23	14.48
RP	3,840	49.47	0.24	14.71
BP	3,843	50.66	0.26	16.28
GH	3,844	50.10	0.27	16.87
VT	3,844	53.71	0.25	15.35
SF	3,843	51.37	0.22	13.93
RE	3,835	51.44	0.21	13.12
MH	3,828	54.27	0.21	13.28
PCS	3,828	49.22	0.24	15.13
MCS	3,828	53.78	0.21	13.14

Abbreviations: SF-36v2, Medical Outcomes Study 36-Item Short Form Version 2; NHMS, National Health Measurement Study; SE, standard error; SD, standard deviation; PF, physical functioning; RP, role-physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role-emotional; MH, mental health; PCS, physical component summary; MCS, mental component summary.

general population, the corresponding NHMS coefficients were slightly higher for physical functioning, bodily pain, general health, vitality, and role emotional (Table 6). The mental summary orthogonally rotated coefficients in the NHMS were slightly higher for social functioning and mental health compared with the Ware coefficients. In addition, the same scales had negative coefficients for Ware et al. [5,7] and NHMS, but were smaller for the latter. For obliquely rotated physical summary coefficients, values were higher by a small degree for physical functioning, bodily pain, and general health when comparing coefficients obtained in the NHMS to Farivar et al. [9]. Obliquely rotated mental summary coefficients were also slightly higher for social functioning, role emotional, and mental health in the NHMS. Furthermore, although the size of the negative loadings was not large, more obliquely rotated physical and mental summary coefficients had negative values for the NHMS compared with Farivar et al. [9].

Correlated physical and mental health component summary (PCS-C and MCS-C) scores, using the obliquely

Table 5. SF-36v2 factor loadings with orthogonal (uncorrelated) and oblique (correlated) rotation in the NHMS

Scale	Orthogonal		Oblique	
	PCS	MCS	PCS-C	MCS-C
PF	0.87	0.25	0.93	−0.06
RP	0.80	0.37	0.81	0.12
BP	0.77	0.32	0.79	0.07
GH	0.74	0.31	0.76	0.07
VT	0.55	0.61	0.41	0.51
SF	0.48	0.71	0.29	0.65
RE	0.34	0.78	0.09	0.8
MH	0.20	0.90	−0.12	0.99

Abbreviations: SF-36v2, Medical Outcomes Study 36-Item Short Form Version 2; NHMS, National Health Measurement Study; PF, physical functioning; RP, role-physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role-emotional; MH, mental health; PCS, physical component summary; MCS, mental component summary.

Table 6. SF-36v2 factor score coefficients with orthogonal and oblique rotation from the NHMS

Scale	Physical summary				Mental summary			
	Ware-orthogonal	NHMS-orthogonal	Farivar-oblique	NHMS-oblique	Ware-orthogonal	NHMS-orthogonal	Farivar-oblique	NHMS-oblique
PF	0.42	0.43	0.20	0.33	−0.23	−0.24	−0.02	−0.09
RP	0.35	0.32	0.31	0.27	−0.12	−0.12	0.03	−0.01
BP	0.32	0.33	0.23	0.27	−0.10	−0.14	0.04	−0.03
GH	0.25	0.32	0.20	0.26	−0.02	−0.13	0.10	−0.03
VT	0.03	0.04	0.13	0.10	0.24	0.19	0.29	0.20
SF	−0.01	−0.05	0.11	0.05	0.27	0.30	0.14	0.27
RE	−0.19	−0.17	0.03	−0.03	0.43	0.43	0.20	0.35
MH	−0.22	−0.32	−0.03	−0.12	0.49	0.58	0.35	0.45

Abbreviations: SF-36v2, Medical Outcomes Study 36-Item Short Form Version 2; NHMS, National Health Measurement Study; PF, physical functioning; RP, role-physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role-emotional; MH, mental health.

rotated factor score coefficients from Farivar et al. [9], are presented in Tables 1–16 in the Appendix, on the journal's Web site at www.jclinepi.com. Additionally, physical and mental health component summary scores, standardized to the 1998 US general population and using NHMS orthogonal (uncorrelated) and oblique (correlated) factor score coefficients (abbreviated as NPCS and NMCS for uncorrelated factors and NPCS-C and NMCS-C for correlated factors), are provided in Tables 1–16 in the Appendix, on the journal's Web site at www.jclinepi.com. These scores can be compared with the published 1998 age- and gender-stratified norms to explore how different factor score coefficients may impact the norm-based PCS and MCS scores.

4. Discussion

We offer new SF-36v2 norms for the US general population aged 35–89 years. The standard (orthogonally rotated) MCS score and nearly all scale scores for the 2005–2006 US general population were higher than the 1998 US general population norms. The physical and mental summary orthogonally rotated factor score coefficients in this sample were similar in magnitude and direction to those reported by Ware et al. [5,7]. Likewise, most of the physical and mental summary obliquely rotated factor score coefficients in the NHMS were similar to those derived by Farivar et al. [9].

There were some limitations in this analysis. The 2005–2006 US general population norms for the SF-36v2 were not computed for any age groups between 18 and 34 years. Further, the response rate was less than 60%. Our survey sample was not completely representative of the US population. They were fairly educated with moderately high household income in relation to national norms. Better-educated and wealthier individuals tend to have better health [16]. However, ratings of general health in this sample were similar to other national samples, suggesting that the sample is comparable.

A major difference between this study and previous SF-36v2 norming studies is mode of survey administration [17]. The 2005–2006 norms were obtained via telephone administration, but the 1998 norms were obtained by mail administration. In the Hanmer et al.'s study [17], respondents who were 70 years or older reported better HRQOL on

telephone-administered questionnaires compared with self-administered questionnaires, including the EuroQol (EQ-5D), Health Utilities Index Mark 3 (HUI-3), and general health item. McHorney et al. [18] observed similar results for older respondents on the first version of the SF-36. Even after adjusting for sociodemographic differences between telephone and mail respondents, McHorney et al. showed that mean scale scores remained higher for telephone respondents. Moreover, when the SF-36v2 was administered both by mail and telephone, Hays et al. [19] showed that telephone administration resulted in significantly higher MCS scores regardless of whether the telephone administration occurred before or after the mail administration. In addition, higher PCS scores were found for telephone administration when it followed mail administration.

5. Conclusions

Comparisons to other general population studies indicate that the norms are representative of the 2005–2006 national population, but they are most applicable to telephone administration of the SF-36v2. Researchers conducting telephone-administered surveys that include the SF-36v2 can use the 2005–2006 norms as a reference to interpret scale and component summary scores.

Acknowledgments

This research was supported by a grant from the National Institute on Aging (P01-AG020679). Hays and Kaplan were also supported in part by NIA grants (P30AG021684 and P30-AG028748), and Hays had partial support from NCMHD (P20MD000182).

The authors would like to thank Dennis G. Fryback for his thoughtful review of this article.

Appendix

Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jclinepi.2011.09.008](https://doi.org/10.1016/j.jclinepi.2011.09.008).

References

- [1] Hays RD, Stewart AL. The structure of self-reported health in chronic disease patients. *Psychol Assess* 1990;2:22–30.
- [2] Stewart AL, Sherbourne CD, Hays RD, Wells KB, Nelson EC, Kamberg CJ, et al. Summary and discussion of MOS measures. In: Stewart AL, Ware JE, editors. *Measuring functioning and well-being: the Medical Outcomes Study approach*. Durham, NC: Duke University Press; 1992:345–71.
- [3] Ware JE. SF-36 Health Survey update. *Spine* 2000;25:3130–9.
- [4] Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey manual and interpretation guide. Boston, MA: The Health Institute; 1993.
- [5] Ware JE, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. User's manual for the SF-36v2 Health Survey. 2nd ed. Lincoln, RI: QualityMetric Incorporated; 2007.
- [6] Kazis LE, Lee A. Measurement comparisons of the Medical Outcomes Study and Veterans SF-36 Health Survey. *Health Care Financ Rev* 2004;25(4):43–58.
- [7] Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston, MA: The Health Institute; 1994.
- [8] Hays RD, Prince-Embury S, Chen H. R-36 HSI: RAND-36 Health Status Inventory. San Antonio, TX: The Psychological Corporation; 1998.
- [9] Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36v2 and SF-12 Health Survey, V.1. *Health Qual Life Outcomes* 2007;5:54.
- [10] U.S. Census Bureau. The 2009 Statistical Abstract: PDF version. Retrieved from <http://www.census.gov/compendia/statab/2009edition.html>; Accessed May 26, 2009.
- [11] Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *J Am Med Assoc* 2006;295:1549–55.
- [12] Weisfeldt ML, Ziemann SJ. Advances in the prevention and treatment of cardiovascular disease. *Health Aff* 2007;26(1):25–37.
- [13] Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherapanov D, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement Study. *Med Care* 2007;45(12):1162–70.
- [14] SAS Institute Inc. SAS software version 9.1. Cary, NC: SAS Institute Inc; 2003.
- [15] Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol* 2010;63(11):69–78.
- [16] Andersen RM, Davidson PL. Improving access to care in America: individual and contextual indicators. In: Andersen RM, Rice TH, Kominski GF, editors. *Changing the U.S. healthcare system*. 2nd ed. San Francisco, CA: Jossey-Bass; 2001:3–30.
- [17] Hanmer J, Hays RD, Fryback DG. Mode of administration is important in U.S. national estimates of health-related quality of life. *Med Care* 2007;45(12):1171–9.
- [18] McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36v2 Health Survey collected by mail versus telephone interview: results from a national survey. *Med Care* 1994;32:551–67.
- [19] Hays RD, Kim S, Spritzer KL, Kaplan RM, Tally S, Feeny D, et al. Effects of mode and order of administration on generic health-related quality of life scores. *Value Health* 2009;12:1035–9.