Schmollers Jahrbuch 127 (2007), 171 – 182
Duncker & Humblot, Berlin

Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF-12v2

By Hanfried H. Andersen, Axel Mühlbacher, Matthias Nübling, Jürgen Schupp, and Gert G. Wagner

1. Computation of Scale Values for SF-12v2

In the year 2002, the Socio-Economic Panel (SOEP) fundamentally revised its questions on the general physical condition of respondents (Schupp / Wagner, 2007). Since then, indicators of a generally accepted and internationally applied inventory of health measures – the so-called SF-12v2 indicators – have been surveyed in two-year intervals (Wagner et al., 2007).[1] The "SF-12v2™ Health Survey" is a 12-item subset of the SF-36v2™ that measures the same eight domains of health.[2] As a brief, reliable measure of overall health status, it is frequently embedded in longer, condition-specific surveys because of its brevity.

The goal of the present paper is to develop an algorithm to compute physical and mental scale scores using SOEP data (which includes a specific version of the SF-12v2 questionnaire) and to establish representative health scores for the population in Germany on this basis, which may serve as a benchmark values for similar analyses.[3] The SOEP is one of the largest representative surveys in Germany and is therefore provides a very appropriate basis for developing an algorithm to calculate scales analogous to SF-12v2. The SOEP results for the year 2004 can also be used as German reference values (norm sample) for all further surveys conducted using this questionnaire, just as the 1998 US population forms the norm sample for calibration of almost all original SF instruments. In comparison to the vast majority of comparable surveys both in Germany and internationally, SOEP has the distinct advantage of its large sample size of over 20,000 cases.

In the following, we describe this algorithm. We also provide it as an SPSS syntax file that can be imported and used directly, along with accompanying descriptive comments.

---

1 An objective non-invasive measurement of grip strength was added as well in 2006 (cf. Hank et al., 2006).

2 See http: // www.qualitymetric.com / products / sf-12v2.aspx.

3 See Nübling et al. 2006 for a German version of the technical details of the SF-12v2.

2. SF-12v2 in the SOEP

The SF-12v2 questionnaire (called SF-12 in this paper) was developed on the basis of its predecessor, the SF-36 short questionnaire on health-related quality of life (see Ware et al., 2001). SF-36 consists of a total of 36 individual questions that can be grouped first into eight subscales, and second into the two superordinate dimensions of physical and mental health. SF-12 contains only 12 of the original 36 items, but nevertheless encompasses the complete field covered by the eight subscales and two superordinate dimensions "physical health" and "mental health" with either one or two items each.

The SOEP questionnaires (up to now: in the years 2002, 2004 and 2006) contain a battery of questions on health-related quality of life (2004: questions 83 – 86), which are modeled on version 2 of the SF-12 questionnaire (SF-12v2). The SOEP version deviates from the original SF-12v2 to some degree in the formulation and order of questions and in general layout.[4] Furthermore, SOEP question 86_5, "severe physical pain", was taken from SF36 and was not contained in the original SF-12v2. In contrast, the item in the original SF-12v2 "work interference due to pain" is missing in the SOEP version. The questions in the SOEP questionnaire are very similar, but not identical, to SF-12v2.

Because of these differences between SOEP-SF-12 and the original version, we develop our computational algorithm based on the empirical results of the 2004 SOEP data and not based on the 1998 US norm sample. Furthermore, the 1998 US data are probably only applicable to a very limited extent to the population of the Federal Republic of Germany today. Finally, the SOEP sample, with more than 20,000 respondents, is larger than most of the older norm samples in both Germany and abroad.

3. On the Selectivity of the Norm Data Set

The SOEP, just like any other cross-sectional or longitudinal survey, is plagued by the problem of some individuals declining to participate. This may be relevant to the present study since the possibility cannot be ruled out that unit non-response is due to individual health problems. We argue, however, that the risk of selectivity is, in all likelihood, small – or at least smaller than in comparable surveys.

First, in the stage of sampling first wave respondents, the SOEP invests enormous efforts in obtaining a non-selective response – more than comparable surveys – for example, by using multiple contact trials. Second, previous

[4] For the documentation of the questionnaires see: http://www.diw.de/english/sop/service/fragen/index.html.

SOEP-based research suggests that participants in an ongoing panel have a very high probability of staying in the sample even when their living conditions worsen, a finding which would seem to apply to health status as well. Corroborating this conclusion, the annual SOEP attrition analysis suggests that the health status of respondents has only a marginal effect on their probability of dropping out or remaining in the sample (cf. Kroh / Spiess, 2006). Hence, the longitudinal structure of SOEP ensures a good representation of respondents with poor health than any cross-sectional study.

To avoid distortions due to missing values and thus differing case numbers in the individual steps of the analysis, we only used the data provided by SOEP2004 respondents who had given valid answers for all 12 variables.[5] Overall the item non-response quota on the SF-12 items for the whole sample in 2004 is 3.5 %. In order to check the magnitude of the non-randomness of the process that determined missing values, we ran a logit analysis. Table 1 displays the result of this analysis, where the dependent variable indicates item non-response in at least one of the 12 health variables.

Three characteristics were identified as playing an important role in item non-response. First, men showed a significantly higher probability than women of answering the twelve health items. Second, individuals living in the eastern part of Germany were more likely to respond than those in the western part of the country. Third, people with a lower subjective health satisfaction level showed a higher probability of answering the SF-12 categories without item non-response. However, the explained variance is very low, and thus, the magnitude of the non-random component is small. Furthermore, the share of missing values is low as well, so the overall impact of missing values on the norm sample is negligible. Table 1b shows the mean values of the relevant health-related variables for the full sample and the norm sample. The differences are marginal, although the respondents with missing values have a few characteristics that differ from the overall sample. One should keep in mind too that the norm values will be computed for men and women separately, so this difference in the levels of missing values will not affect the results.

[5] The US manual mentions the use of special online software for the estimation and imputation of missing values, but it neither provides the relevant algorithms nor refers to any other sources where they have been published or made available.

Table 1a

Logit Analysis on Item Non-Response in the SF-12v2 Health Indicators

	Exp(B)
Age	0.985
Age (squared)	1.000
Men	0.830*
Living in East Germany	0.682**
Highest post-secondary Degree[a)]	
Hauptschule – or no Degree	0.981
Fachhochschule	1.078
Abitur	0.890
Other Degree	1.151
Satisfaction with Health	0.911**
Health Rating, Subjective[b)]	
Very Good	0.779
Sufficient	1.315*
Not so Good	1.343
Bad	1.332
Disabled[c)]	1.017
Number of Doctor Visits[d)]	
3 to 4	0.905
5 or More	0.896
Size of Household[e)]	
Two to Three Persons	0.945
Four Persons or More	0.835
Nagelkerkes R^2	0.028

Reference categories: [a)] Realschule degree; [b)] good; [c)] not disabled; [d)] one or two; [3)] single person.

Table 1b

Distribution of Population[a)] in 2004 by Item Non-Response of SF-12v2 Health Indicators

	SF-12 Population (norm sample)	Total SOEP Population	Subpopulation with Item-Non-Response in SF-12v2
Age (Average)	48	49	54
			in %
Women	53	54	62
Living in West Germany	81	81	86
Highest Post-Secondary Degree			
Hauptschule or no Degree	42	42	49
Realschule	28	28	26
Fachhochschule	5	5	4
Abitur	19	18	15
Other Degree	6	6	6
Satisfaction with Health (Average)	6,5	6,4	5,2
Health Rating, Subjective[4]			
Very Good	10	10	6
Good	38	38	25
Sufficient	32	32	34
Not so Good	15	15	26
Bad	5	5	9
Disabled	13	14	21
Number of Doctor Visits			
Never	31	31	28
One to Two Times	35	35	33
3 Times or More	33	34	38
Size of Household			
One Person	22	22	30
Two to Three Persons	55	55	54
Four Persons or More	23	23	16

[a)] weighted results.

4. Developing a Computational Algorithm

The method used to compute scale values for the eight subscales of SF-12v2 and the two main dimensions (PCS and MCS)[6] was designed to correspond as closely as possible to the procedure used with the original SF-12v2 and described in the US manual (Ware et al., 2002). In the personal questionnaire of the SOEP not every item does correspond to the original item set of the SF-12; rather, the SOEP group developed functionally equivalent questions that fit into the health section of the longitudinal survey. It should be noted that the evaluation of Version 2 of the SF-12 follows new rules of computation and thus does not correspond to the methodologies used for SF-12v1 and SF-36.

4.1 Preparing for the Computation

If necessary, the numerical values of three individual questions were inverted, as in the original. We refrained from recalibrating the item "general health" as was done on the US original since no reliable calibration data were available.

4.2 Creating the Values for the Eight Subscales

Four of the eight subscales in SF-12 consist of one item each; the other four consist of two items each (Table 1). For each subscale, a mean value was first computed and transformed to a position on a scale ranging from 0 to 100 (*z*-transformation). Then the four individual items were each transformed directly into a value between 0 and 100, and for the subscales with two items each, the mean value of the two items was computed (arithmetic mean).

For the $N = 21248$ persons surveyed in SOEP2004 who had provided valid answers for all the questions in SF-12, we arrived at the mean values and standard deviations given in Table 2.

As with the original SF-12v2, we standardized the raw values of the mean scale values through *z*-transformation (mean value = 0, standard deviation = 1). To improve manageability, we then subjected them to linear transformation to a mean value of 50 and a standard deviation of 10 –. With this, the process of "norm-based scoring" of the eight subscales was completed.

In the SOEP sample, the total mean value for each scale is 50 points and the standard deviation is 10 points; in the subgroups and future studies, the values determined can be measured against these values. A value of 50 would mean that the empirical comparative value lies precisely at the mean value of SOEP2004; a value of, for example, 60 would mean that it lies exactly one standard deviation beyond the norm mean value.

[6] PCS means "Physical Component Summary Scale" ("Physical Health"), MCS means "Mental Component Summary Scale ("Mental Health").

Table 2

Mean Scale Values in SF-12 (SOEP Version SF-12v2)

	N	Mean Value	Standard Deviation
Physical Functioning (0 – 100 score value, 2 vars)	21248	73.14	32.145
Role Physical (0 – 100 score value, 2 vars)	21248	74.29	26.460
Bodily Pain (0 – 100 score value, 1 var)	21248	73.27	27.142
General Health (0 – 100 score value, 1 var)	21248	60.57	24.083
Vitality (0 – 100 score value, 1 var)	21248	52.94	22.842
Social Functioning (0 - 100 score value, 1 var)	21248	83.22	23.568
Role Emotional (0 – 100 score value, 2 vars)	21248	81.94	22.350
Mental Health (0 – 100 score value, 2 vars)	21248	61.96	20.465

4.3 Creating the Two Superordinate Scales PCS and MCS

We then grouped the eight subscales under the two superordinate scales "physical health" (PCS) and "mental health" (MCS). We followed the same procedure here as described in the US manual, using explorative factor analysis (PCA, varimax rotation) with the eight z-transformed subscales. This resulted – as with the US norm population – in a two-factor solution, where four scales were assigned to each of two factors.

The resulting structure again confirmed that with the subscales, the assumed and repeatedly confirmed classification applies to the SOEP data as well: the four physical scales clearly belong to the physical factor (Factor 1), and the four mental scales clearly belong to the mental factor (Factor 2). Sixty-eight percent of the variance is explained by the two-factor solution. The rotated factor loadings of the eight subscales on the two superordinate factors are provided in Table 3.

The results of this factor analysis with the SOEP2004 data as the norm population serve as the basis for weighting the eight subdimensions within the two main dimensions. In future studies, the "factor score coefficients" determined in the present analysis can be used rather than the factor loading in Table 3.[7]

In contrast to previous studies by Grabka and Schupp (2005) who conducted separate analyses for each of four subscales, the secondary loading – that is, the loading from the subscales on the non-primary factors – were taken into account for computing the factor values as described in the US manual.

[7] This procedure is unfortunately not described adequately in the US handbook. Our personal investigation of "Qualitymetric" revealed, however, that the same procedure was used in the US case as well.

Table 3

Factor Loadings of the Eight Subscales on the Two Main Dimensions

	Component	
	1	2
Physical Fitness (factor, 2 vars)	**.857**	.152
General Health (inverted, 1 var)	**.789**	.285
Bodily Pain (1 var)	**.788**	.276
Role Physical (factor. 2 vars)	**.779**	.405
Mental Health (factor, 2 vars)	.091	**.839**
Role Emotional (factor, 2 vars)	.311	**.772**
Social Functioining (1 var)	.358	**.727**
Vitality (inverted, 1 var)	.303	**.596**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser normalization.

The raw values of the two superordinate scales (1 = physical health, 2 = mental health) are computed based on the values given in Table 4.

Corresponding to the procedure used for the eight subscales (see step 1, above), the values of the main dimensions were *z*-transformed to a mean value of 50 and a standard deviation (SD) of 10. Thus, for the total population, the mean value of each of the two dimensions is 50 points and the SD is 10 points.

Table 4

Factor Score Coefficients of the Eight Subscales on the Two Main Dimensions

	Component	
	1	2
Physical Functioning (0 – 100 score value, 2 vars)	.414	–.209
Role Physical (0 – 100 score value, 2 vars)	.279	–.021
Bodily Pain (0 – 100 score value, 1 var)	.331	–.105
General Health (0 – 100 score value, 1 var)	.330	–.103
Vitality (0 – 100 score value, 1 var)	–.041	.258
Social Functioning (0 – 100 score value, 1 var)	–.068	.333
Role Emotional (0 – 100 score value, 2 vars)	–.110	.378
Mental Health (0 – 100 score value, 2 vars)	–.244	.489

Source: SOEP2004 (unweighted); own calculations.

This step concludes the scoring algorithm. In the appendix of the data documentation in Nübling et al. (2006), the SPSS syntax is provided for all the values required for analyses with SF-12, that is, for both raw values as well as *z*-transformed values and the values after norm-based scoring.

5. Scale Values for Physical and Mental Health According to Age and Gender

Table 5 displays the mean values for physical and mental health for all SOEP2004 respondents (who are representative for all of Germany), broken down according to age and gender. We only report the unweighted figures as weighting for unit non-response produces practically identical results.

Table 5

SOEP SF-12 Values (normed to SOEP full sample 2004)

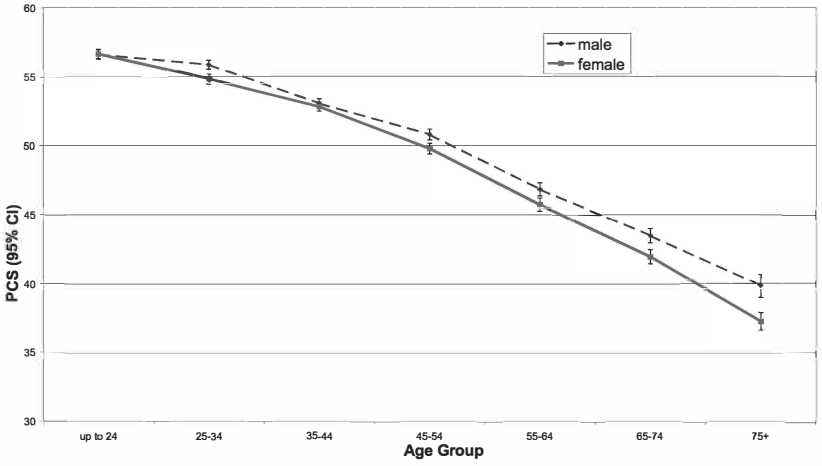
Population	*Physical Health (mean (SD))*	*Mental Health (mean (SD))*
Total (*N* = 21248)	50.00 (9.99)	50.00 (10.00)
Gender		
Males (10236)	50.55 (9.72)	51.14 (9.63)
Females (11012)	49.49 (10.22)	48.94 (10.21)
Age groups		
18 – 24 (2552)	56.62 (6.20)	50.03 (9.37)
25 – 34 (3021)	55.30 (6.74)	48.89 (9.47)
35 – 44 (4441)	52.95 (8.02)	49.27 (9.67)
45 – 54 (4000)	50.28 (8.88)	49.57 (9.79)
55 – 64 (3267)	46.29 (9.80)	51.18 (10.14)
65 – 74 (2590)	42.73 (9.57)	51.99 (10.43)
75 + (1377)	38.28 (9.66)	49.46 (11.82)

Source: SOEP2004 (unweighted); own calculations.

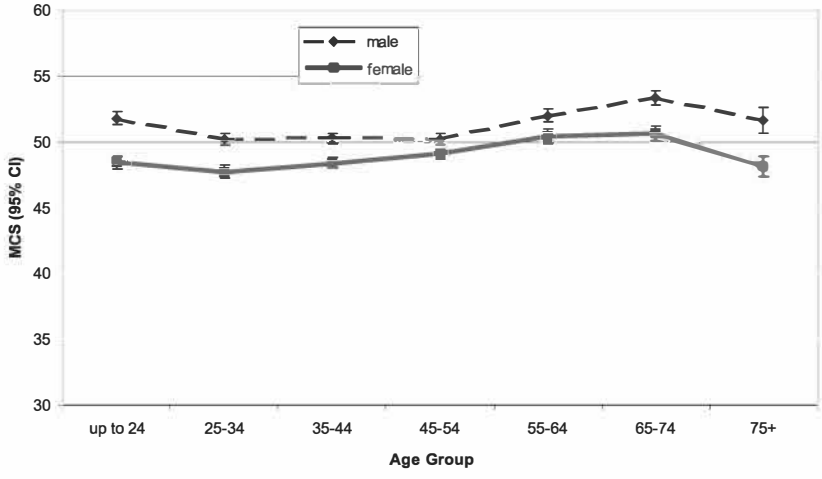
Overall, the differences according to age group and gender are extremely close to those found for the US norm population (SF-36).

For the physical health scale, we find a strong and constant decrease in values according to age of respondents, from nearly 57 points among those under the age of 25 to just 38 points among 75-year-olds.

With mental health, however, we find very low variation with age and, in contrast to the US values for SF-36, no clear linear age effect.



Graph 1: PCS by Age Groups and Gender



Graph 2: MCS by Age Groups and Gender

However, gender differences are more marked in the mental health dimension: men show total mean values two points higher than women (Table 4) which are consistent across all age groups. With physical health, there is just a one-point difference (again favoring men), but this increases with age, and in the lower age groups, few to no differences can be identified.

6. Discussion and Conclusions

The "SF family" of questionnaires is used extensively worldwide. They are thoroughly tested and established instruments for measuring the health-related quality of life.

In the SOEP – which is not just a representative sample of the German population but also a large-scale random sample – a specific version of SF-12v2 was used in 2002, 2004 and 2006 that may also be employed again in the future. In the present study, the scoring algorithm used with the original SF-12v2 has been adapted to the specific SOEP version. Thus we have derived firmly established norm values for the subsamples and main scales of SF-12 (SOEP version of SF-12v2).

The algorithm is also provided as open access code in SPSS syntax. With this algorithm, it will be possible for future studies using the questions from the SOEP version of SF-12v2 to compare the results for their survey populations directly with the SOEP survey data, that is, with the German mean values. Furthermore, when using the technique of norm-based scoring, these values derived with other SF versions (where NBS has been used) are also comparable with these values, which would not be the case under use of raw scale values.

References

Grabka, M. / *Schupp,* J. (2005): Befragungsgestützte Messung von Gesundheit – Bestandsaufnahme und Ausblick. Kapitel 6 der DIW Eventdokumentation 2 – 2005, Berlin.

Hank, K. / *Jürges,* H. / *Schupp,* J. / *Wagner,* G. G. (2006): Die Messung der Greifkraft als objektives Gesundheitsmaß in sozialwissenschaftlichen Bevölkerungsumfragen, DIW Discussion Papers No. 577, Berlin.

Kroh, M. / *Spiess,* M. (2006): Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2005), DIW Data Documentation No. 15, Berlin.

Nübling, M. / *Andersen,* H. H. / *Mühlbacher,* A. (2006): Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP-Version des SF-12. DIW Data Documentation 16, Berlin.

Rendtel, Ulrich (1995): Lebenslagen im Wandel – Panelausfälle und Panelrepräsentativität, Frankfurt am Main / New York.

Schupp, J. / *Wagner,* G. G. (2007): New Concepts and Survey Questions in the German Socio-Economic Panel Study (SOEP) – Principles, DIW Berlin Data Documentation, Berlin.

Wagner, G. G. / *Frick,* J. R. / *Schupp,* J. (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements, Schmollers Jahrbuch 127 (1), in this volume.

Ware, J. E. / *Kosinski,* M. / *Dewey,* E. D. (2001): How to Score Version Two of the SF-36® Health Survey (standard and acute forms), Lincoln, RI: Qualitymetric Incorporated, 3rd edition.

Ware, J. E. / *Kosinski,* M. / *Turner-Bowker,* D. M. / *Gandek,* B. (2002): How to Score Version 2 of the SF-12® Health Survey (with a supplement documenting Version 1), Lincoln, RI: Qualitymetric Incorporated.