

¹ Zentrum für Psychosoziale Medizin, Institut und Poliklinik für Medizinische Psychologie, Universitätsklinikum Hamburg-Eppendorf, Deutschland

² Selbständige Abteilung für Medizinische Psychologie und Soziologie, Universität Leipzig, Deutschland

Die Version 2.0 des SF-36 Health Survey – Ergebnisse einer bevölkerungsrepräsentativen Studie

Submitted: 3 September 2004

Accepted: 17 June 2005

Published Online First: 21 September 2005

Summary

The version 2.0 of the SF-36 Health Survey: results of a population-representative study

Objectives: The SF-36 is an internationally approved instrument to assess subjective health. In its second version the SF-36 contains modifications to version 1.0 in several item formulations and response categories. The second version is translated in German but a psychometric analysis and standardisation are still lacking.

Methods: The SF-36, version 2.0, was administered in a representative German sample ($n = 2043$), analysed concerning its scale structure, reliability, dissemination characteristics and differences in traits of populations and compared to version 1.0 (German health survey).

Results: Version 2.0 shows high reliability in subscales and has a scale structure which is comparable to that of version 1.0. The scale values in the actual population-representative sample survey are rather distributed obliquely and effects of sex, age and residence (Eastern/Western Germany) have been found in regression analyses.

Conclusion: The psychometric comparability of both versions is given but should be interpreted with caution as regards content because of the different enquiry periods. Though the expected gain in psychometric quality and discrimination of version 2.0 to version 1.0 is evident, but does not justify a preference to version 2.0.

Key words: SF-36 – Version 2.0 – Representative German sample – Reliability – Validity.

Der SF-36 Health Survey, der aus einer langen Tradition von Skalenentwicklungen im Rahmen der Medical Outcomes Study

entstanden ist, stellt eines der international meist untersuchten und weit gebräuchlichsten Verfahren zur Erfassung der subjektiven Gesundheit bzw. der gesundheitsbezogenen Lebensqualität dar (Kirchberger 2000). Eine aktuelle Literaturübersicht zählt im Zeitraum zwischen 1993 bis 2004, d. h. in den letzten 10 Jahren eine Anzahl von über 3 300 Arbeiten mit dem SF-36 (Zugriff Medline, Januar 2005). Bereits früh in der Entwicklungsgeschichte des Instruments wurde seine Übersetzung, psychometrische Prüfung und Normierung in verschiedenen Sprachen betrieben, wobei inzwischen die SF-36 Version 1.0 in 31 Sprachen vorhanden ist und aus über 20 Ländern Normdatensätze vorhanden sind.

Die Erfahrungen mit dem SF-36, Version 1.0, in Deutschland zeigen, dass es sich um ein psychometrisch ausgesprochen robustes Verfahren handelt, sowohl hinsichtlich der Reliabilität als auch der Validität (Bullinger 1955; Bullinger et al. 2003). Dennoch bestanden in der Version 1.0 des SF-36 einige sprachliche Unklarheiten und wie sich empirisch zeigte, eine zum Teil mangelnde Differenzierung der Antwortkategorien, die eine Überarbeitung des SF-36, Version 1.0, nahe legte (Zwingmann et al. 1998). Die Veränderungen sind allerdings eher ergänzender oder präzisierender Natur, es wurden weder neue Items eingefügt noch die Zuordnung der Items zu den Skalen geändert.

Die International Quality of Life Assessment Group (IQOLA) hat als internationale SF-36-Arbeitsgruppe in einer Serie von Artikeln die Übersetzung, Prüfung und Normierung des SF-36, Version 1.0, in verschiedenen Sprachen dargestellt (IQOLA 1998). Es zeigte sich im Rahmen der internationalen Arbeit mit dem SF-36, dass einige der Items suboptimal formuliert und in den Antwortkategorien verbesserungswürdig sind. Dies wurde unterstützt von empirischen Analysen des wachsenden Datensatzes, deren Ergebnisse in verschiedenen IQOLA-Arbeitsitzungen thematisiert wurden. Dies führte zu einer Veränderung des englischen Originals, die in Tabelle 1, Vergleich Version

Tabelle 1 Änderung im SF-36 V2.0 gegenüber der Version 1.0

Item	SF-36 Version 1.0	SF-36 Version 2.0
3.h	„mehrere Strassenkreuzungen“	ersetzt durch „mehrere hundert Meter“
3.i	„eine Strassenkreuzung“	ersetzt durch „einhundert Meter“
4 a–d	Hatten Sie in den vergangenen 4 Wochen... → geändert in	Wie oft hatten Sie in den vergangenen 4 Wochen...
4 a–d	„Ja/Nein“-Antwortkategorien „1–2“	„Immer/Meistens/Manchmal/Selten/Nie“ „1–5“
5 a–c	Hatten Sie in den vergangenen 4 Wochen... → geändert in	Wie oft hatten Sie in den vergangenen 4 Wochen...
5 a–c	„Ja/Nein“-Antwortkategorien „1–2“	„Immer/Meistens/Manchmal/Selten/Nie“ „1–5“
9.a	„voller Schwung“ →	„voller Leben“
9 a–i	Immer/Meistens/Ziemlich oft/ Manchmal/Selten/Nie (1–6)	Immer/Meistens/ – / Manchmal/Selten/Nie (1–5)
11.b	„Ich bin genauso gesund wie alle anderen , die ich kenne“	„Ich bin genauso gesund wie andere Menschen , die ich kenne“
11.c	„Ich erwarte, dass meine Gesundheit nachlässt“	„Ich erwarte, dass mein Gesundheitszustand sich verschlechtert“
11.d	„Ich erfreue mich ausgezeichneter Gesundheit“	„Mein Gesundheitszustand ist ausgezeichnet“

2.0 vs. 1.0 aufgeführt ist (Ware et al. 2000; 2003). Im Rahmen der IQOLA-Gruppe wurde dann auch die Übersetzung der revidierten Items vorangetrieben, wobei das Übersetzungs-procedere (zwei Vorwärts-, eine Rückwärts-Übersetzung) mit internationaler Einigung eingehalten wurde; es liegen Übersetzungen in 12 Sprachen vor.

Wie Tabelle 1 zeigt, wurden gegenüber der Version 1.0 folgende Änderungen vorgenommen: Im Item 3.h wurde die Formulierung „mehrere Strassenkreuzungen“ durch „mehrere hundert Meter“ sowie in 3.i „eine Strassenkreuzung“ durch „einhundert Meter“ ersetzt und somit eher dem europäischen Verständnis von Distanzen entsprochen.

Die in den Items der Skalen Körperliche (Items 4a–d) und Emotionale (Items 5a–c) Rollenfunktion erfragten Einschränkungen wurden von einer reinen Abfrage des grundsätzlichen Vorhandenseins in eine quantitative Abfrage verändert. Es heisst nunmehr in der Frage nicht mehr „Hatten“ sondern „Wie oft hatten Sie ...“. Die entsprechenden Antwortkategorien in den Items 4 und 5 wurden von einem dichotomen „Ja/Nein“ in die Kategorien „Immer/Meistens/Manchmal/Selten/Nie“ überführt.

Darüber hinaus ist die Formulierung „voller Schwung“ in der Frage 9.a in „voller Leben“ verändert worden. Die diese Fragen betreffenden Antwortkategorien sind in der Version 2.0 um die Möglichkeit „Ziemlich oft“ reduziert worden.

Weiterhin wurden die Items der Skala Allgemeine Gesundheitswahrnehmung grundlegend überarbeitet: Die Formulierung in 11.b wurde dahingehend verändert, dass „alle anderen“ durch „andere Menschen“ ersetzt wurde, in 11.c heisst es nun „mein Gesundheitszustand sich verschlechtert“ und 11.d lautet in der Version 2.0 „Mein Gesundheitszustand ist ausgezeichnet“ (Bullinger & Morfeld 2004). Zusätzlich wurden Veränderungen am Layout des Fragebogens sowie einige Hervorhebungen in den Itemformulierungen modifiziert (Ware 2004).

Obwohl die SF-36 Version 2.0 bereits Ende der 1990er-Jahre in mehreren Sprachen vorlag, begann der Einsatz des SF-36, Version 2.0, in Studien nur zögerlich. Gründe dafür waren einerseits die Frage, inwieweit die mit der Version 1.0 vorhandenen Ergebnisse und Referenzwerte auch auf Version 2.0 übertragbar sind und andererseits die Skepsis, inwieweit die Version 2.0 tatsächlich deutliche Verbesserungen zu Version 1.0 sowohl psychometrisch als auch inhaltlich darstellt.

In Deutschland lag die Übersetzung der Items der Version 2.0 bereits 1998 vor, dennoch wurde aufgrund noch unklarer Copyright-Situation und fehlender Datenbasis für die Version 2.0 die Entscheidung getroffen, im deutschen Gesundheitssurvey die Version 1.0 einzusetzen (Radoschewski & Bellach 1999), die damals bereits sehr viel gebräuchlicher und besser dokumentiert war. Mit leichter zeitlicher Verzögerung wurde es im Jahre 2002 möglich, im Rahmen einer Serie repräsentativer Bevölkerungsstudien zum Thema „psychologische Messinstrumente“ auch den SF-36, Version 2.0, einzusetzen und zu prüfen.

Der folgende Artikel beschreibt Methodik und Ergebnisse dieses Einsatzes und diskutiert diese Ergebnisse vor dem Hintergrund vorliegender Informationen zur psychometrischen Qualität und inhaltlichen Aussagekraft des SF-36, Version 2.0.

Folgende Fragestellungen wurden untersucht:

1. Welche psychometrischen Kennwerte besitzt der SF-36, Version 2.0 (Reliabilität, faktorielle Validität und Skalencharakteristika)?
2. Gibt es Unterschiede in der Lebensqualität nach soziodemographischen/psychosozialen Faktoren, die bereits mit Version 1.0 des SF-36 gefunden wurden, unter Berücksichtigung des Wohnortes (alte vs. neue Bundesländer)?
3. Wie lässt sich die subjektive Gesundheit der deutschen Bevölkerung mit Hilfe des SF-36, Version 2.0, im Vergleich mit den Ergebnissen der vorhergehenden repräsentativen Befragung mit dem SF-36, Version 1.0, beschreiben?

Bezüglich der letzten Frage ist bereits an dieser Stelle zu betonen, dass ein Vergleich nur sehr stark eingeschränkt möglich ist. Beide Untersuchungen wurden zu unterschiedlichen Zeitpunkten bei unterschiedlichen Populationen durchgeführt. Trotzdem erscheint ein Vergleich sinnvoll, um mögliche Unterschiede darstellen zu können.

Methoden

Der SF-36, Version 2.0, wurde im Rahmen einer Mehrthemenbefragung der in Privathaushalten lebenden deutsche Wohnbevölkerung im Alter ab 14 der Universität Leipzig (Leitung: Prof. Dr. Elmar Brähler) eingesetzt. Die Daten der vorliegenden Untersuchung wurden im Herbst 2002 im Auftrag der Universität Leipzig von dem Meinungsforschungsinstitut

Tabelle 2 Soziodemographische Charakteristika der Untersuchungsstichprobe

Studie	Mehrthemenumfrage SF-36 v 2.0	Bundesgesundheitsurvey RKI 1998
N	2043	7124
Alter	48,93 ± 18,13 Jahre	45,72 ± 15,91 Jahre
Geschlecht		
männlich	960 (47,0 %)	3450 (48,4%)
weiblich	1083 (53,0 %)	3674 (51,6%)
Altersgruppen		
≤ 20 J	130 (6,4%)	368 (5,2%)
21 - 30 J.	262 (12,8%)	1054 (14,8%)
31 - 40 J.	331 (16,2%)	1559 (21,9%)
41 - 50 J.	350 (17,1%)	1309 (18,4%)
51 - 60 J.	318 (15,6%)	1353 (19,0%)
61 - 70 J.	376 (18,4%)	1004 (14,1%)
71 - 80 J.	225 (11,0%)	477 (6,7%)
81 und älter	51 (2,5%)	–
Familienstand		
ledig	500 (24,5 %)	1549 (22,3%)
verheiratet	1012 (49,5 %)	4432 (63,9%)
getrennt lebend	24 (1,2 %)	118 (1,7%)
geschieden	220 (10,8 %)	405 (5,8%)
verwitwet	287 (14,0 %)	427 (6,2%)
Ost-West-Zugehörigkeit		
Ost	1009 (49,4 %)	2419 (34%)
West	1034 (50,6 %)	4705 (66%)
Haushalts-Nettoeinkommen		
bis unter 500 /Monat	43 (2,2 %)	146 (2,6%)
500 bis unter 750 /Monat	170 (8,7 %)	314 (5,6%)
750 bis unter 1.000 /Monat	241 (12,3 %)	433 (7,8%)
1.000 bis unter 1.250 /Monat	318 (16,2 %)	545 (9,8%)
1.250 bis unter 1.500 /Monat	294 (15,0 %)	708 (12,7%)
1.500 bis unter 2.000 /Monat	414 (21,2 %)	1324 (23,8%)
2.000 bis unter 2.500 /Monat	273 (13,9 %)	1011 (18,2%)
2.500 und mehr/Monat	204 (10,4 %)	1086 (19,5%)

Bemerkung: Zugrundegelegt wurden für die Analysen mit dem BGS die vom RKI bereit gestellten fallbezogenen Gewichte.

USUMA Berlin im Rahmen einer bevölkerungsrepräsentativen Mehrthemenumfrage erhoben. Die in die Studie einbezogenen Personen wurden von geschulten Interviewern zu Hause aufgesucht und dort befragt. Im Rahmen der Interviews wurden den Probanden Fragebögen zur selbständigen Beantwortung vorgelegt. Die Ausschöpfungsquote der Erhebung lag bei 66,5%. An der Untersuchung nahmen 2066 Personen im Alter von 14 bis 95 Jahren teil, wobei sich die folgende Analyse auf diejenigen Befragten mit deutscher Staatsangehörigkeit beschränkt (es wurden 23 ausländische Personen befragt) (Albani et al. 2005). Angaben zu den soziodemographischen Merkmalen der verbleibenden 2043 Probanden (1009 Befragte in Ostdeutschland und 1034 in Westdeutschland) sind Tabelle 2 zu entnehmen. Nach Dateneingabe und Plausibilitätscheck wurden die Daten dem Institut für Medizinische Psychologie der Universität Hamburg zur Verfügung gestellt. Die Daten des Bundesgesundheits surveys wurden als Public Use File angefordert (zur SF-36 Version 1.0 im Bundesgesundheits survey (Radoschewski et al. 1999; Thefeld et al. 1999). Dort erfolgte eine sowohl deskriptiv statistische Auswertung mit Hilfe des SPSS-Programmes als auch eine psychometrische Analyse speziell der Reliabilität und der faktoriellen Validität mit Hilfe des Multitrait Analysis Program (MAP) (Hays et al. 1998). Dieses Programm basiert auf dem „Multi-trait/Multi-item-Ansatz“ von Campbell und Fiske (1959) und erlaubt eine konfirmatorische Prüfung der hypothetischen Skalenstruktur (Skalenfit) der Items eines Fragebogens. Es wird angegeben, in wie viel Prozent der Fälle ein Item mit seiner eigenen Skala höher korreliert als mit anderen Subskalen. Der sogenannte Skalenfit prüft die Anzahl der Fälle, in denen ein Item mit seiner Skala signifikant höher oder signifikant höher korreliert als mit einer anderen Skala. Der Skalenfit ist somit ein Indikator für die faktorielle Validität und ist bei 100% optimal. Weiterhin liefert das Programm Angaben über die interne Konsistenz (Cronbachs Alpha) der Subskalen, deren Charakteristika und

die Präsenz von Boden- und Deckeneffekten (Bullinger & Kirchberger 1998).

Die Skalenwerte des SF-36, Version 2.0, wurden transformiert (Campbell & Fiske 1959) auf einer Skala von 0–100 nach der Formel

$$\text{Transformierte Skala} = \frac{(\text{tatsächlicher Rohwert} - \text{niedrigstmöglicher Rohwert})}{\text{mögliche Spannweite der Rohwerte}} \times 100$$

(Bullinger & Kirchberger 1998).

Darüber hinaus wurden Vergleiche hinsichtlich z. B. Geschlechts- und Altersunterschiede im SF-36 mit Hilfe parametrischer Verfahren (t-Test, Varianz- und Regressionsanalyse u. a.) berechnet.

Die Stichprobe bestand aus n = 2043 Personen ab 14 Jahren mit einem Durchschnittsalter von 48,9 (+/-18,13) Jahren. 960 Personen (47,0%) waren männlich, 1083 (53,0%) waren weiblich. Die Altersgruppe 61–70 Jahre war mit 18,4% der Befragten am häufigsten vertreten. Etwa die Hälfte der Personen (1009 = 49,4%) kamen aus den alten Bundesländern, 1034 (50,6%) aus den neuen Bundesländern, 49,5% der Population war verheiratet. Beim Haushaltseinkommen zeigte sich, dass sich der Hauptteil der Befragten einer Einkommensspanne Haushaltsnetto-Einkommen von 1250 bis 2499 Euro pro Monat (48%) einordnete. Einen Überblick über soziodemographische Charakteristika gibt Tabelle 2, in der auch die Stichprobenverteilung aus dem Bundesgesundheits survey dargestellt ist, in dem der SF-36, Version 1.0, eingesetzt wurde (Radoschewski & Bellach 1999).

Der statistische Vergleich zwischen beiden Stichproben weist darauf hin, dass in der hier im Mittelpunkt stehenden Untersuchung zum SF-36 V 2.0 signifikant mehr Personen in den neuen Bundesländern ($\chi^2 = 161,52$, $df = 1$, $p \leq .001$) befragt wurden. Darüber hinaus besteht ein signifikanter Altersunterschied zwischen den Teilnehmern der Untersuchungen, im

Tabelle 3 Fehlende Werte und psychometrische Kennwerte im SF-36 V2.0

Subskalen	Fehlende Daten %	Boden %	Decke %	Skalenfit %	Interne Konsistenz
Körperliche Funktionsfähigkeit (KöFu) ¹	0,2 (2,5)	0,2 (0,5)	51,9 (36,7)	100,0 (100)	0,94 (0,92)
Körperliche Rollenfunktion (KöRo)	0,1 (3,0)	0,6 (9,9)	46,5 (71,8)	100,0 (100)	0,92 (0,89)
Körperliche Schmerzen (Schm)	0,1 (2,3)	0,0 (0,0)	45,0 (27,3)	100,0 (100)	0,89 (0,82)
Allg. Gesundheitswahrnehmung (AGes)	0,1 (2,7)	0,1 (0,1)	3,1 (0,9)	97,5 (100)	0,81 (0,71)
Vitalität (Vita)	0,2 (2,5)	0,2 (0,1)	1,9 (1,0)	100,0 (100)	0,81 (0,77)
Soziale Funktionsfähigkeit (SoFu)	0,6 (2,2)	0,2 (0,3)	57,5 (55,8)	100,0 (100)	0,81 (0,77)
Emotionale Rollenfunktion (EmRo)	0,2 (3,0)	0,7 (6,0)	53,8 (82,9)	100,0 (100)	0,93 (0,83)
Psychisches Wohlbefinden (Psych)	0,1 (2,5)	0,1 (0,1)	1,7 (2,2)	97,5 (97,5)	0,83 (0,81)

In Klammern die entsprechenden Werte der Version 1.0 aus dem Bundesgesundheits survey

¹ In den Klammern befinden sich die im deutschen Sprachraum eingeführten Kurzbenennungen der Skalen, international sind andere Kurzformen gebräuchlich, in Reihenfolge wie in Tabelle 3 lauten diese: pfi, rolph, pain, ghp, vital, social, rolem, mhi.

Mittel sind die Teilnehmer am Bundesgesundheitsurvey fast drei Jahre jünger ($F = 60,89$; $df = 1$; $p \leq .001$). Bezüglich der Geschlechterverteilung zeigt sich kein Unterschied.

Ergebnisse

Wie die Tabelle 3 zeigt, ist der Anteil fehlender Werte in den Subskalen des SF-36, Version 2.0, in der Gesamtpopulation ausgesprochen gering, meist um 0,1%, nur bei der Skala Soziale Funktionsfähigkeit liegt er bei 0,3%. Während die Bodeneffekte (d. h. Einschränkungen in der Ausnutzung der gesamten Skalenbreite) ebenfalls ausgesprochen gering sind, zeigt sich in den meisten Skalen ein deutlicher Deckeneffekt und damit eine als hoch selbstberichtete Gesundheit sowohl hinsichtlich körperlicher Funktionsfähigkeit und Rollenfunktion als auch hinsichtlich der emotionalen Rollenfunktion, der sozialen Funktionsfähigkeit und der körperlichen Schmerzen. Bei den in Klammern angegebenen Kennwerten der Version 1.0 aus dem Bundesgesundheitsurvey zeigen sich Differenzen zur Version 2.0, besonders der Skalen KöRo und EmRo (beide hervorgehoben). Bei diesen Skalen wurde in der Version 2.0 das Antwortformat von zwei auf fünf Stufen geändert. Ob dies mit den nun deutlich niedrigeren Deckeneffekten der Version 2.0 ursächlich zusammenhängt, bleibt noch zu diskutieren.

Bezüglich der psychometrischen Charakteristika ergibt sich ein Skalenfit von 100% für 6 der 8 Skalen, für allgemeine Gesundheitswahrnehmung und psychisches Wohlbefinden liegt der Skalenfit nur leicht darunter (97,5%). Die interne Konsistenz, berechnet über Cronbach's Alpha, ist für alle Skalen über $\alpha = 0,80$, für drei Skalen über $\alpha = 0,90$. Im Vergleich zu den Reliabilitäten der Skalen für die Version 1.0 aus dem Bundesgesundheitsurvey zeigt sich, dass die Subskalen von Version 2.0 deutlich höher liegen (vgl. rechte äussere Spalte in Tabelle 3).

Korrelative Validitätshinweise müssen zukünftigen Analysen vorbehalten bleiben und können aufgrund der noch laufenden Auswertung der Untersuchung noch nicht vorgelegt werden. Eine detaillierte (hier nicht dargestellte) Analyse der Kennwerte und psychometrischen Charakteristika der Skalen nach Geschlecht zeigt etwas niedrigere Reliabilitäten für Männer (aber immer noch über 0,75) bei einer insgesamt niedrigen Rate fehlender Daten und hohen Skalenfits für beide Geschlechter.

Bezüglich der Skalen des SF-36, Version 2.0, zeigten sich Werte zwischen 64,4 (allgemeine Gesundheitswahrnehmung) bis 87,5 (soziale Funktionsfähigkeit) mit Standardabweichungen von 15–20 auf einer Skala von 0–100, die den Werten aus

Morfeld M, Bullinger M, Nantke J, et al.
Die Version 2.0 des SF-36 Health Survey –
Ergebnisse einer bevölkerungsrepräsentativen Studie

dem Bundesgesundheitsurvey mit Version 1.0 vergleichbar sind (vgl. Tab. 4).

Die Skalenwerte beider Versionen und Erhebungen zeigen eine hohe Bewertung der Lebensqualität besonders in den Skalen Körperliche (KöFu) und Soziale Funktionsfähigkeit (SoFu) und eine relative Einschränkung in den Skalen Allgemeine Gesundheitswahrnehmung und Vitalität, wobei Diffe-

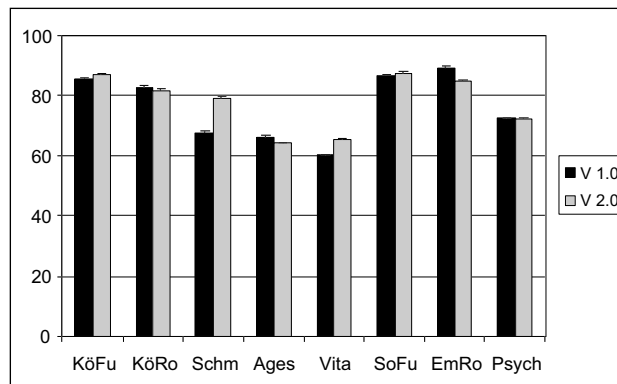


Abbildung 1 D Die acht Subskalen des SF-36 Version 2.0 im Vergleich mit denen der Version 1.0 (angegeben sind die Mittelwerte sowie der Standardmessfehler des Mittelwerts)

Tabelle 4 Skalenwerte der Subskalen des SF-36 V2.0

Subskalen	MW	SD
Körperliche Funktionsfähigkeit (KöFu) ²	87,2 (85,8)	20,36 (20,4)
Körperliche Rollenfunktion (KöRo)	81,8 (82,8)	23,09 (32,4)
Körperliche Schmerzen (Schm)	79,3 (67,7)	25,29 (25,9)
Allg. Gesundheitswahrnehmung (AGes)	64,4 (66,5)	15,17 (18,2)
Vitalität (Vita)	65,3 (60,3)	18,31 (17,8)
Soziale Funktionsfähigkeit (SoFu)	87,5 (86,8)	19,25 (19,6)
Emotionale Rollenfunktion (EmRo)	84,7 (89,4)	22,70 (26,5)
Psychisches Wohlbefinden (Psych)	72,3 (72,6)	17,15 (16,6)

In Klammern die entsprechenden Werte der Version 1.0 aus dem Bundesgesundheitsurvey

Tabelle 5 Überprüfung der SF-36 Skalen aus beiden Surveys

	t	p	Cohen's d
Körperliche Funktionsfähigkeit	-2,712	≤.001	-0,07
Körperliche Rollenfunktion		n. s.	
Körperliche Schmerzen	-17,82	≤.001	-0,45
Allg. Gesundheitswahrnehmung	5,25	≤.001	0,12
Vitalität	-10,95	≤.001	-0,28
Soziale Funktionsfähigkeit		n. s.	
Emotionale Rollenfunktion	7,84	≤.001	0,18
Psychisches Wohlbefinden		n. s.	

² In den Klammern befinden sich die im deutschen Sprachraum eingeführten Kurzbenennungen der Skalen, international sind andere Kurzformen gebräuchlich, in Reihenfolge wie in Tabelle 3 lauten diese: pfi, rolph, pain, ghp, vital, social, rolem, mhi.

renzen bis zu über 5 Punkte auf der Skala von 0–100 bereits klinisch als bedeutsam gelten können.

Auf eine altersgruppenspezifische Darstellung wurde hier verzichtet. Aus Gründen der Übersichtlichkeit wird dies Gegenstand in einer Neuauflage des Handbuches zum SF-36 sein. Der Einfluss des Alters auf die Skalen des SF-36 wird weiter unten für beide zugrundegelegten Untersuchungen in schrittweisen Regressionsanalysen untersucht.

Die teststatistische Überprüfung zeigt im t-Test für unverbundene Stichproben Unterschiede in den meisten Skalen (siehe Tab. 5).

Wie in Tabelle 5 zu sehen, ergeben sich für fünf der acht Skalen im t-Test statistisch signifikante Unterschiede. Zur Einschätzung der Bedeutsamkeit dieser Unterschiede – vor allem hinsichtlich der substantiellen Stichprobengrößen – finden sich in der rechten Spalte die nach Cohen berechneten Effektstärken (fallzahlabhängig). Nach der mittlerweile wissenschaftlich akzeptierten Klassifikation der Effektstärken zeigt sich allenfalls für die Skala Körperliche Schmerzen ein Unterschied, alle anderen Effekte sind als klein zu bezeichnen.

In Bezug auf die Unterschiede zwischen den Geschlechtern

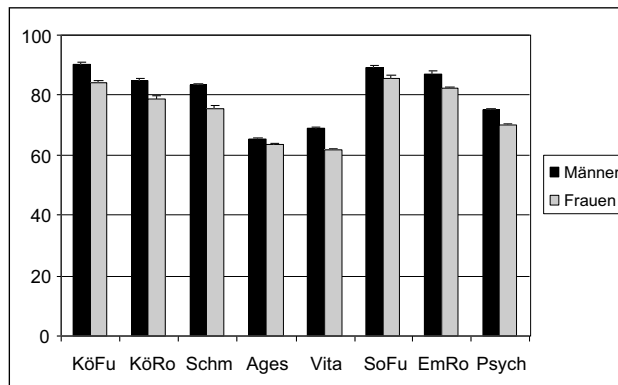


Abbildung 2 Graphische Gegenüberstellung der Subskalenwerte im SF-36 V2.0 getrennt nach Männern (n = 959) und Frauen (n = 1.081) (angegeben sind die Mittelwerte sowie der Standardmessfehler des Mittelwerts)

zeigen sich hoch signifikante Unterschiede für alle Subskalen des SF-36 dahingehend, dass Frauen eine geringere Lebensqualität als Männer berichten (vgl. Abb. 2). Dies wurde bereits für den SF-36 Version 1.0 berichtet (Campbell & Fiske 1959). Auch finden sich deutliche Unterschiede bezüglich des Alters.

Tabelle 6 Regressionsanalytische Prüfung von Einflussfaktoren auf die Subskalen des SF-36 V 1.0 im Bundesgesundheitsurvey

	Geschlecht	Alter	Ost – West	Haushalts- Nettoeinkommen	Korr. R- Quadrat
	std. Beta	std. Beta	std. Beta	std. Beta	
Körperliche Funktionsfähigkeit	-0,10	-0,40	-0,03	0,14	0,20
Körperliche Rollenfunktion	-0,07	-0,22	-0,06	0,09	0,07
Körperliche Schmerzen	-0,12	-0,16	-0,06	0,06	0,05
Allg. Gesundheitswahrnehmung		-0,29		0,13	0,10
Vitalität	-0,12		-0,06	0,08	0,02
Soziale Funktionsfähigkeit	-0,09	-0,04	-0,09	0,10	0,03
Emotionale Rollenfunktion	-0,07	-0,07	-0,07	0,10	0,02
Psychisches Wohlbefinden	-0,14		-0,07	0,11	0,04

Tabelle 7 Regressionsanalytische Prüfung von Einflussfaktoren auf die Subskalen des SF-36 V 2.0

	Geschlecht	Alter	Ost – West	Haushalts- Nettoeinkommen	Korr. R- Quadrat
	std. Beta	std. Beta	std. Beta	std. Beta	
Körperliche Funktionsfähigkeit	-0,11	-0,47		0,14	0,28
Körperliche Rollenfunktion	-0,09	-0,34	0,05	0,13	0,16
Körperliche Schmerzen	-0,18	-0,37	0,06	0,08	0,18
Allg. Gesundheitswahrnehmung	-0,04	-0,27	0,06	0,08	0,09
Vitalität	-0,17	-0,27		0,11	0,13
Soziale Funktionsfähigkeit	-0,08	-0,16		0,10	0,05
Emotionale Rollenfunktion	-0,08	-0,21		0,11	0,07
Psychisches Wohlbefinden	-0,13	-0,10	0,07	0,13	0,05

Unterschiede zwischen Ost und West finden sich im SF-36 V 2.0 tendenziell dahingehend, dass die Werte in den neuen Bundesländern etwas niedriger sind.

In Bezug auf das Haushaltseinkommen als ein Indikator für soziale Ungleichheit zeigt sich ebenfalls kein systematischer Unterschied in den Skalenwerten. Darüber hinaus zeigt sich der soziale Gradient dahingehend, dass die Personen aus einkommensstärkeren Haushalten eine deutlich höhere Lebensqualität berichten als Personen aus einkommensniedrigeren Haushalten. Dies wird aus Gründen der Übersichtlichkeit hier nicht dargestellt.

In einer für beide Versionen/Studien getrennt durchgeführten schrittweisen Regressionsanalyse wurde der Einfluss von Geschlecht, Alter sowie des Wohnortes „Neue vs. Alte Bundesländer“ und Haushalts-Nettoeinkommen auf die Skalenwerte (vgl. Tab. 6 und 7) geprüft.

In Tabelle 6 zeigt sich für die Varianzaufklärung in Version 1.0 der vier eingegangenen unabhängigen Variablen Geschlecht, Alter, Ost/West und Haushalts-Nettoeinkommen nur ein relativ geringer Beitrag, der nur für die Subskala Körperliche Funktionsfähigkeit etwas höher ist. (In den Tabellen sind jeweils die standardisierten Beta-Gewichte dargestellt, die den relativen Beitrag zur Aufklärung skalenweise untereinander darstellen). Während das Alter einen (negativen) Einfluss auf die eher körperlich ausgerichteten Subskalen des SF-36 (grau hinterlegt) hat, hat das (weibliche) Geschlecht einen Einfluss auf die eher psychisch ausgerichteten Skalen. Der Einfluss des Wohnortes – ob neue oder alte Bundesländer – ist eher gering. Das Haushalts-Nettoeinkommen wiederum hat eher einen Einfluss auf die körperlich orientierten Skalen, der jedoch auch eher gering ist.

In der folgenden Darstellung wird das Modell auf die Untersuchung angewendet, die den SF-36 in der Version 2.0 eingesetzt hat. Bezüglich der Variable Ost/West wurde eine Gewichtung dahingehend vorgenommen, dass das Verhältnis Populationsgröße Ost zu West (34 : 66 %) in der neuen Stichprobe berücksichtigt wurde.

In Tabelle 7 zeigt sich im Vergleich zur Version 1.0 der Einfluss des Alters auf die gesundheitsbezogene Lebensqualität. Ebenso ergibt sich hier eine Differenzierung zwischen eher körperlich (grau hinterlegt) und eher psychisch orientierten Skalen. Diese fällt jedoch nicht so deutlich aus wie bei der Version 1.0. Mit dem etwas höheren Altersdurchschnitt ist dies nicht zu erklären. Erwähnenswert bleibt die insgesamt etwas höhere Varianzaufklärung im Modell mit dem SF-36 Version 2.0.

Insgesamt ist es nicht möglich, aus dieser Analyse zu schließen, dass Unterschiede in den Skalenwerten auf die unterschiedlichen Versionen des SF-36 zurückzuführen sind.

Diskussion

Die Gegenüberstellung der Ergebnisse einer bevölkerungsrepräsentativen Studie in Deutschland mit dem SF-36, V 2.0, im Vergleich zur Durchführung des Bundesgesundheits surveys, in dem erstmals der SF-36 V 1.0 bei $n = 7\,124$ Personen eingesetzt wurde, erbringt psychometrisch durchaus vergleichbare Ergebnisse. Dies bezieht sich sowohl auf die psychometrischen Eigenschaften als auch auf die soziodemographischen Determinanten der Skalenwerte. Die typischen Unterschiede in Bezug auf Geschlecht (Frauen berichten niedrigere Lebensqualität), Alter (hohe Altersklassen berichten in Bezug auf die Funktionsfähigkeit geringere Werte) und soziale Ungleichheit (Personen mit niedrigem Haushalts-Nettoeinkommen als Indikator für soziale Ungleichheit berichten eine schlechtere Lebensqualität als Personen mit hohem Haushalts-Nettoeinkommen) sind in beiden Versionen und Erhebungen vergleichbar (vgl. auch (Taft et al. 2004), die für die schwedische Version 2.0 des SF-36 vergleichbares berichten). Ost-/West-Unterschiede sind demgegenüber nicht so stark ausgeprägt, möglicherweise ein Effekt der vollzogenen sozialen Angleichung in den alten und den neuen Bundesländern.

Das Ziel, mit der Version 2.0, SF-36, eine höhere Präzision und leichtere Ausführbarkeit des SF-36 gegenüber der Version 1.0 zu erreichen, ist nach den vorliegenden Analysen als gelungen anzusehen. Ein Hinweis stellt diesbezüglich die erhebliche Anpassung der Antwortformate gegenüber der Version 1.0 dar. Beim Vergleich der Boden- und Deckeneffekte zeigt sich gerade für die Skalen, bei denen die Antwortkategorien von nur zwei auf fünf angehoben wurden, eine differenziertere Ausschöpfung der Antwortmöglichkeiten, dargestellt durch geringere Deckeneffekte. Unterstützt wird dieses Ergebnis durch vergleichbare Resultate bei der Überprüfung der britischen Version 2.0 des SF-36 (Jenkinson et al. 1999).

Allerdings gestaltet sich ein direkter Vergleich SF-36, Version 1.0 vs. Version 2.0 als schwierig, weil beide Skalen in unterschiedlichen Populationen und zu unterschiedlichen Zeitpunkten vorgegeben und beantwortet wurden. Der Grund in der vorliegenden Befragung Version 1.0 und 2.0 nicht simultan vorzulegen, liegt einerseits in der Belastung der Probanden, die spezifische Fragen doppelt hätten beantworten müssen, zum anderen in der Natur einer repräsentativen Befragung, in der spezifische methodische Fragestellungen schwer einzubinden sind.

Die psychometrische Testung des SF-36 Version 2.0 zeigt, dass die Eigenschaften bezogen auf die Validität und Reliabilität gut sind. Auf eine Prüfung der Skalenstruktur mittels Faktorenanalyse wurde verzichtet.

Auch das von Ware (2004) vorgeschlagene so genannte „norm based scoring“ wurde in den folgenden Analysen nicht durch-

geführt. Bereits für die Version 1.0 liegen bisher ausschliesslich die amerikanischen Gewichte aus der Normstichprobe vor, primär aus Gründen der internationalen Vergleichbarkeit. Hier würde eine Adaptation an deutsche Gewichte zum jetzigen Zeitpunkt aus Gründen der Vergleichbarkeit bezüglich der vielen bisher bereits durchgeführten Analysen auch nicht mehr ratsam erscheinen. Für die nun hier vorgestellte Version 2.0 wird dies jedoch dann ein drängendes Thema sein, wenn absehbar wird, dass es in Deutschland zu einem Versionenwechsel von 1.0 zu 2.0 kommt und damit ein neues Instrument eingeführt wird. Das Problem der internationalen Vergleichbarkeit besteht aber auch in diesem Falle weiter. Bezüglich des Einsatzes des SF-36, Version 2.0, ist also davon auszugehen, dass sie der Version 1.0 weitgehend psychometrisch und inhaltlich äquivalent ist. Unter Nutzung eines Algorithmus ist es möglich, von Version 2.0 auf Version 1.0 zurückzurechnen³. Die Frage, welche Version einzusetzen ist, ist nicht nur empirisch, sondern auch pragmatisch zu diskutieren. Während die Nutzungsrechte von Version 1.0 in Deutschland verfügbar sind, ist die Nutzung von Version 2.0

an die amerikanische Lizenzstelle gebunden. Insofern können Referenzwerte für Version 2.0 zwar zur Verfügung gestellt werden, sie sind aber nur dann nutzbar, wenn die deutsche Version über die amerikanischen Lizenzgeber bezogen wird. Für die aktuelle Diskussion um den Einsatz des SF-36 legen die vorliegenden Ergebnisse nahe, dass Version 2.0 zwar genutzt werden kann, Version 1.0 aber ebenso einsetzbar ist. Weitere Untersuchungen werden zeigen, inwiefern weitere Kurzversionen wie z. B. SF-8, der allerdings nicht aus dem SF-36 bzw. -12 abzuleiten ist, eine weitere praktikable Möglichkeit zur ökonomischen Erfassung der gesundheitsbezogenen Lebensqualität darstellen. Darüber hinaus ist anzumerken, dass die Entwicklung von Outcome-Messinstrumenten nicht in Richtung Abgrenzung verschiedener Messinstrumente, sondern Identifikation optimaler Items zur Erfassung eines Konstruktbereichs ist und in diesem Sinne werden weitere Untersuchungen und empirische Ergebnisse zeigen, welche Items aus SF-36, Version 1.0 oder 2.0 bzw. beide gleich (oder ganz andere) geeignet sind, um subjektive Gesundheit reliabel, valide und veränderungssensitiv zu erfassen.

Zusammenfassung

Fragestellung: Der SF-36 ist ein international anerkanntes Verfahren zur Erfassung der subjektiven Gesundheit. In seiner zweiten Version beinhaltet der SF-36 gegenüber Version 1.0 Veränderungen in einigen Item-Formulierungen und Antwortkategorien. Version 2.0 liegt in deutscher Übersetzung vor, eine psychometrische Prüfung und Normierung fehlt aber.

Methoden: Der SF-36, Version 2.0, wurde in einer repräsentativen bundesdeutschen Stichprobe eingesetzt, hinsichtlich seiner Skalenstruktur, Reliabilität, Verteilungscharakteristika sowie bezüglich Unterschieden in Populationsmerkmalen untersucht und verglichen mit Version 1.0 (Bundesgesundheitsurvey).

Ergebnisse: Version 2.0 zeigt eine hohe Reliabilität der Subskalen und eine Skalenstruktur, die der Version 1.0 vergleichbar

ist. Die Skalenwerte sind in der aktuellen bevölkerungsrepräsentativen Stichprobe eher schief verteilt, wobei regressionsanalytisch Effekte von Geschlecht, Alter und Wohnort (Ost/West) nachweisbar sind. Eine Gegenüberstellung der Kennwerte von Version 1.0 und 2.0, unter Konstanthaltung obiger Confounder, zeigt eine hohe Konkordanz.

Schlussfolgerung: Die psychometrische Vergleichbarkeit der beiden Versionen ist gegeben, sollte inhaltlich wegen der unterschiedlichen Erhebungszeiträume aber vorsichtig interpretiert werden. Der erwartete Zugewinn an psychometrischer Qualität und Differenzierungsfähigkeit von Version 2.0 gegenüber Version 1.0 ist zwar vorhanden, rechtfertigt aber nicht eine Bevorzugung von Version 2.0.

Schlüsselworte: SF-36 – Version 2.0 – repräsentative deutsche Stichprobe – Reliabilität – Validität.

³ Qualimetric bietet die (kommerzielle) Möglichkeit, einen Algorithmus zu erwerben, der eine Umrechnung beider Versionen des SF-36 ermöglicht (<http://www.qualitymetric.com/products/ScoringSoftwareBasic.aspx>)

Résumé

La version 2.0 de l'Enquête sur la santé SF-36 : Résultats d'une étude représentative de la population allemande

Objectifs: Comparer les versions 1.0 et 2.0 du SF-36 lors de leur utilisation pour une population allemande. La SF-36 est une méthode internationalement reconnue pour l'identification de la santé subjective. La seconde version a été modifiée au niveau de la formulation de certaines questions et de la catégorisation de certaines réponses.

Méthodes: La SF-36, version 2.0, a été utilisée dans le cadre d'une étude représentative (n=2043) en Allemagne. Sa structure d'échelle, sa fiabilité, ses caractéristiques de distribution, et les traits de la population ont été analysés et comparés à la version 1.0 utilisée dans le cadre du German Health Survey.

Morfeld M, Bullinger M, Nantke J, et al.
Die Version 2.0 des SF-36 Health Survey –
Ergebnisse einer bevölkerungsrepräsentativen Studie

Résultats: La version 2.0 montre une bonne fiabilité des sous-échelles et une structure d'échelle comparable à la version 1.0. Dans cet échantillon représentatif, les valeurs d'échelle sont distribuées asymétriquement et les effets du sexe, de l'âge et du lieu de résidence en Allemagne (ouest/est) sont détectables.

Conclusions: Les deux versions sont comparables sur le plan psychométrique. Leur contenu doit cependant être interprété prudemment puisque les deux enquêtes ont été menées à des moments différents. Il y a des améliorations au niveau psychométrique et au niveau de la différenciation entre les versions 1.0 et 2.0. Elles ne justifient cependant pas de préférer la version 2.0.

Mots-clé: SF-36 – version 2.0 – expérience représentative allemande – fiabilité – validité.

Literaturverzeichnis

Albani C, Blaser G, Geyer M, et al. (2005). Überprüfung der Gütekriterien des Fragebogens "Profile of Mood States" (POMS) in einer repräsentativen Bevölkerungsstichprobe. *Z Klin Psychol Psychiatr Psychother*, eingereicht.

Bullinger M (1995). German translation and psychometric testing of the SF-36 Health Survey: preliminary results from the IQOLA project. *Soc Sci Med* 41: 1359–66.

Bullinger M, Kirchberger I (1998). SF-36 Fragebogen zum Gesundheitszustand: Manual. Göttingen: Hogrefe.

Bullinger M, Morfeld M (2004). Der Health Survey SF-36/SF-12: Darstellung und aktuelle Entwicklungen. In: Maurischat C, Morfeld M, Bullinger M, eds. *Lebensqualität: Nützlichkeit und Psychometrie des Short Form 12/36 in der medizinischen Rehabilitation*. Pabst: Lengerich: 15–27.

Bullinger M, Morfeld M, Kohlmann T, et al. (2003). Der SF-36 in der rehabilitationswissenschaftlichen Forschung: Ergebnisse aus dem Norddeutschen Verbund für Rehabilitationsforschung (NVRF). *Rehabilitation* 42: 218–25.

Campbell DT, Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 56: 81–105.

Hays RD, Hayashi T, Carson S, et al. (1998). User's guide for the multitrait analysis program (MAP). Santa Monica: Rand Corporation.

IQOLA, International Quality of Life Assessment Group (1998). *J Clin Epidemiol* 51 (11).

Jenkinson C, Stewart-Brown S, Petersen S, et al. (1999). Assessment of the SF36 version 2 in the UK. *J Epidemiol Community Health* 53: 46–50.

Kirchberger I (2000). Der SF-36-Fragebogen zum Gesundheitszustand: Anwendung, Auswertung und Interpretation. In: Ravens-Sieberer U, Cieza A, eds. *Lebensqualität und Gesundheitsökonomie in der Medizin*. Landsberg/Lech: ecomed: 73–85.

Radoschewski M, Bellach B-M (1999). Der SF-36 im Bundesgesundheitsurvey: Möglichkeiten und Anforderungen der Nutzung auf der Bevölkerungsebene. *Gesundheitswesen* 61 (Sonderh. 2): S191–S199.

Taft C, Karlsson J, Sullivan M (2004). Performance of the Swedish SF-36 Version 2.0. *Qual Life Res* 13: 251–6.

Thefeld W, Stolzenberg H, Bellach B-M (1999). Bundes-Gesundheitssurvey: Response, Zusammensetzung der Teilnehmer und Non-Responder-Analyse. *Gesundheitswesen* (Sonderh. 2): 57–61.

Ware JE (2004). SF-36 Health Survey update. <http://www.sf-36.org/tools/sf36.shtml>.

Ware JE, Kosinski M, Dewey J (2000). How to score version two of the SF-36 Health Survey. Lincoln, RI: QualiMetric.

Ware JE, Kosinski M, Dewey J (2003). Version 2 of the SF-36 Health Survey. 2003. Lincoln, RI: QualiMetric.

Zwingmann C, Metzger D, Jäckel W (1998). Short form-36 Health Survey (SF-36): psychometrische Analyse der deutschen Version bei Rehabilitanden mit chronischen Rückenschmerzen. *Diagnostica* 44: 209–19.

Korrespondenzadresse

Dr. Matthias Morfeld
Universitätsklinikum Hamburg-Eppendorf
Zentrum für Psychosoziale Medizin
Institut und Poliklinik für Medizinische
Psychologie, S 35
Martinistrasse 52
D-20246 Hamburg
Tel.: +49 40 42803-3871
Fax: +49 40 42803-4940
e-mail: morfeld@uke.uni-hamburg.de