

Methodische Betrachtungen zu den Summenscores des SF-36 anhand der erwachsenen bundesdeutschen Bevölkerung

Der SF-36 ist eines der international gebräuchlichsten generischen Instrumente zur Messung der gesundheitsbezogenen Lebensqualität. International wurde und wird der SF-36 sowohl in zahlreichen großen Bevölkerungsstudien [1, 2, 3, 4, 5, 6, 7] als auch in klinischen oder Interventionsstudien [8, 9, 10] eingesetzt. Mit seinen 36 Fragen ist der SF-36 ein multidimensionales Messinstrument für den selbst eingeschätzten allgemeinen Gesundheitszustand bzw. die gesundheitsbezogene Lebensqualität. Er erzeugt Skalen, die 8 Dimensionen von Gesundheit abbilden [11]:

- Körperliche Funktionsfähigkeit (PF),
- Körperliche Rollenfunktion (RP),
- Körperliche Schmerzen (BP),
- Allgemeiner Gesundheitszustand (GH),
- Vitalität (VT),
- Soziale Funktionsfähigkeit (SF),
- Emotionale Rollenfunktion (RE),
- Psychisches Wohlbefinden (MH).

Ausgehend von einer Hauptkomponentenanalyse der amerikanischen Normstichprobe konnten 2 Summenscores identifiziert werden, von denen der eine als Dimension der körperlichen Gesundheit (physical health component summary score, PCS) und der andere als Dimension der psychischen Gesundheit (mental health component summary score, MCS) interpretiert werden konnte [12]. Der hohe Anteil erklärter Varianz und die hohen Korrelationen zwischen dem Summenscore PCS mit den Komponenten für körperliche Gesundheit des SF-36 (PF, RP, BP, GH) einerseits und dem Sum-

menscore MCS mit den Komponenten des SF-36, die die psychische Gesundheit abbilden (VT, SF, RE, MH), andererseits belegten die Konstruktvalidität des Instruments [13]. Analoge Hauptkomponentenanalysen in 9 europäischen Staaten unter Nutzung ihrer Normstichproben zeigten die Stabilität dieses Konstrukts: In allen 9 Staaten (und in den USA) ergaben sich im Ergebnis dieselben Hauptkomponenten; minimale Unterschiede zeigten sich in den Gewichten (Faktorladungen), mit denen die 8 Skalen des SF-36 in die beiden Summenscores eingehen [13, 14, 15, 16, 17]. Da diese Abweichungen aber vergleichsweise geringfügig sind, wurde im Interesse einer internationalen Vergleichbarkeit die Empfehlung abgegeben, für die Konstruktion der beiden Summenscores jeweils die amerikanischen Gewichte zu verwenden. Dies hat neben der Vergleichbarkeit den Vorteil, dass auch für Populationen, für die keine Normstichprobe zur Gewinnung eigener Faktorladungen erhoben wurde, die Scores für physische und psychische Gesundheit unter Nutzung der amerikanischen Gewichte dennoch berechnet werden können. Inwieweit die amerikanischen Gewichte bei internationalen Vergleichen Sinn machen, wurde beispielsweise für Großbritannien von Jenkinson [14] untersucht. Hier wurden keine nennenswerten Unterschiede zwischen den amerikanischen und britischen Gewichten gefunden.

In den erwähnten Ergebnisvergleich für 9 europäische Staaten wurde auch eine 1994 für Deutschland erhobene Normstichprobe einbezogen [18]. Die in [4] be-

schriebene aktuelle Normstichprobe für den SF-36 wurde bislang noch nicht einer analogen vergleichenden Hauptkomponentenanalyse unterzogen.

Auch wenn es mittlerweile generelle methodische Vorbehalte gegen die Verwendung der Summenscores für physische und psychische Gesundheit gibt bzw. andere Verfahren zur Berechnung der Faktoren etwas andere Ergebnisse liefern [19, 20], finden die Summenscores nach wie vor Anwendung. Daher ist es Ziel der vorliegenden Arbeit zu untersuchen, wie sich die Faktorstruktur des SF-36 im deutschen Datensatz des Bundes-Gesundheitssurveys 1998 von der amerikanischen Normstichprobe unterscheidet. Darüber hinaus werden die analogen Hauptkomponententransformationen für Männer und Frauen durchgeführt, um der Frage nachzugehen, ob geschlechtsspezifische Auswertungen im Sinne eines Gender Mainstreaming zusätzliche Informationen liefern. Für die resultierenden PCS und MCS werden alters- und geschlechtsspezifische Normwerte für die bundesdeutsche Bevölkerung angegeben.

Material und Methode

Stichprobe

Grundlage für die neuen Berechnungen der deutschen Hauptkomponenten und deren Faktorgewichte ist der Bundes-Gesundheitssurvey 1998. Darin wurden 7.124 Personen einer repräsentativen Stichprobe der 18- bis 79-jährigen Wohnbevölkerung in Deutschland untersucht und befragt [21]. Der Anteil der Frauen war entsprechend

der Bevölkerungsstruktur mit 51,3% etwas höher als der der Männer mit 48,7%. Das Durchschnittsalter der Befragten lag bei 46,1 Jahren. Der Sozialstatus wurde im Bundes-Gesundheitssurvey gemäß den Empfehlungen der Deutschen Arbeitsgemeinschaft für Epidemiologie (DAE) erfasst [22]. Der Oberschicht gehörten 21,6%, der Mittelschicht 55,4% und der Unterschicht 22,2% der Studienteilnehmer an.

Im Selbstausfüllfragebogen wurden die Probanden auch nach dem Auftreten von Schmerzen in verschiedenen Körperregionen im Lauf des vergangenen Jahres und in den letzten 7 Tagen befragt. Zusätzlich wurde die Lokalisation des stärksten

Schmerzes in den letzten 7 Tagen erfragt, und der Proband erhielt die Möglichkeit, die empfundene Schmerzintensität auf einer Skala von 1–9 (beginnend mit 1 für kaum spürbare Schmerzen und endend mit 9 für unerträgliche Schmerzen) zu quantifizieren [23].

Zur Messung der gesundheitsbezogenen Lebensqualität wurde die deutschsprachige Übersetzung des SF-36 Version 1 in der Standardversion (4 Wochen) eingesetzt [18, 24]. Von 6.964 Probanden wurde der Fragebogen des Bundes-Gesundheitssurveys so vollständig ausgefüllt, dass diese Angaben zu SF-36 vorlagen. Um die Skalenwerte PF, RP, BP, GH, VT, SF, RE und MH zu berech-

nen, wurde mit den Daten, wie in der Literatur beschrieben, verfahren (s. [4]).

Bildung der Summenskalen

Für die richtige Interpretation der körperlichen und psychischen Summenskala ist deren standardisierte Berechnung von grundlegender Bedeutung. Nachdem die Skalen des SF-36 einer Hauptkomponentenanalyse mit anschließender orthogonaler Rotation unterzogen wurden [25], um Koeffizienten zu ermitteln, werden die Summenskalen des SF-36 in 3 Schritten gebildet:

- In einem ersten Schritt wird jede SF-36-Skala „z-transformiert“, indem von den einzelnen Skalenwerten der aus der jeweiligen Normstichprobe gewonnene Mittelwert der Skala subtrahiert und das Ergebnis durch die Standardabweichung geteilt wird.
- Nach der z-Transformation werden für PCS und MCS jeweils Rohwerte durch Multiplikation der 8 einzelnen Skalen mit den jeweiligen Koeffizienten für den körperlichen bzw. psychischen Faktor und anschließende Aufsummierung berechnet.
- Schließlich werden diese beiden Rohwerte jeweils so transformiert, dass in der jeweiligen Normstichprobe der Mittelwert 50 und die Standardabweichung 10 beträgt. Im Ergebnis liegen die körperliche und die psychische Summenskala vor. Eine einfache Interpretation der Werte ist nun möglich: Ein Wert

Tabelle 1

Korrelation der 8 SF-36-Skalen mit den orthogonalen Hauptkomponenten

Skala	PCS			MCS		
	USA	Deutschland 94	Deutschland 98	USA	Deutschland 94	Deutschland 98
PF	0,85	0,85	0,84	0,12	0,16	0,32
RP	0,81	0,79	0,76	0,27	0,28	0,19
BP	0,76	0,82	0,76	0,28	0,24	0,15
GH	0,69	0,72	0,64	0,37	0,38	0,31
VT	0,47	0,52	0,40	0,64	0,64	0,66
SF	0,42	0,33	0,35	0,67	0,75	0,70
RE	0,17	0,20	0,16	0,78	0,71	0,71
MH	0,17	0,16	0,14	0,87	0,87	0,86

PF Körperliche Funktionsfähigkeit, RP Körperliche Rollenfunktion, BP Schmerzen, GH Allgemeine Gesundheit, VT Vitalität, SF Soziale Funktionsfähigkeit, RE Emotionale Rollenfunktion, MH Psychisches Wohlbefinden, PCS physical health component summary score, MCS mental health component summary score.

Tabelle 2

Koeffizienten für PCS und MCS aus der amerikanischen und der deutschen Normstichprobe

Subskala	USA	Deutschland gesamt	Deutschland Frauen	Deutschland Männer	USA	Deutschland gesamt	Deutschland Frauen	Deutschland Männer
PF	0,424	0,451	0,464	0,436	–0,230	–0,192	–0,201	–0,179
RP	0,351	0,310	0,301	0,317	–0,123	–0,041	–0,034	–0,043
BP	0,318	0,375	0,371	0,382	–0,097	–0,092	–0,086	–0,100
GH	0,250	0,286	0,294	0,274	–0,016	–0,006	0,005	0,027
VT	0,029	0,028	0,018	0,040	0,235	0,284	0,292	0,276
SF	–0,008	–0,047	–0,041	–0,050	0,269	0,338	0,329	0,345
RE	–0,192	–0,156	–0,156	–0,148	0,434	0,360	0,355	0,354
MH	–0,221	–0,105	–0,103	–0,109	0,486	0,390	0,382	0,400

PF Körperliche Funktionsfähigkeit, RP Körperliche Rollenfunktion, BP Schmerzen, GH Allgemeine Gesundheit, VT Vitalität, SF Soziale Funktionsfähigkeit, RE Emotionale Rollenfunktion, MH Psychisches Wohlbefinden.

Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 2004 · 47:1027–1032
DOI 10.1007/s00103-004-0933-1
© Springer Medizin Verlag 2004

U. Ellert · B.-M. Kurth

Methodische Betrachtungen zu den Summenscores des SF-36 anhand der erwachsenen bundesdeutschen Bevölkerung

Zusammenfassung

Der SF-36 ist ein weit verbreitetes Instrument zur Messung der gesundheitsbezogenen Lebensqualität, das mit seinen 36 Fragen 8 Dimensionen [Körperliche Funktionsfähigkeit (PF), Körperliche Rollenfunktion (RP), Körperliche Schmerzen (BP), Allgemeiner Gesundheitszustand (GH), Vitalität (VT), Soziale Funktionsfähigkeit (SF), Emotionale Rollenfunktion (RE), Psychisches Wohlbefinden (MH)] von selbst berichteter Gesundheit abdeckt. Um die Anzahl der Dimensionen zu reduzieren, wurde mit den Daten des US-amerikanischen SF-36-Surveys eine Hauptkomponentenanalyse durchgeführt. Die 2 orthogonalen Hauptkomponenten, die durch eine gewichtete Aufsummierung der Originalskalen des SF-36 definiert sind, lassen sich als körperliche (PCS) bzw. psychische (MCS) Komponenten interpretieren. Die Durchführung analoger Hauptkomponentenanalysen für 9 europäische Länder-Surveys aus den Jahren 1994–1996 ergaben jeweils dieselben Hauptkomponen-

ten mit landesspezifischen Gewichten bei der Aufsummierung der SF-36-Skalen. Da sich diese Gewichte jedoch nur geringfügig von den US-amerikanischen unterscheiden, wurde im Interesse einer internationalen Vergleichbarkeit und einer Vereinfachung der Berechnungsprozeduren empfohlen, generell die amerikanischen Gewichte zur Kalkulation der beiden Summenscores PCS und MCS zu verwenden. In dieser Arbeit werden die Berechnungen für eine neue Bevölkerungsstichprobe aus dem Jahr 1998 nachvollzogen und neue deutsche Gewichte berechnet. Durch einen Vergleich mit den amerikanischen Gewichten aus dem Jahr 1994 und mit Ergebnissen für die Bevölkerungsstichprobe aus dem Jahr 1998 werden Pro und Kontra der Verwendung amerikanischer Gewichte erneut diskutiert. Im Ergebnis erscheinen die amerikanischen Gewichte für internationale Vergleiche weiterhin akzeptabel. Die getrennte Betrachtung von körperlicher und psychischer Gesundheit für Män-

ner und Frauen vermittelt unter Umständen mehr Einblick in geschlechtsspezifische Veränderungen. Für die deutsche Normpopulation von 1998 werden in dieser Arbeit die beiden Summenscores geschlechtsspezifisch berechnet und die Gewichte zur Anwendung in weiteren Untersuchungen publiziert. Im Sinne des Gender Mainstreamings sollte auch bei der Lebensqualitätsforschung dieser Zugang Berücksichtigung finden. In Anbetracht der vielfältigen methodischen und inhaltlichen Schwierigkeiten bei der Verwendung der Summenskalen für körperliche und psychische Gesundheit erhebt sich die Frage nach der Sinnhaftigkeit der Verwendung der Summenskalen anstelle der Originalskalen des SF-36.

Schlüsselwörter

Lebensqualität · SF-36 · Summenscores · Länderspezifische Gewichte · Gender Mainstreaming · Bundes-Gesundheitssurvey 1998

Methodological views on the SF-36 summary scores based on the adult German population

Abstract

The SF-36 is a widely used instrument to measure health-related quality of life that provides a profile of eight scales [Physical Functioning (PF), Role-Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role-Emotional (RE) and Mental Health (MH)]. To reduce the dimension of the profile, a principal component analysis with the data of the US SF-36 Survey was undertaken. As a result, the Physical Component Summary (PCS) and the Mental Component Summary (MCS) were gained by weighting and summing up of the original scales of the SF-36. The weights were assumed to be country specific. Comparing the weights from the representative samples of nine European countries with those from the USA it has been suggested that the weights gained from the US sam-

ple could be applied to all datasets for purposes of comparability and simplicity. In this paper, for a new representative German population sample completed in 1998, the calculation results are compared with those for the older German population sample from 1994. PCS and MCS are calculated using the weights from the new German population sample as well as using the American weights. In this paper, the calculations for the 1998 representative population sample were repeated based on the US weights developed in 1994. The German weights were also calculated based on the 1998 sample. The results were compared and the pro and cons of using the American weights are discussed. In essence, the American weights continue to be acceptable for international comparisons. The separate evalu-

ation of physical and mental health for men and women allows additional insight into gender-specific changes. In this paper, the summary scales for the 1998 German normative population are calculated gender specific. The weights to be used in future studies are also published. With respect to gender mainstreaming, this approach should be taken into consideration in quality of life research. Regarding the methodological and factorial difficulties, the questions arise if the summary scales should be applied instead of the original scales of the SF-36 questionnaire.

Keywords

Quality of life · SF-36 · Summary health scores · Country-specific weighting · Gender mainstreaming · German National Health Survey

Tabelle 3

Werte für MCS und PCS in Altersgruppen unter Zugrundelegung der amerikanischen und deutschen Faktorladungen für die aktuelle deutsche Normpopulation 1998

	USA-Gewichte		Deutsche Gewichte	
	PCS	MCS	PCS	MCS
Gesamt	48,36 (9,42)	50,87 (8,82)	50,00	50,00
Männer	49,26 (8,87)	51,92 (8,00)	51,22 (9,83)	51,34 (9,08)
Frauen	47,49 (9,85)	49,84 (9,43)	49,00 (10,91)	48,76 (10,65)
<i>Frauen</i>				
18–19	53,51 (5,74)	47,29 (8,88)	55,50 (6,38)	46,83 (9,91)
20–29	51,69 (6,72)	49,10 (8,62)	53,75 (7,41)	48,51 (9,79)
30–39	51,03 (7,31)	48,60 (9,23)	53,00 (8,00)	47,82 (10,46)
40–49	48,65 (8,59)	49,39 (9,61)	50,28 (9,64)	48,44 (10,91)
50–59	45,27 (10,08)	50,02 (10,13)	46,49 (11,19)	48,68 (11,56)
60–69	43,78 (10,53)	52,16 (8,30)	44,83 (11,71)	50,94 (9,38)
70–79	40,80 (11,87)	51,03 (10,26)	41,50 (13,02)	48,97 (11,60)
<i>Männer</i>				
18–19	53,98 (5,44)	49,37 (8,64)	56,48 (6,04)	49,03 (9,80)
20–29	52,66 (6,17)	51,43 (7,17)	55,07 (6,74)	51,27 (7,95)
30–39	52,15 (6,34)	51,41 (7,37)	54,42 (6,97)	51,19 (8,29)
40–49	50,33 (7,51)	51,78 (8,02)	52,40 (8,29)	51,40 (9,09)
50–59	46,70 (9,55)	51,68 (8,91)	48,29 (10,72)	50,71 (10,34)
60–69	44,96 (9,99)	53,25 (8,51)	46,44 (11,05)	52,20 (9,73)
70–79	42,71 (11,68)	53,98 (7,39)	44,06 (12,90)	52,49 (8,70)

über 50 liegt über dem Mittelwert und einer unter 50 unter dem Mittelwert in der jeweiligen Normstichprobe.

Während die Ergebnisse der Berechnung von PCS und MCS für die amerikanische Normpopulation publiziert sind [25], wurden die Berechnungen für die deutsche Normpopulation von 1998 neu durchgeführt. Im Ergebnis wurden die deutschen Faktorladungen erzeugt und PCS und MCS nach obiger Vorschrift berechnet.

Parallel dazu wurden PCS und MCS unter Nutzung der aus der amerikanischen Stichprobe festgelegten Koeffizienten berechnet [25] und die Unterschiede ermittelt. Analog wurde mit den 2 nach Geschlecht getrennten Teilstichproben verfahren.

Die Faktoranalysen wurden mit dem Statistiksoftwareprogramm SAS 9.1 (proc factor), alle anderen Berechnungen mit SPSS Version 12.0 durchgeführt.

Ergebnisse

Die Korrelationen der 8 SF-36-Skalen mit den orthogonalen Hauptkomponenten unterstreichen deren Interpretation als kör-

perliche und psychische Dimension von Gesundheit. So ist beispielsweise die Skala Körperliche Funktionsfähigkeit (PF) am stärksten mit der physischen Komponente und am wenigsten mit der psychischen Komponente korreliert. Dies trifft sowohl auf die amerikanische als auch auf die deutsche Normstichprobe von 1994 zu. In der deutschen Normstichprobe von 1998 ist es hingegen die Schmerzskala, die am wenigsten mit der psychischen Komponente korreliert. ■ **Tabelle 1** vergleicht die Korrelationen der Skalen des SF-36 mit den jeweils separat anhand der amerikanischen Koeffizienten berechneten orthogonalen Hauptkomponenten für die beiden deutschen Normstichproben von 1994 und 1998 sowie für die amerikanische Normstichprobe.

Sowohl mit der deutschen Gesamtstichprobe von 1998 als auch mit je einer Unterstichprobe für Frauen bzw. Männer wurden Faktoranalysen zur Gewinnung der spezifischen deutschen Gewichte durchgeführt. Die durch die Faktoranalyse gewonnenen Koeffizienten aus der deutschen Stichprobe von 1998 unterscheiden sich nicht wesentlich von denen der amerikanischen Stichprobe, die gleichen Dimensionen laden

Tabelle 4

PCS und MCS nach Alter und Geschlecht unter Zugrundelegung von deutschen geschlechtsspezifischen Werten

	Geschlechtsspezifische Gewichte	
	PCS	MCS
<i>Frauen</i>		
18–19	56,14 (6,26)	48,03 (9,37)
20–29	54,47 (7,22)	49,60 (9,28)
30–39	53,73 (7,80)	48,95 (9,91)
40–49	51,05 (9,28)	49,57 (10,34)
50–59	47,37 (10,88)	49,82 (10,94)
60–69	45,74 (11,39)	52,00 (8,88)
70–79	42,53 (12,65)	50,14 (10,95)
<i>Männer</i>		
18–19	55,50 (6,31)	47,53 (10,88)
20–29	54,07 (7,04)	49,94 (8,78)
30–39	53,42 (7,31)	49,81 (9,15)
40–49	51,31 (8,71)	49,98 (10,06)
50–59	46,98 (11,31)	49,13 (11,48)
60–69	45,01 (11,73)	50,72 (10,78)
70–79	42,41 (13,70)	51,03 (9,64)

(wie auch in der Literatur beschrieben) auf den 2 Faktoren (■ **Tabelle 2**).

Auch wenn die Koeffizienten nicht stark voneinander abweichen, zeigt sich, dass die Abweichungen zwischen den amerikanischen und den deutschen Koeffizienten insgesamt größer als die zwischen Frauen und Männern innerhalb der deutschen Stichprobe sind. Bis auf die Skala Körperliche Rollenfunktion (RP) sind die Koeffizienten für die körperliche Summenskala in allen körperlichen Dimensionen des SF-36, die sich aus der deutschen Stichprobe ergeben, höher als die der amerikanischen Normstichprobe. Bei den Frauen ist dies sogar noch deutlicher ausgeprägt als bei den Männern. Für 2 der 4 psychischen Dimensionen [Vitalität (VT) und Soziale Funktionsfähigkeit (SF)] ergeben sich aus der deutschen Normstichprobe höhere Werte als aus der amerikanischen Normstichprobe. Für die Dimensionen Emotionale Rollenfunktion und Psychisches Wohlbefinden sind hingegen die aus der deutschen Normstichprobe gewonnenen Werte niedriger als die der amerikanischen Normstichprobe.

Um einen Eindruck über die Auswirkungen der Verwendung der amerikani-

schen Gewichte auf die deutschen Daten zu gewinnen, wurden die alters- und geschlechtsspezifischen Mittelwerte der Summenscores PCS und MCS unter Verwendung der amerikanischen und der aktuellen deutschen Gewichte berechnet. Die unterschiedlichen Mittelwerte sind in **■ Tabelle 3** zusammengefasst.

Die Summenscores für die 1994er-Normstichprobe wurden mit den amerikanischen Gewichten berechnet und sind im Handbuch angegeben [18]. Vergleicht man die mit den amerikanischen Koeffizienten berechneten Summenscores der beiden deutschen Stichproben von 1994 und 1998, so ergibt sich 1998 für beide Geschlechter ein um ca. zwei Punkte niedriger körperlicher und ein etwas niedrigerer psychischer Summenscore (s. [18]). Möglicherweise würden diese Veränderungen mit speziellen deutschen Gewichten noch deutlicher zu Tage treten.

Aus den Berechnungswerten ergibt sich, dass die deutschen Männer sowohl mit amerikanischen als auch mit deutschen Gewichten höhere Mittelwerte in beiden Skalen erreichen als die Frauen, dass sich die Relationen der Einzelskalen zueinander aber in Abhängigkeit von den verwendeten Gewichten unterscheiden: So erreichen die deutschen Frauen innerhalb der deutschen Population für die körperliche Gesundheit einen höheren Mittelwert als für die psychische Gesundheit. Diese Relation dreht sich bei amerikanischer Gewichtung um. Männer liegen in der deutschen Bewertung in beiden Skalen über dem Populationsmittel, dies ist bei der amerikanischen Gewichtung nicht mehr ersichtlich.

Die Bewertung des Altersganges bei Männern und bei Frauen mit amerikanischen Gewichten ist ebenfalls erschwert, weil hier der Bezug auf die amerikanische Bevölkerungsstruktur eine innerdeutsche Interpretation unmöglich macht.

Unabhängig von der zugrunde liegenden Scorebildung erreichen Männer in der körperlichen und auch in der psychischen Summenskala höhere Werte als Frauen. Ein deutlicher Altersgang mit abnehmenden Werten bei zunehmendem Alter ist für die körperlichen Summenskalen sowohl bei den Männern als auch bei den Frauen zu beobachten. Dies gilt gleichermaßen für die mit amerikanischen wie für

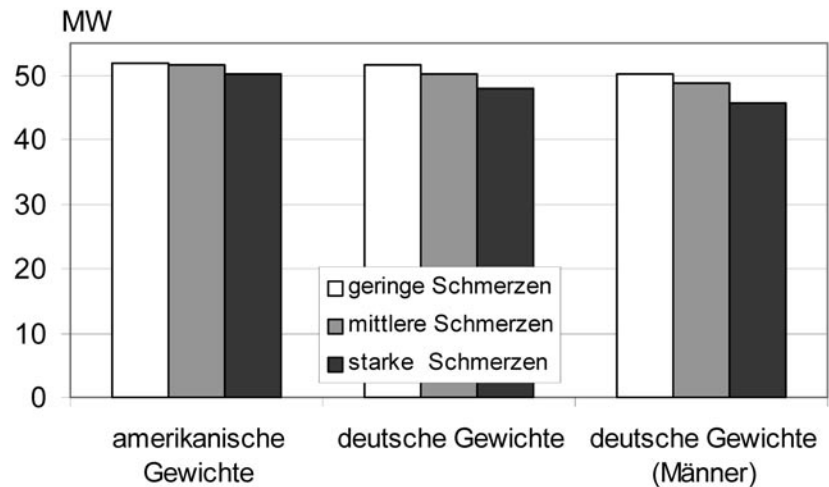


Abb. 1 ▲ Die psychische Summenskala des SF-36 (MCS) bei Männern in Abhängigkeit von der Schmerzintensität und unter Zugrundelegung von amerikanischen, deutschen oder deutschen geschlechtsspezifischen Werten

die mit deutschen Gewichten der Gesamtstichprobe gewonnenen Scores.

Die Tatsache, dass die in **■ Tabelle 3** für Männer und Frauen dargestellten Altersgänge in der körperlichen und psychischen Gesundheit jeweils auf die Gesamtbevölkerung Bezug nehmen, erschwert die geschlechtsspezifische Bewertung der Veränderungen mit zunehmendem Alter. Um die männliche und die weibliche Population in ihrer Geschlechtsspezifität getrennt zu erfassen [26], wurde daher die Hauptkomponentenanalyse auch für die nach Geschlechtern getrennten Stichproben gerechnet. Die Werte für die beiden Summenscores nach Alter und Geschlecht, die sich jeweils aus einer getrennten Hauptkomponentenanalyse in den beiden Unterstichproben ergeben, sind in **■ Tabelle 4** dargestellt.

Bei dieser Kalkulation wird auf einmal sehr transparent, dass sich die Veränderungen in der körperlichen und psychischen Gesundheit mit dem Alter bei Männern und Frauen fast identisch abbilden. Ausgehend von bekanntermaßen unterschiedlichen Grundniveaus ist der Altersgang analog.

Die Instrumente zur Messung gesundheitlich bedingter Lebensqualität sind entwickelt worden, um die Wirksamkeit von Therapien, die Auswirkungen gesundheitlicher Beeinträchtigungen, die Unterschiede in Risikopopulationen sowie den Einfluss sozialer Komponenten messen zu können. Dementsprechend sind theoretisch methodische Fragen auch nicht aus-

schließlich auf der Populationsebene zu beantworten, sondern auch im Hinblick auf die ursprüngliche Zielrichtung der Messung von Wirkungen.

Dass Schmerzen einen großen Einfluss auf die gesundheitsbezogene Lebensqualität haben, wurde auch schon in [6] gezeigt. Schmerz beeinflusst ausnahmslos alle Skalen des SF-36 negativ. Je geringer die Schmerzintensität, desto höher auch die Werte der beiden Summenskalen. Dies zeigt sich unabhängig davon, welche Gewichtungsfaktoren zugrunde gelegt werden. Was allerdings sehr deutlich wird, ist die unterschiedliche Differenzierung. Die Unterschiede sind in der körperlichen Summenskala mit allen 3 Gewichten signifikant. Erst wenn mit den deutschen Gewichten gerechnet wird, werden die Unterschiede in der psychischen Summenskala signifikant, bei Bezugnahme auf die männliche Population (also unter Verwendung der spezifischen männlichen Gewichte) werden die Auswirkungen maximal deutlich (**■ Abb. 1**). Bei den Frauen zeigt sich ein ähnliches Bild, auf eine diesbezügliche Abbildung wurde hier verzichtet.

Diskussion

Da eine Bewertung von 8 Einzelskalen oft unübersichtlich und aufwändig ist, stellt die Reduzierung auf nur 2 Summenscores eine Erleichterung dar. Um zu diesem sicher wünschenswerten Ergebnis zu kommen, muss allerdings in Kauf genommen

werden, dass für die Bildung dieser Summenscores Gewichte benötigt werden, die auf unterschiedlichen Wegen gewonnen werden können. Von den Entwicklern des Instruments wird eine Hauptkomponentenanalyse mit orthogonaler Rotation zur Gewinnung dieser Gewichte empfohlen. Dieses Verfahren wurde in verschiedenen europäischen Ländern überprüft [15] und führte dort zu ähnlichen Ergebnissen. Nutzt man also die etablierten Gewichte, so ist durch eine einheitliche Standardisierung die internationale Vergleichbarkeit und Darstellungsmöglichkeit der Ergebnisse gegeben. Welchen Einfluss die Verwendung bestimmter Gewichte auf die Hervorhebung oder Verringerung bestehender Unterschiede hat, bleibt zu klären.

Entsprechend war es ein Ziel dieser Arbeit, mit dem vorgegebenen Bildungsalgorithmus deutsche Gewichte zu bilden und diese mit den amerikanischen zu vergleichen. Die Werte aus der amerikanischen Stichprobe und auch die deutschen Werte führen zu ähnlichen Ergebnissen und keinesfalls zu widersprüchlichen Grundaussagen. Sowohl unter Nutzung der amerikanischen als auch der deutschen Gewichte erreichen Männer in beiden Summenskalen höhere Werte als Frauen. Zusammen liegen beide Summenskalen bei den Männern über der Norm für die gesamte Bevölkerung bzw. bei den Frauen unterhalb dieser Werte. Allerdings wird die Interpretation von Unterschieden und Veränderungen innerhalb eines Landes durch Nutzung der nationalen Werte erleichtert. Die Transformation auf einen Mittelwert von 50 und eine Standardabweichung von 10 innerhalb der Stichprobe macht es möglich, Abweichungen von der Norm nach oben oder unten zu erkennen und zu quantifizieren. So wird unter Verwendung der deutschen Gewichte als Vergleich der Wert 50 in der deutschen Stichprobe herangezogen. Unter Verwendung der amerikanischen Gewichte wären diese Vergleichswerte dann 48,36 für PCS und 50,87 für MCS.

Wie am Beispiel der Auswirkungen von Schmerz demonstriert, können bestimmte Bewertungen innerhalb einer Population mit den amerikanischen Gewichten gar nicht bzw. mit nicht geschlechtsdifferenzierten Werten schlechter vorgenommen werden. Bei Betrachtung der zugrunde liegenden Subskalen war aber ein signifikant negativer Einfluss von Schmerzen

auch auf die psychische Gesundheit von Männern zu erwarten, der sich in diesem Fall erst durch die Verwendung der geschlechtsspezifischen Gewichte für Männer gezeigt hat. Die getrennte Betrachtung von körperlicher und psychischer Gesundheit für Männer und Frauen vermittelt möglicherweise mehr Einblick in geschlechtsspezifische Veränderungen.

Eine weitere Schwierigkeit bei der Bildung von Summenscores sind die in der Mehrzahl der psychometrischen Studien gefundenen Mischladungen, beispielsweise zwischen der Skala PF und dem psychischen Summenscore oder der Skala VT und dem körperlichen Summenscore. Aus diesem Grund ist eine Zuordnung der Subskalen zu den Summenscores nicht ganz eindeutig möglich. Wie in [20] gezeigt, führt die Verwendung von Faktorenextraktion nach Maximum-Likelihood oder von Strukturgleichungsmodellen zur Bildung der Summenscores zu anderen – teilweise sogar besser mit den zugrunde liegenden Subskalen übereinstimmenden – Ergebnissen. Eine inhaltliche Beschränkung auf die Subskalen, zu deren Bildung keine Gewichte benötigt werden, könnte ein Ausweg aus diesem Dilemma sein.

Korrespondierender Autor

Dr. U. Ellert

Abteilung Epidemiologie und Gesundheitsberichterstattung,
Robert Koch-Institut, Nordufer 20, 13353 Berlin
E-Mail: u.ellert@rki.de

Literatur

1. Apolone G, Mosconi P (1998) The Italian SF-36 Health Survey: translation, validation and norming. *J Clin Epidemiol* 51:1025–1036
2. Bjorner JB, Thunelborg K, Kristensen TS et al. (1998) The Danish SF-36 Health Survey: translation and preliminary validity studies. *J Clin Epidemiol* 51:991–999
3. Bullinger M, Alonso J, Apolone G et al. (1998) Translating health status questionnaires and evaluating their quality: the IQOLA project approach. *J Clin Epidemiol* 51:913–923
4. Ellert U, Bellach BM (1999) Der SF-36 im Bundes-Gesundheitssurvey – Beschreibung einer aktuellen Normstichprobe (SF-36 Questionnaire in the Federal Health Survey – Description). *Gesundheitswesen* 61 [Sonderheft 2]:184–190
5. Fukuhara S, Bito S, Green J et al. (1998) Translation, adaption, and validation of the SF-36 Health Survey for use in Japan. *J Clin Epidemiol* 51:1037–1044
6. Kurth BM, Ellert U (2002) The SF-36 questionnaire and its usefulness in population studies: results of the German Health Interview and Examination Survey 1998. *Soz-Präventivmed* 47:266–277

7. Sanson-Fisher RW, Perkins JJ (1998) Adaption and validation of the SF-36 Health Survey for use in Australia. *J Clin Epidemiol* 51:961–967
8. Hlatky MA, Boothroyd D, Vittinghoff E et al. (2002) Quality-of-life and depressive symptoms in postmenopausal women after receiving hormone therapy. *JAMA* 287:591–597
9. King AC, Pruitt LA, Phillips W et al. (2000) Comparative effects of two physical activity programs on measured and perceived physical functioning and other health-related quality of life outcomes in older adults. *J Gerontol* 55A:M74–M83
10. Ruo B, Rumsfeld JS, Hlatky MA et al. (2003) Depressive symptoms and health-related quality of life: the Heart and Soul Study. *JAMA* 290:215–221
11. Ware JE Jr, Snow KK, Kosinski M, Gandek BG (1993) SF-36 Health Survey Manual and Interpretation Guide. The Health Institute, New England Medical Center, Boston, MA
12. Ware JE et al. (1995) Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the medical outcome study. *Medical Care* 33:264–279
13. Ware JE et al. (1998) The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA project. *J Clin Epidemiol* 51:1159–1165
14. Jenkinson C (1999) Comparison of UK and US methods for weighting and scoring the SF-36 summary measures. *J Public Health Med* 21:372–376
15. Ware JE et al. (1998) The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA project. *J Clin Epidemiol* 51:1167–1170
16. Scott KM et al. (2000) A challenge to the cross-cultural validity of the SF-36 health survey: factor structure in Maori, Pacific and New Zealand European ethnic groups. *Soc Sci Med* 51:1655–1664
17. Mishra G, Schofield MJ (1998) Norms for the physical and mental health component summary scores of the SF-36 for young, middle-aged and older Australian women. *Qual Life Res* 7:215–220
18. Bullinger M, Kirchberger I (1998) SF-36 Fragebogen zum Gesundheitszustand (SF-36 Health Survey concerning Health status). Hogrefe, Göttingen Bern Toronto Seattle
19. Kohlmann T (2000) Handle with care! Complex scoring systems in measuring health-related quality of life. *Soz Präventivmed* 45:235–236
20. Wilson D, Parsons J, Tucker G (2000) The SF-36 summary scales: problems and solutions. *Soz Präventivmed* 45:239–246
21. Thefeld W, Stolzenberg H, Bellach BM (1999) Bundes-Gesundheitssurvey: Response, Zusammensetzung der Teilnehmer und Non-Responder-Analyse. *Gesundheitswesen* 61:57–61
22. Ahrens W, Bellach BM, Jöckel KH (1998) Messung soziodemographischer Merkmale in der Epidemiologie. RKI-Schriften 1/1998. MMV Medizin Verlag, München
23. Bellach BM, Ellert U, Radoschewski M (2000) Epidemiologie des Schmerzes – Ergebnisse des Bundesgesundheits surveys 1998. *Bundesgesundheitsblatt* 43:424–431
24. Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36) I. Conceptual framework and item selection. *Medical Care* 30:473–483
25. Ware JE, Kosinski M, Keller S (1994) SF-36 physical and mental summary scales: a user's manual. MA: The Health Institute, New England Medical Center, Boston
26. Council of Europe (1998) Gender Mainstreaming – Konzeptioneller Rahmen, Methodologie und Beschreibung bewährter Praktiken. Schlussbericht über die Tätigkeit der Group of Specialists on Mainstreaming. EG-S-MS, Straßburg