

Erfassung der gesundheitsbezogenen Lebensqualität mit dem deutschen SF-8

Ein Vergleich der telefonischen und postalischen Befragungsmethode

In epidemiologischen Studien haben sich postalische und telefonische Befragungsmethoden bewährt. Die Vor- und Nachteile dieser Methoden sind vielfach beschrieben und für Aspekte der Response, der Ökonomie und der Praktikabilität gut belegt [1, 2, 3]. Deutlich kontroverser diskutiert wird die Frage, inwieweit die verschiedenen Erhebungsmethoden zu äquivalenten Befragungsergebnissen führen. Gegenstand der Betrachtung sind dabei die verschiedensten Aspekte der Datenqualität.

Vielfach wird eine unterschiedliche Teilnahmebereitschaft berichtet. Dies ist insbesondere vor dem Hintergrund teilnahmebedingter Stichprobenverzerrungen bedeutsam. So finden sich Berichte zu unterschiedlichem Teilnahmeverhalten beispielsweise in Abhängigkeit vom Alter [4] und Geschlecht [5] oder des sozioökonomischen Status [6].

Ein Vorteil telefonischer Befragungen liegt in der Vollständigkeit von Datensätzen; diesbezüglich zeichnen sie sich durch äußerst geringe Missingraten aus. Dementsprechend finden sich für telefonische und postalische Befragungen nur in wenigen Studien vergleichbare Missingraten [7, 8], in der Mehrzahl der Studien erweist sich die telefonische Datenerhebung der postalischen als überlegen [2, 4, 9].

Zentral bleibt darüber hinaus die Frage, welchen möglichen Verzerrungen das Antwortverhalten selbst unterliegt. Untersuchungen zeigen, dass persönliche Befragungsmethoden positivere Angaben zu

Morbidität, Inanspruchnahme medizinischer Leistungen oder sozial unerwünschtem Verhalten hervorrufen können als postalische Befragungen [9, 10, 11]. Diese Beobachtungen werden hauptsächlich auf die größere Anonymität postalischer Befragungen zurückgeführt. In einigen anderen Studien werden für unterschiedliche Befragungsmodi hingegen keine relevanten Verzerrungen im Antwortverhalten festgestellt [7, 8]. Diese uneinheitliche Befundlage führt zu der Notwendigkeit, den Einfluss unterschiedlicher Erhebungsmethoden auf genau abgegrenzte Themenbereiche und verschiedene Befragungsinstrumente explizit empirisch zu prüfen. Besonders vielschichtig stellen sich mögliche Methodeneinflüsse auf komplexe Konstrukte wie die gesundheitsbezogene Lebensqualität (HRQoL) dar. Das Konzept der gesundheitsbezogenen Lebensqualität umfasst bei der Betrachtung von Gesundheit physiologische, emotionale, psychische sowie soziale Aspekte und hat sich als fester Bestandteil der Gesundheitsforschung etabliert.

Erhebungen mithilfe des Short Form 36 Health Survey (SF-36), der als das international weitverbreitetste Messinstrument zur Erfassung der gesundheitsbezogenen Lebensqualität gilt [12, 13], können nicht als methodenunabhängig angesehen werden. McHorney et al. [4] dokumentieren für 7 von 8 Skalen des SF-36 signifikant höhere Mittelwerte und damit positivere Einschätzungen der Gesundheit für telefo-

nisch als für postalisch Befragte einer umfangreichen amerikanischen Normstichprobe. Der schriftliche und telefonische Einsatz des australischen SF-36 bei einer populationsbasierten Zufallsstichprobe [3] führte in 4 Skalen zu signifikant positiveren Berichten seitens der telefonisch Befragten. Ware et al. [14] weisen darauf hin, dass sich die Methodeneffekte besonders auf die mit psychischer Gesundheit assoziierten Skalen auswirken, sodass für Telefonbefragungen mit einer Abweichung des psychischen Summenscores (MCS) von bis zu 0,2 Standardabweichungen in positiver Richtung zu rechnen ist.

Der SF-8 (Shortform-8-Questionnaire) von Ware et al. [14] wurde als Kurzform des SF-36 konzipiert und dient ebenso wie dieser der krankheitsübergreifenden Erfassung des subjektiven Gesundheitszustandes von Erwachsenen in Bezug auf psychische, körperliche und soziale Aspekte. Mit dem SF-8 steht ein aufgrund seiner hohen Ökonomie für epidemiologische Studien besonders geeignetes Instrument zur Erfassung der gesundheitsbezogenen Lebensqualität zur Verfügung, das im telefonisch durchgeführten Bundes-Gesundheitssurvey 2002/2003 erstmals eingesetzt wurde [15]. Die 8-Itemversion wurde speziell für bevölkerungsbezogene Umfragen konzipiert, wird darüber hinaus aber auch in Studien mit klinisch relevanten Subgruppen eingesetzt [16, 17, 18, 19, 20]. Jeweils ein Item des SF-8 bildet jeweils eine der 8 Skalen des SF-36 ab: körperliche Funktions-

fähigkeit (PF), körperliche Rollenfunktion (RP), Schmerz (BP), allgemeine Gesundheitswahrnehmung (GH), Vitalität (VT), soziale Funktionsfähigkeit (SF), emotionale Rollenfunktion (RE), psychisches Wohlbefinden (MH). Die Erstellung des 8-dimensionalen Gesamtprofils des SF-8 erfolgt ebenso wie die Bildung der beiden Summenskalen (PCS: körperlicher Summenscore; MCS: psychischer Summenscore), anhand der Standardmetrik des SF-36. Entsprechend ist jeder SF-8-Wert direkt mit dem jeweiligen SF-36-Wert vergleichbar und analog zu interpretieren.

Inwieweit die für den SF-36 bekannten methodischen Implikationen auch für den SF-8 Berücksichtigung finden müssen, ist bislang unzureichend evaluiert. Ware et al. [14] überprüften während der Normierung des SF-8 Effekte verschiedener Datenerhebungsmethoden auf die Skalenmittelwerte und Summenscores des Instruments. Verglichen mit einem aus postalischen und internetbasierten Befragungen kombinierten Sample wurden in den telefonisch erhobenen Daten signifikant höhere Mittelwerte in 6 von 8 Skalen sowie in beiden Summenscores beobachtet. Lediglich die Skalen „Körperliche Funktionsfähigkeit“ und „Allgemeine Gesundheitswahrnehmung“ wiesen äquivalente Befragungsergebnisse unabhängig von der Erhebungsmethode auf. Die gefundenen Unterschiede variierten in einem Bereich von ca. 0,1–0,2 Standardabweichungen und waren im Fall des psychischen Summenwertes etwa doppelt so hoch wie im Fall des physischen Summenwertes. Einschränkungen hinsichtlich dieser Untersuchung ergeben sich aus dem Verfahren der pragmatischen Auswahl zur Stichprobengewinnung ohne die randomisierte Zuteilung der Probanden zu einer Befragungsmethode und aus der bereits erwähnten Kombination postalisch und via Internet erhobener Daten.

Ziel der hier vorliegenden Studie ist es, mögliche Auswirkungen der telefonischen und postalischen Befragungsmethode in einer deutschen Zufallsstichprobe von 1690 Erwachsenen auf den SF-8 zu untersuchen. Verglichen werden neben den Stichprobenzusammensetzungen insbesondere die Datenqualität hinsichtlich bestehender Mittelwerts- und Varianzun-

Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 2005 · 48:1322–1329
DOI 10.1007/s00103-005-1165-8
© Springer Medizin Verlag 2005

M. Erhart · R. Wetzel · A. Krügel · U. Ravens-Sieberer

Erfassung der gesundheitsbezogenen Lebensqualität mit dem deutschen SF-8. Ein Vergleich der telefonischen und postalischen Befragungsmethode

Zusammenfassung

Im Rahmen eines Vergleiches telefonischer und postalischer Befragungsmethoden wurde der SF-8 als Instrument zur Messung der gesundheitsbezogenen Lebensqualität bei Erwachsenen eingesetzt. Die aus 1690 Personen im Alter zwischen 25 und 66 Jahren bestehende Stichprobe wurde randomisiert zur Hälfte telefonisch und zur anderen Hälfte postalisch befragt. Verglichen wurden die je nach Darbietungsmodus erzielten Ausfüllraten, zentralen Tendenzen, Streuungen, Decken- und Bodeneffekte in den Messwerten des SF-8 sowie die Korrelationen zwischen den Items. Die Bedeutung des Alters und des Geschlechtes für die Unterschiede wurde

untersucht. Die Resultate des Methodenvergleiches zeigen je nach Darbietungsmodus und erfragtem Gesundheitsaspekt keine oder geringe statistisch signifikante Unterschiede in der Beantwortung des SF-8. Als Fazit sollte in weiteren Untersuchungen die genaue Natur der Abweichungen untersucht und Korrekturfaktoren ermittelt werden.

Schlüsselwörter

Gesundheitsbezogene Lebensqualität · SF-8 · Computerassistierte Telefoninterviews · Postalische Befragung · Methodenvergleich

Assessment of health-related quality of life with the German SF-8. A comparison of telephone and postal survey modes

Abstract

Within a comprehensive comparison of telephone and postal survey methods the SF-8 was applied to assess adult's health-related quality of life. The 1690 subjects were randomly assigned to a telephone survey and a postal survey. Comparisons across the different modes of administration addressed the response rates, central tendency, deviation, ceiling and floor effects observed in the SF-8 scores as well as the inter-item correlation. The importance of age and gender as moderating factors was investigated. Results indicate no or small statistically signifi-

cant differences in the responses to the SF-8 depending on the actual mode of administration and the health aspect questioned. It was concluded that further investigations should focus on the exact nature of these deviations and try to generate correction factors.

Keywords

Health-related quality of life · SF-8 · Computer assisted telephone interview · Postal survey · Comparison of methods

Tabelle 1

Soziodemografische Charakteristiken nach Darbietungsmodus der Befragung				
	Postalisch (n=899)		Telefonisch (n=791)	
	(n)	[%]	(n)	[%]
Alter				
25–34 Jahre	69	7,7	56	7,1
35–44 Jahre	591	65,9	501	63,3
45–54 Jahre	223	24,9	218	27,6
55+	14	1,6	16	2
χ^2 -Wert (df=3)	2,312, p=0,510			
Altersdurchschnitt (SD)	41,72 (5,43)		41,75 (5,49)	
Geschlecht				
Weiblich	749	84,3	658	83,3
Männlich	140	15,7	133	16,8
χ^2 -Wert (df=1)	0,350, p=0,554			
Familienstand				
Ledig	69	7,7	39	6,5
Verheiratet	682	76,5	644	81,4
Getrennt lebend	35	3,9	23	2,9
Geschieden	97	10,9	78	9,9
Verwitwet	9	1	7	0,9
χ^2 -Wert (df=4)	8,186, p=0,085			
Bildungsstatus				
Keinen Abschluss	20	2,2	8	1
Volks-/Hauptschule	209	23,4	165	20,9
Mittlere Reife	369	41,2	333	42,2
Abitur	114	12,7	111	14,1
Fachhochschulabschluss	81	9,1	70	8,9
Universitätsabschluss	102	11,3	103	13
χ^2 -Wert (df=5)	6,494, p=0,261			
Berufsstatus				
Vollzeit	316	35,5	267	33,8
Teilzeit	347	39	311	39,3
Gelegentlich beschäftigt	37	4,2	47	5,9
Arbeitslos	35	3,9	32	4
Hausfrau/Hausmann	136	15,3	113	14,3
Student/in, Rentner/in	10	1,1	13	1,6
Sonstiges	9	1	8	1
χ^2 -Wert (df=6)	4,171, p=0,653			

terschiede sowie des Vorkommens von Boden- und Deckeneffekten.

Stichprobe und Methode

Die zugrunde liegenden Daten wurden 2003 im Rahmen eines umfassenden Methodenvergleichs erhoben. Verglichen wurden telefonische und postalische Befragungsmethoden, um deren Einfluss auf die mit verschiedenen Instrumenten zur subjektiven Gesundheit erhobenen Daten zu untersuchen. Zugrunde gelegt wurde eine 2-stufig geschichtete Zu-

fallsstichprobe von 4000 Haushalten: Zunächst wurde in Zusammenarbeit mit dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) aus Mannheim eine Auswahl von 42 bundesdeutschen Gemeinden nach Merkmalen der Region und Ortsklassengröße vorgenommen; in einem zweiten Schritt wurde eine Zufallsauswahl aus allen in den jeweiligen Einwohnermelderegistern verzeichneten Familien mit Kindern dieser Gemeinden getroffen. Jeder Befragungsmethode wurden je 2000 der Haushalte zufällig zugewiesen.

Alle Haushalte wurden in einem Informationsbrief über die durchführende Institution sowie den Inhalt der Befragung informiert und – in Abhängigkeit von der zuvor festgelegten Befragungsmethode – um die Teilnahme eines Elternteils und eines Kindes an einem Telefoninterview bzw. um das Ausfüllen der beigefügten Fragebögen gebeten. Um die Ausschöpfung der Telefonstichprobe zu erhöhen, wurden diejenigen Haushalte, deren Telefonnummer nicht im Telefonbuch verzeichnet war, im ansonsten gleich lautenden Informationsbrief um die schriftliche Mitteilung der Telefonnummer gebeten. Die Telefonbefragung wurde mittels computergestützter Telefoninterviews (CATI) durchgeführt. Geschulte Interviewer kontaktierten die Haushalte 1–2 Wochen nach dem Versand des Informationsbriefes bzw. nach Erhalt der Telefonnummer und baten um die Interviewteilnahme. Eltern und Kinder einer Familie wurden jeweils getrennt voneinander befragt. Der Fragebogen umfasste neben Fragen zur Soziodemographie und Akzeptanz Fragenkomplexe zu körperlichen Beschwerden, zum Gesundheitsverhalten und zur selbst eingeschätzten gesundheitsbezogenen Lebensqualität, unter anderem erfasst mit dem SF-8. Die Daten wurden in eine Datenbank eingegeben, plausibilitätsgeprüft und mit dem Programmpaket SPSS ausgewertet.

Die soziodemografischen und sozioökonomischen Charakteristiken der telefonisch befragten Personen wurden mit denen der postalisch befragten Personen verglichen und mittels χ^2 -Tests auf signifikante Unterschiede getestet. Effekte des Darbietungsmodus wurden analysiert, indem die Mittelwerte auf den SF-8-Skalen varianzanalytisch miteinander verglichen wurden. Streuungsunterschiede in den Antworten wurden durch Levene-Tests untersucht. Der Prozentsatz von Befragten beider Stichproben mit maximalem oder minimalem Skalenwert wurde mittels χ^2 -Tests verglichen. Die Konsistenz des Antwortverhaltens bzw. die Korrelationen zwischen den Items wurden getrennt für beide Erhebungsmodi berechnet. Obwohl einige Analysen eine mehrfache Prüfung derselben Hypothese beinhalten, erfolgte keine Korrektur des Alpha-Fehler-Niveaus, da an-

Tabelle 2

Verteilungseigenschaften der SF-8-Items und -Summenwerte nach Darbietungsmodus der Befragung (n=1690)

SF-8-Items und -Skalen	Postalisch (n=899)		Telefonisch (n=791)		Δ -MW Effektgröße ^a ,d ^{ca}	Δ -MW ANOVA	Δ -SD Levene-Test
	MW	SD	MW	SD			
Allgemeine Gesundheitswahrnehmung	48,03	6,34	48,58	6,9	0,08	ns	**
Körperliche Funktionsfähigkeit	49,83	6,27	49,76	6,55	0,01	ns	ns
Körperliche Rollenfunktion	50,18	6,13	49,66	6,68	0,08	ns	*
Körperliche Schmerzen	53,15	8,71	52,34	9,64	0,09	ns	***
Vitalität	51,30	7,29	52,28	7,62	0,13	**	ns
Soziale Funktionsfähigkeit	50,98	6,50	51,20	6,38	0,03	ns	ns
Psychisches Wohlbefinden	48,75	8,55	50,13	8,11	0,17	***	ns
Emotionale Rollenfunktion	48,85	5,91	49,39	5,53	0,10	ns	*
Körperlicher Summenwert	51,75	7,64	50,96	8,29	0,10	*	ns
Psychischer Summenwert	49,83	8,95	51,42	8,24	0,18	***	ns

^a Klassifikation des Effektes: 0,20=klein; 0,50=moderat; 0,80=groß, * p<0,05; ** p<0,01; *** p<0,001.

Tabelle 3

Interitemkorrelationen nach Darbietungsmodus der Befragung (n=1690): untere Dreiecksmatrix=postalische, obere Dreiecksmatrix=telefonische Darbietung

SF-8-Items und -Skalen	GH	PF	RP	BP	VT	SF	MH	RE	Telefonische Befragung
GH		0,50 ^a	0,55	0,51	0,46 ^a	0,40	0,39 ^a	0,41	
PF	0,58 ^a		0,76	0,53	0,38	0,29 ^a	0,28	0,41	
RP	0,59	0,76		0,60	0,41	0,38	0,35	0,53	
BP	0,56	0,55	0,57		0,35	0,27	0,22	0,33	
VT	0,56 ^a	0,45	0,46	0,36		0,37 ^a	0,39 ^a	0,37	
SF	0,47	0,40 ^a	0,44	0,26	0,50 ^a		0,50 ^a	0,51 ^a	
MH	0,47 ^a	0,27	0,32	0,22	0,48 ^a	0,58 ^a		0,50	
RE	0,48	0,49	0,55	0,33	0,44	0,59 ^a	0,57		

Postalische Befragung

GH Allgemeine Gesundheitswahrnehmung, PF Körperliche Funktionsfähigkeit, RP Körperliche Rollenfunktion, BP Körperliche Schmerzen, VT Vitalität, SF Soziale Funktionsfähigkeit, MH Psychisches Wohlbefinden, RE Emotionale Rollenfunktion, ^a signifikanter Unterschied.

sonsten das Risiko, einen auf Populations-ebene bestehenden Unterschied nicht zu erkennen, vergrößert würde. Von einem signifikanten Unterschied wurde ausgegangen, wenn die Überschreitungswahrscheinlichkeit einer Prüfgröße geringer als 5% war. Die Antworten auf die SF-8-Items wurden mit den ihnen in der amerikanischen Normstichprobe durchschnittlich entsprechenden SF-36-Skalenwerten kodiert. Der körperliche und der psychische SF-8-Summenwert wurden berechnet, indem die kodierten Itemantworten mit aus der amerikanischen Normstichprobe resultierenden Regressionsgewichten zur Vorhersage der SF-36-Summen-

werte multipliziert und anschließend zusammen mit der Regressionskonstante aufaddiert wurden [14]. Diese Prozedur entspricht bei normal verteilten SF-36-Skalenwerten annäherungsweise einer Normalisierungstransformation.

Ergebnisse

Von den 1928 kontaktierbaren Haushalten der postalischen Stichprobe nahmen 916 Haushalte an der Befragung teil, was einer Responserate von 47,5% entspricht. Innerhalb der Telefonstichprobe konnten 1066 Haushalte kontaktiert werden. Die Teilnahme von 825 Haushalten entspricht

einer Responserate von 77,4%. Insgesamt konnten die SF-8-Gesundheitsangaben von 791 telefonisch und 899 postalisch befragten Personen ausgewertet werden.¹ Die soziodemografischen und sozioökonomischen Charakteristika dieser beiden Stichproben werden in der [Tabelle 1](#)

¹ Da in die oben genannten Responseraten bei der Stichproben auch diejenigen Haushalte einbezogen wurden, in denen nur eine Befragung, also eine Eltern- oder eine Kinderbefragung, durchgeführt werden konnte, weicht die Zahl der zur Verfügung stehenden Elternbefragungen von der Summe der insgesamt befragten Haushalte ab.

Tabelle 4

Verteilungseigenschaften der SF-8-Items und -summenwerte nach Darbietungsmodus der Befragung bei Frauen (n=1407)

SF-8-Items und -Skalen	Postalisch (n=749)		Telefonisch (n=658)		Δ-MW Effektgröße „d“ ^a	Δ-MW ANOVA	Δ-SD Levene-Test
	MW	SD	MW	SD			
Allgemeine Gesundheitswahrnehmung	47,90	6,3	48,55	6,80	0,10	ns	**
Körperliche Funktionsfähigkeit	49,88	6,27	49,83	6,31	0,01	ns	ns
Körperliche Rollenfunktion	50,18	6,08	49,71	6,41	0,08	ns	ns
Körperliche Schmerzen	52,97	8,81	52,40	9,59	0,06	ns	**
Vitalität	51,18	7,33	51,96	7,71	0,10	ns	ns
Soziale Funktionsfähigkeit	50,92	6,51	51,15	6,41	0,04	ns	ns
Psychisches Wohlbefinden	48,52	8,56	49,81	8,16	0,15	**	ns
Emotionale Rollenfunktion	48,92	5,86	49,42	5,43	0,09	ns	*
Körperlicher Summenwert	51,72	7,56	51,10	7,99	0,08	ns	ns
Psychischer Summenwert	49,61	9,01	51,09	8,33	0,17	**	ns

^a Klassifikation des Effektes: 0,20=klein; 0,50=moderat; 0,80=groß; * p<0,05; ** p<0,01; *** p<0,001.

Tabelle 5

Verteilungseigenschaften der SF-8-Items und -Summenwerte nach Darbietungsmodus der Befragung bei Männern (n=273)

SF-8-Items und -Skalen	Postalisch (n=140)		Telefonisch (n=133)		Δ-MW Effektgröße „d“ ^a	Δ-MW ANOVA	Δ-SD Levene-Test
	MW	SD	MW	SD			
Allgemeine Gesundheitswahrnehmung	48,68	6,51	48,73	7,41	0,01	ns	ns
Körperliche Funktionsfähigkeit	49,73	6,33	49,44	7,68	0,04	ns	ns
Körperliche Rollenfunktion	50,22	6,37	49,45	7,93	0,11	ns	*
Körperliche Schmerzen	54,26	7,96	52,06	9,94	0,24	*	**
Vitalität	51,89	6,98	53,86	7,00	0,28	*	ns
Soziale Funktionsfähigkeit	51,30	6,56	51,42	6,26	0,02	ns	ns
Psychisches Wohlbefinden	50,12	7,92	51,73	7,68	0,21	ns	ns
Emotionale Rollenfunktion	48,56	6,20	49,29	6,01	0,12	ns	ns
Körperlicher Summenwert	51,84	7,99	50,29	9,66	0,18	ns	ns
Psychischer Summenwert	50,90	8,50	53,06	7,57	0,26	**	ns

^a Klassifikation des Effektes: 0,20: klein; 0,50: moderat; 0,80: groß; * p<0,05; ** p<0,01; *** p<0,001.

präsentiert. Die hinter den Stichproben stehenden Populationen unterscheiden sich nicht in ihrer Alters- und Geschlechtsverteilung. In beiden Gruppen sind die Befragten durchschnittlich 41,5 Jahre alt, bei einer Standardabweichung von ungefähr 5,5 Jahren. Das Alterspektrum reicht von 25–66 Jahre. Sowohl telefonisch als auch postalisch wurden mit einem Anteil von jeweils über 80% mehrheitlich Frauen befragt. Auch hinsichtlich des Familienstandes, des Bildungsstatus und des beruflichen Status der Probanden finden sich keine signifikanten Unterschiede.

Nach der Durchführung einfaktorier Varianzanalysen ergaben sich zwischen den beiden Darbietungsmodi die folgenden in **■ Tabelle 2** dargestellten Unterschiede im Antwortverhalten: Die postalisch Befragten erzielen mit 51,75 Punkten im Mittel einen höheren körperlichen Summenwert als die telefonisch Befragten, die im Durchschnitt 50,96 Punkte erreichen ($F_{[df_{1;1676}]}=4,111; p=0,043$). Die telefonisch Befragten geben durchschnittlich eine höhere Vitalität ($F_{[df_{1;1687}]}=7,289; p=0,007$) und ein besseres psychisches Wohlbefinden ($F_{[df_{1;1687}]}=11,566; p=0,001$)

an als die postalisch Befragten. Darüber hinaus erzielen sie mit 51,42 Punkten (postalisch 49,83 Punkte) im Mittel auch einen höheren psychischen Summenwert, ($F_{[df_{1;1676}]}=14,227; p<0,001$). Nach Transformation in das „d“-Effektstärkemaß können die gefundenen Mittelwertdifferenzen maximal als kleine Effekte klassifiziert werden [21].

Neben den Unterschieden in der zentralen Tendenz der Beantwortung wurde überprüft, ob sich das Antwortverhalten je nach Darbietungsmodus auch hinsichtlich der Varianz in den Gesund-

Tabelle 6

Überprüfung der Interaktion zwischen Alter und Darbietungsmodus der Befragung: Kovarianzanalysen nach dem allgemeinen linearen Modell (basierend auf multiplen linearen Regressionen)

SF-8-Items und -Skalen	GH	PF	RP	BP	VT	SF	MH	RE	PCS	MCS
Prädiktoren										
<i>Modus (tel.)</i>										
b	3,19	0,68	-0,33	3,11	1,36	1,91	4,17	3,06	0,74	4,46
95%-KI unter	-1,73	-4,10	-5,09	-3,71	-4,18	-2,89	-2,04	-1,23	-5,24	-2,01
95%-KI ober	8,11	5,45	4,43	9,92	6,90	6,72	10,39	7,35	6,72	10,93
<i>Alter</i>										
b	-0,03	-0,01	0,01	0,03	0,07	0,08*	0,07	0,06	0,01	0,11*
95%-KI unter	-0,05	-0,09	-0,06	-0,08	-0,02	<0,01	-0,03	-0,01	-0,09	<0,01
95%-KI ober	0,11	0,06	0,09	0,13	0,16	0,15	0,16	0,13	0,10	0,21
<i>Modus (tel.)* Alter</i>										
b	-0,06	-0,02	-0,01	-0,09	-0,01	-0,04	-0,07	-0,06	-0,04	-0,07
95%-KI unter	-0,18	-0,13	-0,12	-0,26	-0,14	-0,15	-0,21	-0,16	-0,18	-0,22
95%-KI ober	0,05	0,10	0,11	0,07	0,12	0,07	0,08	0,04	0,11	0,09
Konstante	46,85	50,28	49,7	52,07	48,35	47,81	46,11	46,39	51,39	45,41
95%-KI unter	43,58	47,11	46,53	47,54	44,67	44,62	41,98	43,52	47,40	41,09
95%-KI ober	50,12	53,45	52,86	56,59	52,03	50,99	50,24	49,25	55,38	49,73
R ²	<0,01	<0,01	<0,01	<0,01	0,01	<0,01	0,01	<0,01	<0,01	0,01

GH Allgemeine Gesundheitswahrnehmung, PF Körperliche Funktionsfähigkeit, RP Körperliche Rollenfunktion, BP Körperliche Schmerzen, VT Vitalität, SF Soziale Funktionsfähigkeit, MH Psychisches Wohlbefinden, RE Emotionale Rollenfunktion, b Regressionsgewichte (Rohgewichte) geben den mit der Ausprägung der jeweiligen Prädiktorvariablen verbundenen Effekt in den SF-8-Skalen wieder, Signifikanz der b Regressionsgewichte * p<0,05; ** p<0,01; *** p<0,001.

heitsangaben sowie der Benutzung der extremen Antwortkategorien unterscheidet: Die allgemeine Gesundheitswahrnehmung variiert bei den telefonisch Befragten stärker als bei den postalisch Befragten (Levene-Statistik $_{[df_{1;1687}]}=8,751$; $p=0,003$). Auch auf den Items zur körperlichen Rollenfunktion (Levene-Statistik $_{[df_{1;1686}]}=5,006$; $p=0,025$) und zu körperlichen Schmerzen (Levene-Statistik $_{[df_{1;1686}]}=15,429$; $p<0,001$) weisen die telefonisch Befragten in ihren Antworten eine signifikant größere Varianz auf. Lediglich die Antworten auf das Item emotionale Rollenfunktion variieren bei den postalisch Befragten stärker (Levene-Statistik $_{[df_{1;1685}]}=6,216$; $p=0,013$). Die durchgeführten χ^2 -Tests zeigen, dass die telefonisch Befragten bei dem Item Allgemeine Gesundheitswahrnehmung mit 15,2%, im Gegensatz zu 9,6% der postalisch Befragten, signifikant häufiger die höchste Antwortkategorie „ausgezeichnet“ wählten ($\chi^2_{[df_{1}]}=12,348$; $p<0,001$). Für das Item Vitalität wurde die höchste Antwortkategorie „ausgezeichnet“ von den telefonisch Befragten mit 19,1% sogar fast doppelt so häufig gewählt wie von den postalisch Befragten (10,1%) ($\chi^2_{[df_{1}]}=27,579$,

$p<0,001$). Ebenfalls signifikant häufiger gaben die telefonisch Befragten die höchste Antwortkategorie des Items psychisches Wohlbefinden an ($\chi^2_{[df_{1}]}=22,418$; $p<0,001$). Das Prozentverhältnis betrug 44,4% zu 33,1%. Die niedrigste Antwortkategorie des Items Körperliche Schmerzen wurde mit einem Anteil von 2% signifikant häufiger von den telefonisch Befragten gewählt ($\chi^2_{[df_{1}]}=10,799$; $p=0,001$).

Weiterhin wurde die Konsistenz des Antwortverhaltens bzw. die Korrelation zwischen den Items des SF-8 für beide Darbietungsformen berechnet. Die Resultate sind in **Tabelle 3** dargestellt. Bei den postalisch Befragten fallen die Korrelationen zwischen den Items geringfügig höher aus, einige Unterschiede sind nach dem auf der Fishers-Z-Transformation basierenden Test signifikant. Die größten Unterschiede betreffen die Korrelationen zwischen den Items Vitalität und Soziale Funktionsfähigkeit sowie Vitalität und Allgemeine Gesundheitswahrnehmung und können nach Umrechnung in Fishers Z-Werte als kleine Effekte klassifiziert werden [21].

In einem nächsten Schritt wurde untersucht, inwieweit Geschlecht und Al-

ter einen moderierenden Einfluss auf den Effekt des Darbietungsmodus haben. Zunächst wurden die Lage- und Streuungsunterschiede in den Antworten beider Gruppen getrennt für Frauen und Männer untersucht. **Tabelle 4** stellt die Resultate für Frauen dar. Signifikant höhere Werte erzielten die telefonisch befragten Frauen auf dem Item Psychisches Wohlbefinden ($F_{[df_{1;1404}]}=8,195$; $p=0,004$) und dem Psychischen Summenwert ($F_{[df_{1;1397}]}=10,096$; $p=0,002$). Die Stärke der gefundenen Effekte kann erneut maximal als klein klassifiziert werden. Die Antworten der untersuchten Frauen variieren bei telefonischer Befragung stärker als bei postalischer Befragung auf den Items Allgemeine Gesundheitswahrnehmung (Levene-Statistik $_{[df_{1;1405}]}=7,741$; $p=0,005$) und Körperliche Schmerzen (Levene-Statistik $_{[df_{1;1403}]}=8,638$; $p=0,003$). Eine geringere Varianz in den Antworten der telefonisch befragten Frauen ist dagegen bei dem Item Emotionale Rollenfunktion festzustellen (Levene-Statistik $_{[df_{1;1403}]}=5,271$; $p=0,022$).

Betrachtet man nun die Angaben der untersuchten Männer für das Item Körperliche Schmerzen, zeigen sich signifi-

kant höhere Angaben bei den postalisch befragten ($F_{[df\ 1;271]}=4,061; p=0,045$). Dagegen geben telefonisch befragte Männer überzufällig bessere Einschätzungen auf dem Item Vitalität ($F_{[df\ 1;271]}=5,446; p=0,020$) an und erzielen im Durchschnitt auch einen höheren psychischen Summenwert als die postalisch befragten ($F_{[df\ 1;269]}=4,813; p=0,029$). Die Stärke der gefundenen Effekte kann maximal als klein klassifiziert werden, durchschnittlich fallen die Effektstärken allerdings höher aus als bei den Frauen. Die Antworten der telefonisch befragten Männer variieren auf den Items Körperliche Rollenfunktion (Levene-Statistik ($df\ 1;271$)= $4,494; p=0,035$) und Körperliche Schmerzen (Levene-Statistik ($df\ 1;271$)= $8,925; p=0,003$) stärker als die der postalisch befragten Männer (■ **Tabelle 5**).

Die moderierende Funktion des Alters auf den Effekt des Darbietungsmodus wurde in Kovarianzanalysen für jeden SF-8-Messwert überprüft. Die Varianzaufklärung der spezifischen Interaktion zwischen der Kovariate Alter und dem Faktor Darbietungsmodus wurde auf Signifikanz geprüft. Den in ■ **Tabelle 6** dargestellten Ergebnissen ist zu entnehmen, dass für keinen SF-8-Messwert eine signifikante Interaktion zwischen dem Alter und dem Darbietungsmodus nachgewiesen werden konnte: Die Auswirkungen des Darbietungsmodus auf die zentrale Tendenz der Beantwortung bleibt über das Alter gleich.

In Kovarianzanalysen für den körperlichen und den psychischen Summenwert mit der Erhebungsmethode, dem Alter, dem Geschlecht und der Schulbildung (kodiert nach Fachhochschulabschluss/Abitur/Studium, Mittlere Reife und der Referenzgruppe ohne Abschluss/Grundschulabschluss/Hauptschulabschluss) sowie deren 2-Wege- und 3-Wege-Wechselwirkungen als erklärende Faktoren bzw. Kovariaten wurde überprüft, ob diese Einflussgrößen in den beiden durch die Erhebungsmethode definierten Teilstichproben gleichartig wirken. Keine der getesteten Wechselwirkungen erreichte statistische Signifikanz.

Diskussion

In dieser Studie wurden die Auswirkungen unterschiedlicher Befragungsmethoden auf die Beantwortung des SF-8-Frage-

bogens untersucht. Eine Zufallsstichprobe wurde zur Hälfte telefonisch und zur anderen Hälfte postalisch befragt. Die bereits berichtete theoretisch zu erwartende höhere Responserate telefonischer Befragungen zeigte sich auch in dieser Studie. Diese hatte jedoch ebenso wie die insgesamt höhere Ausschöpfungsrate bei der schriftlichen Befragung keine signifikanten Unterschiede in den soziodemografischen und sozioökonomischen Eigenschaften zwischen den beiden Teilstichproben zur Folge.

Da beide Befragungsmethoden zu vergleichbaren Stichproben führten und außerdem keine spezifische Interaktion zwischen den überprüften Merkmalen hinsichtlich der Beantwortung des SF-8 vorliegt, können in dieser Studie Abweichungen in der Beantwortung des SF-8 auf die unterschiedlichen Darbietungsformen bzw. deren Interaktion mit anderen Merkmalen zurückgeführt werden. Die telefonisch Befragten machen überzufällig bessere Gesundheitsangaben auf einigen der psychologischen SF-8-Skalen. Dieses Resultat steht in Übereinstimmung mit anderen Untersuchungen [4, 14], wobei positivere Angaben bei persönlichen Befragungsmethoden oft mit dem Wegfallen der größeren Anonymität postalischer Befragungen in Zusammenhang gebracht werden. Die absolute Abweichung entspricht jedoch maximal 1–1,6 Punkte, was bei einer Standardabweichung von ungefähren 8 Punkten bestenfalls einem kleinen Effekt entspricht [21]. Im Gegensatz zu bisherigen Befunden [4, 14] und den nun vorliegenden Ergebnissen bezüglich der psychischen Items und Skalen berichten die postalisch Befragten auf der körperlichen Summenskala einen besseren Gesundheitszustand. Möglicherweise führt die geringere Anonymität bei telefonischer Befragung zwar bei den die psychische Verfassung betreffenden Fragen zu einem positiveren Antwortverhalten, jedoch nicht bei Fragen, die die körperliche Verfassung betreffen. Dies könnte mit einer größeren Scham hinsichtlich der Berichte psychischer Probleme erklärt werden, während körperliche Beschwerden weniger tabuisiert sind. Eine negativere Beantwortung der körperlichen Fragen als eine Reaktion auf geschönte Angaben zu den psychischen

Fragen ist unwahrscheinlich, da die körperbezogenen Items als Erste dargeboten werden. Da die gefundene Effektstärke auch hier nur klein ausfällt, bleibt zu berücksichtigen, dass die praktische Bedeutsamkeit des beschriebenen Unterschieds eher gering ist.

Insgesamt streuen die Gesundheitsangaben eher bei telefonischer Befragung stärker als bei postalischer Befragung, d. h., die Befragten nutzen am Telefon eher den gesamten Bereich der Antwortvorgaben. Dies ist möglicherweise eine Folge der wiederholten mündlichen Darbietung der Antwortvorgaben.

Die Korrelationen zwischen den einzelnen Items fallen bei den postalisch Befragten geringfügig höher aus. Dieser Befund könnte dadurch erklärt werden, dass die visuelle Vorlage der Item-Antwortkategorien dazu verleitet, ein konsistenteres Antwortmuster zu geben. Praktisch würde dies allerdings bedeuten, dass die Beantwortung eines Items die Antwort auf andere Items beeinflusst, das heißt, die lokale Unabhängigkeit der Item-Antworten, die eine bedeutsame Annahme verschiedener Testmodelle darstellt [22], wäre verletzt.

Die getrennten Analysen für Frauen und Männer können aufgrund der wegen kleinerer Fallzahlen verringerten Teststärke bestehende Methodeneffekte nur mit einer geringeren Wahrscheinlichkeit nachweisen. Ihre Interpretationsmöglichkeit ist dadurch eingeschränkt. Sofern signifikante Resultate vorliegen, bestätigen diese die generellen Befunde. Tendenziell fallen die Unterschiede zwischen den Darbietungsmodi bei Männern stärker aus als bei Frauen, allerdings ist hierzu anzumerken, dass die geringere Anzahl befragter Männer zu unsicheren Schätzungen der Effektstärken führt. Das Alter hat, bei Prüfung der linearen Auswirkung auf den Methodeneffekt, nach den vorliegenden Ergebnissen keine Bedeutung für die Ausprägung des Methodeneffektes.

Fazit

Als Fazit führt die Anwendung verschiedener Darbietungsmodi zu einigen überzufälligen Unterschieden in der Beantwortung des SF-8. Die tatsächlichen Effekte sind jedoch eher von geringer praktischer Bedeutung. Welcher Darbietungs-

modus die validere und reliablere sowie am wenigsten verzerrte Messung ermöglicht, kann im Rahmen dieser Studie nicht geklärt werden und sollte Gegenstand weiterer Untersuchungen sein. Ware et al. [14] empfehlen aufgrund einer ähnlichen Untersuchung die Anwendung von Korrekturfaktoren bei telefonischer Befragung. Nach den vorliegenden Resultaten könnten zu diesem Zweck die gefundenen Mittelwertdifferenzen von den telefonisch erhobenen Messwerten subtrahiert werden. Die vorliegenden Resultate weichen jedoch in Teilen von den Befunden von Ware et al. ab [14] und können aufgrund des Untersuchungsdesigns nur auf eine eingeschränkte Population verallgemeinert werden. Aus diesem Grund sollen an dieser Stelle keine Empfehlungen für Korrekturen ausgesprochen werden. Beim Einsatz des SF-8 muss bei telefonischen Befragungen generell mit Abweichungen von postalischen Resultaten gerechnet werden. Eine direkte Vergleichbarkeit der Ergebnisse ist daher nur mit Einschränkungen möglich.

Korrespondierender Autor

PD Dr. U. Ravens-Sieberer

Forschungsgruppe Psychosoziale Gesundheit, Robert Koch-Institut, Postfach 650261, 13302 Berlin
E-Mail: Ravens-SiebererU@rki.de

Literatur

1. Battistutta D, Byth K, Norton R, Rose G (1983) Response rates: a comparison of mail, telephone and person interview strategies for an Australian population. *Comm Health Svcs* 7:309–313
2. O’Toole BI, Battistutta D, Long A, Crouch K (1986) A comparison of costs and data quality of three health survey methods: mail, telephone and personal home interview. *Am J Epidemiol* 124:317–328
3. Perkins JJ, Sanson-Fisher RW (1998) An examination of self- and telephon-administered modes of administration for the Australian SF-36. *J Clin Epidemiol* 51:969–973
4. McHorney CA, Kosinski M, Ware JE (1994) Comparisons of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: results from a national survey. *Med Care* 32:551–567
5. Galobardes B, Sunyer J, Antó JM et al. (1998) Effect of the method of administration, mail or telephone, on the validity and reliability of a respiratory health questionnaire. The Spanish centers of the European asthma study. *J Clin Epidemiol* 51:875–881

6. Brambilla DJ, McKinley SM (1987) A comparison of responses to mailed questionnaires and telephone interviews in a mixed mode health survey. *Am J Epidemiol* 126:962–971
7. Needle RH, Jou SC, Su SS (1989) The impact of changing methods of data collection on the reliability of self-reported drug use of adolescents. *Am J Drug Alcohol Abuse* 15:275–289
8. Hawthorne G (2003) The effect of different methods of collecting data: mail, telephone and filter data collection issues in utility measurement. *Qual Life Res* 12:1081–1088
9. Brøgger J, Bakke P, Eide GE, Gulsvik A (2002) Comparison of telephone and postal survey modes on respiratory symptoms and risk factors. *Am J Epidemiol* 155:572–576
10. Gmel G (2000) The effect of mode of data collection and of non-response on reported alcohol consumption: a split-sample study in Switzerland. *Addiction* 95:123–134
11. Siemiątycki J, Campbell S, Richardson L, Aubert D (1984) Quality of response in different population groups in mail and telephone surveys. *Am J Epidemiol* 120:302–314
12. Ware JE, Snow KK, Kosinski M, Gandek B (2000) SF-36 Health Survey: Manual and Interpretation Guide. Quality Metric Incorporated, Lincoln (RI)
13. Bullinger M, Kirchberger I (1998) Der SF-36-Fragebogen zum Gesundheitszustand: Handbuch für die deutschsprachige Fragebogenversion. Hogrefe-Verlag für Psychologie, Göttingen
14. Ware JE Jr, Kosinski M, Dewey JE, Gandek B (1999) How to score and interpret single-item health status measures: a manual for users of the SF-8™ Health Survey. Quality Metric Incorporated, Lincoln (RI)
15. Ellert U, Lampert T, Ravens-Sieberer U (2005) Messung der gesundheitsbezogenen Lebensqualität mit dem SF-8 – eine Normstichprobe für Deutschland. *Bundesgesundheitsblatt* 48
16. Aoki FY, Fleming DM, Griffin AD et al. (2000) Impact of Zanamivir treatment on productivity, health status and healthcare resource use in patients with influenza. *Pharmacoeconomics* 2:187–195
17. Bayliss MS, Dewey JE, Dunlap I et al. (2003) A study of the feasibility of Internet administration of a computerized health survey: the headache impact test (HIT). *Qual Life Res* 12:953–961
18. Ramos-Remus CR, Hunsche E, Mavros P et al. (2004) Evaluation of quality of life following treatment with etoricoxib in patients with arthritis or low-back pain: an open label, uncontrolled pilot study in Mexico. *Curr Med Res Opin* 20:691–698
19. Turner-Bowker DM, Bayliss MS, Ware JE Jr, Kosinski M (2003) Usefulness of the SF-8 Health Survey for comparing the impact of migraine and other conditions. *Qual Life Res* 12:1003–1012
20. Ware JE Jr, Kosinski M, Bjorner JB et al. (2003) Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res* 12:935–952
21. Cohen J (1988) *Statistical power analysis for the behavioural sciences*. Erlbaum, New York
22. Embretson SE, Reise SP (2000) *Item response theory for psychologists*. Lawrence Erlbaum Associates, Mahwah

**Th. Eikmann, U. Heinrich, B. Heinzow, R. Konietzka (Hrsg.)
Gefährdungsabschätzung von Umweltschadstoffen (GvU)**

Ergänzbare Handbuch toxikologischer Basisdaten und ihrer Bewertung. Loseblattwerk in zwei Spezialordnern. 9. und 10. Ergänzungslieferung August 2004 und Februar 2005; Gesamtwerk 1992 S., 98 €. Erich Schmidt Verlag Berlin Bielefeld München. ISBN 3-503-05083-3

Das Handbuch dient als Nachschlagewerk für humantoxikologische Bewertungsmaßstäbe und ihrer Begründungen vorwiegend für die Anwendung des Bundesbodenschutzgesetzes. Es werden Grundsatzfragen der Risikobewertung erörtert sowie Methoden zur Ableitung gesundheitlich tolerierbarer Schadstoffzufuhren entwickelt. Die Bewertungen der Einzelsubstanzen beruhen zum einen auf den zum Bundesbodenschutzgesetz erarbeiteten „Basisdaten Toxikologie“. Erweitert werden die Abschätzungen durch Empfehlungen der „Kommission Human-Biomonitoring“ sowie der „Kommission Innenraumluftthygiene“.

In den Ergänzungslieferungen werden die Ergebnisse neuer Risikobewertungen und von Neubewertungen vorliegender Monographien veröffentlicht. Es werden erstmals die „tolerierbaren resorbierten Dosen (TRD-Werte)“ zu Chloroform, Chlorbenzol, Benzol, Ethylbenzol, Vinylchlorid und 1,2,4-Trichlorbenzol vorgelegt. Außerdem werden Neubewertungen von Tetrachlorethen (PER, Kennziffer D 892), Toluol (D 916), Nickel (D 714) und Beryllium (D 127) geboten.

Die Kommission „Human-Biomonitoring“ erarbeitete eine Bewertung bzw. Neubewertung von Nickel, Cadmium (D173), und Quecksilber (831). Aus der Arbeit der „Kommission Innenraumluftthygiene“ sind die Bewertungen von Kohlenmonoxid und von Diisocyanaten zu finden (dazu auch Hinweise zum nicht-beruflichen Umgang mit diisocyanathaltigen Produkten [Anstriche, Klebstoffe, Ortschaftäme]).

Gerald Milde, Berlin