

## Eine kliniktaugliche Version des deutschsprachigen SF-36 und ihr psychometrischer Vergleich mit dem Originalfragebogen

### Summary

#### A hospital suited version of the German SF-36 and its psychometric comparison with the original questionnaire

**Objectives:** The application of the Short-Form 36 Health Survey (SF-36) in hospitals is sometimes problematic, because some formulations like "at work" or "at home" in some items are not adequate in a hospital context. To account for this problem, we deleted 23 words like "at work" and "at home" from the SF-36 and generated thus a slightly modified version, the SF-36m. The present paper deals primarily with two questions: is the SF-36m well applicable in hospitals and is it still equivalent with the original version?

**Methods:** The SF-36m was applied in more than 300 patients in two different hospitals. In one hospital a randomised design was used to assess the equivalence of SF-36 and SF-36m before treatment.

**Results:** The reported high percentages of missing values in four critical scales were no more observed, when the SF-36m was applied at the end of hospital stays. Only minor differences between SF-36 and SF-36m with effect sizes  $< 0,20$  were observed, although the scale "physical role function" and especially item 4d showed unexpectedly significantly higher values in the SF-36m ( $p < 0,10$ ). Additionally, higher and significant differences were observed between the applications of the questionnaires at home and in the hospital.

**Conclusions:** Opposed to the original SF-36 the SF-36m is well applicable at the end of hospital stays. SF-36 and SF-36m show only minor differences in the comparison of their equivalence. The application of SF-36 and SF-36m in hospitals seems to yield more favourable scores of health related quality of life than the application at home.

**Key-Words:** SF-36 – Quality of life – Assessment – Rehabilitation.

Die Erfassung der gesundheitsbezogenen Lebensqualität ist zu einem zentralen methodischen Thema der medizinischen Outcome-Forschung geworden. Dieses Interesse begründet sich in dem Wunsch bzw. dem Erfordernis, neben physiologischen und klinischen Parametern auch die Patientenperspektive und die Alltagskonsequenzen von Therapien zu evaluieren. Neben zahlreichen anderen Entwicklungen hat der MOS 36-Item Short-Form Health Survey (SF-36) in jüngerer Zeit bisher wohl die meiste Beachtung und Verbreitung gefunden<sup>1,2</sup>. Ähnlich bekannte generische Messinstrumente aus dem angloamerikanischen Sprachraum sind das Nottingham Health Profile<sup>3,4</sup>, die Quality of Wellbeing Scale<sup>5</sup> und das Sickness Impact Profile<sup>6</sup>. Der SF-36 ist nicht nur sehr aufwendig entwickelt worden, sondern liegt auch in zahlreichen verschiedenen Sprachen vor und ist psychometrisch ausgiebig und mit weitgehend zufriedenstellenden Resultaten untersucht worden<sup>7-11</sup>. Die Beschränkung auf 36 Items, die in der deutschen Version auf gerade drei Seiten untergebracht sind, erlaubt eine sehr ökonomische – wenn auch nicht immer ganz präzise – Einschätzung von acht Dimensionen der Lebensqualität, die sich an der WHO-Definition von Gesundheit orientieren: Körperliche Funktionsfähigkeit, körperliche Rollenfunktion, körperlicher Schmerz, allgemeine Gesundheitswahrnehmung, Vitalität, soziale Funktionsfähigkeit, emotionale Rollenfunktion und psychisches Wohlbefinden<sup>12</sup>.

Da die Entwicklung und psychometrische Testung des SF-36 primär auf epidemiologischen Erhebungen in der relativ gesunden Allgemeinbevölkerung basierte, ist sein Einsatz bei verschiedenen Erkrankungen und vor allem am

Ende eines Klinikaufenthalts nicht unproblematisch. Einige Fragen sind so eindeutig und konkret auf den Alltagskontext bezogen, dass sie am Ende eines Klinikaufenthalts so nicht sinnvoll gestellt werden können. Tut man dies dennoch, so reagieren viele Patienten verständlicherweise ablehnend auf den Fragebogen und es resultieren im besten Fall etliche fehlende Werte, die für die weitere Auswertung problematisch sind. In einer Untersuchung von Zwingmann et al. bezifferte sich der Anteil fehlender Werte am Ende eines Aufenthalts in einer Klinik bei den problematischen vier Skalen auf bis zu 25 %<sup>10</sup>.

Als Lösungen für dieses Problem wurden bisher diskutiert, den SF-36 erst wieder im Alltagskontext mit zeitlichem Abstand vom Klinikaufenthalt einzusetzen<sup>10</sup> oder aber die Probanden auf die problematischen Fragen vorzubereiten und sie zu instruieren, den in den betreffenden Fragen spezifizierten Kontext zu ignorieren, und die Items selbst so gut wie eben möglich zu beantworten<sup>12</sup>. Die erste Lösung ist insofern unbefriedigend, als damit in vielen Untersuchungen auf einen sehr wichtigen Messzeitpunkt verzichtet werden müsste: Der kurzfristige Therapie- bzw. Rehabilitationserfolg würde schlicht nicht im SF-36 abgebildet. Die zweite Lösung erhöht den organisatorischen Untersuchungsaufwand erheblich, weil alle Patienten aufwendiger instruiert werden müssten, und fordert auch eine kognitive Flexibilität auf Seiten der Befragten, die wohl nicht immer vorausgesetzt werden kann.

Eine nähere Betrachtung des SF-36 führte uns zu der Einschätzung, dass die problematischen Items selbst durchaus auch am Ende eines Klinikaufenthalts sinnvoll zu beantworten sind, während vor allem die in einigen einleitenden Formulierungen erfolgende konkrete Spezifizierung des Kontexts bei der Beantwortung Schwierigkeiten bereitet: Durch Formulierungen wie „bei der Arbeit“ und „zu Hause“ werden den Patienten im Klinikkontext ganz offensichtlich unpassende Fragen gestellt (vgl. Tab. 1). Eine Inspektion der problematischen Fragen zeigt ebenso deutlich, dass diese Konkretisierungen durchaus aus den Frageformulierungen entfernt werden können, ohne den Sinn der Items gravierend zu verändern. Die in Tabelle 1 unterstrichenen 23 Wörter wurden deshalb von uns für eine modifizierte Version des SF-36 entfernt. Damit wurde eine minimal veränderte Variante des SF-36 erzeugt, die nach unserer Einschätzung auch am Ende eines Klinikaufenthalts relativ problemlos einsetzbar sein sollte.

Im Folgenden berichten wir über den Einsatz dieser modifizierten Version erstens unter dem Aspekt ihrer Kliniktauglichkeit, die im Rahmen von Qualitätssicherungsmaßnahmen in zwei verschiedenen Kliniken untersucht wurde, und zweitens unter dem Aspekt ihrer psychometrischen Ver-

4. Hatten Sie in der vergangenen Woche aufgrund Ihrer körperlichen Gesundheit irgendwelche Schwierigkeiten bei der Arbeit oder anderen alltäglichen Tätigkeiten im Beruf bzw. zu Hause?
  - a) Ich konnte nicht so lange wie üblich tätig sein
  - b) Ich habe weniger geschafft als ich wollte
  - c) Ich konnte nur bestimmte Dinge tun
  - d) Ich hatte Schwierigkeiten bei der Ausführung
  
5. Hatten Sie in der vergangenen Woche aufgrund seelischer Probleme irgendwelche Schwierigkeiten bei der Arbeit oder anderen alltäglichen Tätigkeiten im Beruf bzw. zu Hause (z. B. weil Sie sich niedergeschlagen oder ängstlich fühlten)?
  - a) Ich konnte nicht so lange wie üblich tätig sein
  - b) Ich habe weniger geschafft als ich wollte
  - c) Ich konnte nicht so sorgfältig wie üblich arbeiten
  
8. Inwieweit haben die Schmerzen Sie in der vergangenen Woche bei der Ausübung Ihrer Alltagstätigkeiten zu Hause und im Beruf behindert?

**Tabelle 1** Die problematischen und in der modifizierten Version des SF-36 geänderten Items. Angegeben sind die Fragen der Originalversion mit einem Zeitfenster von einer Woche, wobei die unterstrichenen Wörter für besonders kritisch erachtet und in der modifizierten Version entfernt wurden (Hervorhebungen in Fett- und Kursiv-Schrift entsprechen dem Original)

gleichbarkeit mit dem Originalfragebogen, die in einer der erwähnten Kliniken speziell untersucht wurde.

## Methoden

Bei der psychometrischen Untersuchung der modifizierten Version des SF-36 – im Folgenden kurz SF-36m genannt – gilt es, zwei grundsätzlich verschiedene Fragen zu beantworten. Erstens stellt sich die Frage, ob die beschriebenen minimalen Änderungen ausreichen, um den SF-36m am Ende eines Klinikaufenthalts einsetzen zu können, ohne dass es zu den bekannten für die Originalversion beschriebenen Problemen kommt. Zweitens ist zu klären, inwiefern der SF-36m nach den vorgenommenen Änderungen noch äquivalent zum SF-36 ist. Die dritte Frage, die ausserdem noch untersucht werden soll, ist generellerer Art: Wird der SF-36 zuhause anders ausgefüllt als in der Klinik?

### *Beurteilung der Kliniktauglichkeit des SF-36m*

Neben allgemeinen Erfahrungen beim Einsatz des SF-36m, wie z.B. Probandenäußerungen zum Fragebogen, ist die Anzahl fehlender Werte bei den vier Skalen, die Items der zweiten Seite des Fragebogens enthalten, ein geeigneter Indikator, um zu beurteilen, ob diese Items von den Probanden in der veränderten Form akzeptiert werden. Dieser Indikator wurde bereits von Zwingmann et al. bei der Analyse des SF-36 gewählt<sup>10</sup>, so dass Ergebnisse für den SF-36m mit den von Zwingmann et al. berichteten Werten verglichen wurden. Weiter wurde die Anzahl fehlender Werte, die am Ende eines Klinikaufenthalts beim SF-36m beobachtet wurden, mit der Anzahl fehlender Werte beim

Original-Fragebogen verglichen, der zu Beginn eines Klinikaufenthalts appliziert wurde. In zwei Kliniken wurde der SF-36m im Rahmen von Qualitätssicherungsmaßnahmen jeweils am Ende eines Klinikaufenthalts eingesetzt und der Originalbogen des SF-36 zu Beginn bzw. vor dem Klinikaufenthalt. Bei der einen Klinik handelte es sich um die Malteser Klinik Dr. v. Weckbecker in Bad Brückenau, eine naturheilkundlich orientierte Klinik mit einem breiten Diagnosespektrum. Hier wurden konsekutiv insgesamt 268 Patienten mit dem SF-36 befragt, die überwiegend Erkrankungen des Muskel-Skelett-Systems, Erkrankungen des Kreislaufsystems, endokrine, Ernährungs- und Stoffwechselerkrankungen oder psychische und Verhaltensstörungen aufwiesen. Aus der Vogtland-Klinik, Bad Elster, lagen 104 ausgefüllte SF-36-Bogen vom Beginn einer Reha-Massnahme vor und 71 bearbeitete Bogen des SF-36m vom Ende der Massnahme. Auch hier handelte es sich um eine konsekutive und nicht selektierte Stichprobe, wobei die Haupt-Indikationsgebiete der Klinik Orthopädie und Kardiologie sind und in zahlenmässig geringerem Umfang Gynäkologie und Nephrologie. Da in dieser Arbeit weniger die Patientencharakteristika von Interesse sind und mehr die Eigenschaften des SF-36 bzw. SF-36m mag diese grobe Charakterisierung des Patientenguts genügen.

#### Beurteilung der Äquivalenz des SF-36m

In dem bereits erwähnten Qualitätssicherungsprojekt in der Vogtland-Klinik, Bad Elster, wurde ein spezielles Design gewählt, um die Äquivalenz von SF-36m und SF-36 beurteilen zu können. Abbildung 1 zeigt dieses Design. Die Probanden erhielten zuhause vor Beginn ihrer Rehabilitationsmassnah-

me entweder den SF-36 oder den SF-36m zugesandt. Die Entscheidung, welcher der Bogen zunächst verwendet wurde, erfolgte randomisiert. Bei Beginn des Klinikaufenthalts wurde dann der jeweils andere Bogen eingesetzt, so dass der SF-36 und der SF-36m vor Beginn der Reha in zufälliger Reihenfolge zuhause und in der Klinik ausgefüllt vorlagen. Unsere Hypothese war, dass sich die SF-36-Werte für den Re-Test zwischen  $t_0$  und  $t_1$  nicht unterscheiden sollten, was umso plausibler erschien, als es sich um ein hoch chronifiziertes Patientengut handelte und der zeitliche Abstand zwischen  $t_0$  und  $t_1$  im Mittel lediglich ca. 13 Tage betrug.

Der Vergleich von SF-36m und SF-36 erfolgte auf drei unterschiedlichen Ebenen, die mit *Itemebene*, *Skalenebene* und *Fragebogenebene* bezeichnet werden können. Auf der *Itemebene* wurde untersucht, ob sich die Antwortverteilungen für die acht Items, bei denen die einleitenden Formulierungen modifiziert worden waren, zwischen SF-36 und SF-36m signifikant unterschieden. Dazu wurde ein eindimensionaler  $\chi^2$ -Test benutzt, bei dem geprüft wurde, ob der SF-36m bei den acht Items jeweils die Antwortverteilung, die beim SF-36 empirisch beobachtet wurde, repliziert. Da eine Äquivalenzfragestellung vorliegt, wurde für alle statistischen Tests ein Alpha-Fehler-Risiko von 0,10 als Kriterium gewählt, unterhalb dessen keine Äquivalenz mehr zwischen SF-36 und SF-36m angenommen werden sollte. Dieses Vorgehen sicherte bei dem für eine Äquivalenzprüfung nicht allzu grossen N eine noch ausreichende statistische Power für die Entdeckung eventueller Unterschiede. Auf *Skalenebene* wurden die Mittelwerte für die drei Skalen, die modifizierte Items enthielten („Körperliche Rollenfunktion“, „Schmerz“

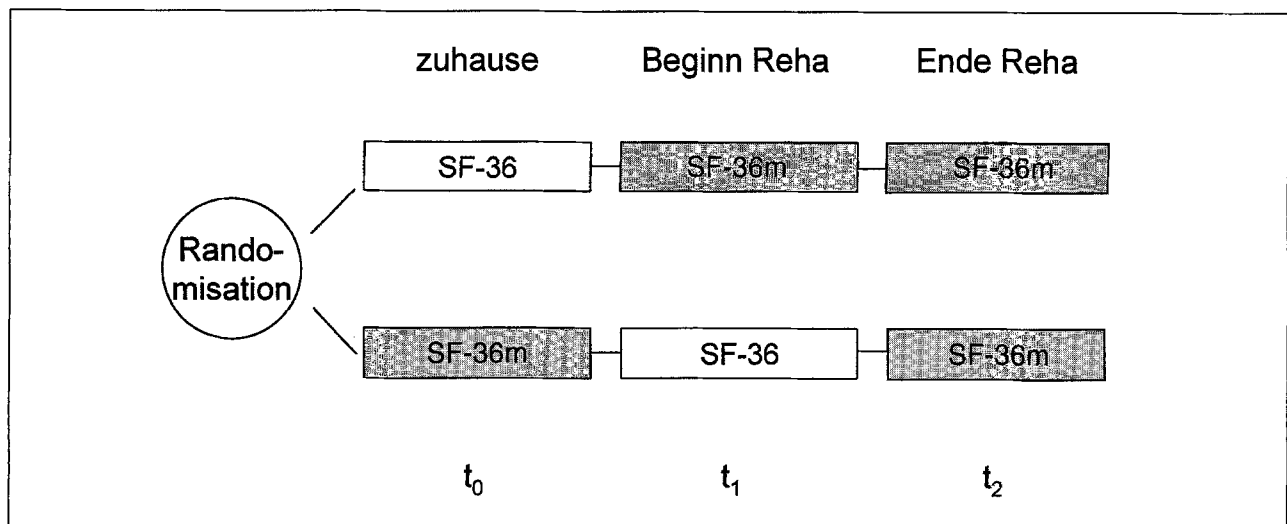


Abbildung 1 Das in der Vogtland-Klinik eingesetzte Design zur Beurteilung der Äquivalenz von SF-36 und SF-36m

Skala	SF-36 (Original)				SF-36m (modifizierte Version)			
	GFI	AGFI	Relia	$\alpha$	GFI	AGFI	Relia	$\alpha$
Körperliche Rollenfunktion	0,95	0,93	0,72	0,89	0,90	0,88	0,69	0,89
Emotionale Rollenfunktion	0,98	0,97	0,84	0,93	1,00	1,00	0,83	0,95
Körperlicher Schmerz	0,96	0,88	0,80	0,88	1,00	1,00	0,82	0,89

GFI = goodness of fit index (0–1), AGFI = adjusted goodness of fit index (0–1), Relia = itemspezifische Reliabilität (0–1),  $\alpha$  = Cronbachs  $\alpha$  (0–1, Interne Konsistenz).

**Tabelle 2** LISREL-Analysen zur Skalenhomogenität und Reliabilitätsschätzung bei den modifizierten Skalen des SF-36 und SF-36m – unter Annahme eines tau-parallelen Messmodells für  $t_0/t_1$

und „Emotionale Rollenfunktion“), mittels t-Test für abhängige Stichproben miteinander verglichen. Zur Deskription der Unterschiede wurden für alle acht Skalen Effektstärken für den Unterschied zwischen SF-36 und SF-36m berechnet, wobei die Mittelwertsunterschiede an den Streuungen der Werte des SF-36 standardisiert wurden.

Zusätzlich erfolgten zu den drei modifizierten Skalen des SF-36 Analysen mit Strukturgleichungsmodellen unter Annahme jeweils eines tau-parallelen Messmodells<sup>13</sup>. Bei tau-parallelen Messmodellen erfolgt die sehr weit gehende Annahme, dass die analysierten Items gleiche True-Score-Varianzen und gleiche Residual-Varianzen besitzen. Für diese Modelle wurden der jeweilige „goodness of fit index“ (GFI), der „adjusted goodness of fit index“ (AGFI) und die jeweilige itemspezifische Reliabilität berechnet. Als vielleicht besser bekannter Reliabilitätsindex wurde die interne Konsistenz (Cronbachs Alpha) bestimmt und angegeben. Auf Fragebogen-Ebene wurde zum Vergleich der beiden Gesamt-Fragebogen eine multivariate Varianzanalyse (MANOVA) mit den acht Skalen als abhängigen Variablen und Messwiederholung auf dem Faktor „Fragebogenversion“ berechnet.

Zusätzlich wurde durch Vergleich der Werte für  $t_0$  und  $t_1$  untersucht, ob das Ausfüllen des SF-36 zuhause zu den gleichen Ergebnissen führte wie die Beantwortung in der Klinik. Hier gingen also in die MANOVA der acht Skalen die Werte von SF-36 und SF-36m „gemischt“ ein und der Messwiederholungsfaktor „Kontext“ (zuhause versus Klinik) bestand aus den Werten von  $t_0$  versus  $t_1$ . Die Berechnung von Effektstärken, den standardisierten Mittelwertsunterschieden zwischen  $t_0$  und  $t_1$ , erfolgte hier durch Relativierung an den Streuungen zu  $t_0$ .

Wenngleich multipel getestet wurde, erfolgte keine Adjustierung des Alpha-Fehler-Niveaus, um die statistische Power für die Entdeckung eventueller Unterschiede zwischen den Fragebogen (Äquivalenzfragestellung!) nicht unnötig zu reduzieren. Alle statistischen Analysen erfolgten mit SPSS 6.1.3 für Windows (SPSS, Chicako Inc.) bzw. mit dem damit vertriebenen LISREL 7.20.

Für eine erste methodische Information zum SF-36m sind in Tabelle 2 die Ergebnisse der LISREL-Analysen zu den drei modifizierten Skalen in Original- und modifizierter Version dargestellt. Die geschätzten Parameter zeigen überwiegend für den SF-36m genauso zufriedenstellende Fit-Indizes und Reliabilitätskoeffizienten wie für den SF-36. Lediglich der AGFI von 0,88 für die Skala „Körperliche Rollenfunktion“ in der modifizierten Version bleibt etwas unterhalb des kritischen Wertes von 0,90. Dabei lieferten liberalere, d.h. ein kongenerisches und ein tau-äquivalentes Messmodell, für diese Skala noch schlechtere Fit-Indizes, was auf Probleme bei der Skala „Körperliche Rollenfunktion“ in der modifizierten Version hinweisen könnte.

## Ergebnisse

### Ergebnisse zur Akzeptanz des SF-36m

Beim Einsatz des SF-36m gab es im Gegensatz zur Originalversion auch am Ende eines Klinikaufenthalts keine nennenswerten Probleme. Tabelle 3 zeigt die Daten, die eine sehr weit gehende Akzeptanz des SF-36m belegen. Tabelle 3 zeigt die Prozentsätze fehlender Werte für die vier kritischen Skalen des SF-36, die bei der Originalversion

Skala	Maltser Klinik Dr. v. Weckbecker		Vogtland-Klinik		Ende SF-36 Zwingm. * (N = 244)
	Beginn SF-36 (N = 268)	Ende SF-36m (N = 256)	Beginn SF-36 (N = 104)	Ende SF-36m (N = 71)	
Körperliche Rollenfunktion	8,3%	8,5%	4,8%	3,2%	25,5%
Körperliche Schmerzen	2,1%	3,1%	4,3%	0,7%	21,5%
Soziale Funktionsfähigkeit	1,7%	4,3%	3,8%	2,1%	18,4%
Emotionale Rollenfunktion	6,1%	7,7%	5,8%	1,4%	23,7%

\* Zwingmann et al. (1998).

**Tabelle 3** Die Akzeptanz der modifizierten Version am Ende eines Klinikaufenthalts: Prozentsätze an fehlenden Werten für die vier kritischen Skalen des SF-36 im Vergleich mit Werten für den SF-36m

des SF-36 am Ende eines Klinikaufenthalts von vielen Probanden nicht beantwortet werden. Dies ist in der rechten Spalte von Tabelle 3 noch einmal anhand der von Zwingmann et al. beobachteten hohen Prozentsätze (>18%) fehlender Werte bei diesen Skalen dargestellt<sup>10</sup>. Die Spalten drei und fünf der Tabelle („Ende SF-36m“) zeigen zunächst, dass die mit dem SF-36m beobachteten Prozentsätze fehlender Werte deutlich unter 10% bleiben und damit wesentlich niedriger sind als die von Zwingmann et al. für die Originalversion berichteten Werte. Zum weiteren Vergleich ist der jeweilige Anteil an fehlenden Werten für den SF-36 in seiner Originalversion zu bzw. vor Beginn des Klinikaufenthalts in den Spalten zwei und vier angegeben. Hier ist zu sehen, dass sich die Werte in den Spalten zwei und drei bzw. vier und fünf – und damit der Anteil der gültigen Antworten zu Beginn und am Ende der Massnahmen in jeder der beiden Kliniken – nicht nennenswert unterscheiden. Kritisch anzumerken ist, dass die Patientenzahlen in der Vogtlandklinik Bad Elster (Spalten vier und fünf) relativ gering waren und die berechneten Prozentsätze fehlender Werte dementsprechend grobe Schätzungen darstellen. Demzufolge sollte die vielleicht auffallende Tatsache, dass die Prozentsätze fehlender Werte in der Vogtland-Klinik für den SF-36m sogar etwas geringer waren als für den SF-36 unseres Erachtens nicht überbewertet werden. Zusammenfassend ist festzuhalten, dass der SF-36m am Ende eines Klinikaufenthalts bei über 300 Patienten problemlos einsetzbar war.

#### Ergebnisse zur Äquivalenz von SF-36m und SF-36

Zunächst wird dargestellt, ob und inwiefern sich das Antwortverhalten bei den acht modifizierten Items zwischen Original- und modifizierter Version unterscheidet (Tab. 4).

Tabelle 4 zeigt, dass die Antwortverteilungen der modifizierten Items in SF-36 und SF-36m bei den meisten Items sehr ähnlich sind. Allerdings zeigt sich bei Item 4d „Ich hatte Schwierigkeiten bei der Ausführung“ ein mit einem p-Wert von 0,03 signifikanter Unterschied zwischen Original- und modifizierter Version der Art, dass die Frage beim SF-36 mit 70% häufiger mit „ja“ beantwortet wird als beim SF-36m mit 61%.

In Abbildung 2 sind die Mittelwerte der acht Skalen des SF-36 und des SF-36m wiedergegeben zusammen mit den Normwerten der altersentsprechenden Normstichprobe. Die Mittelwerte aller acht Skalen des SF-36 unterscheiden sich kaum zwischen SF-36 und SF-36m: Die Effektstärken für die Differenzen zwischen Original- und modifizierter Version sind durchweg kleiner als 0,2, also sehr gering, die meisten Effektstärken sogar = 0,1. Allerdings ist für die Skala

Item	Antwort	SF-36 %	SF-36m %	P*
4a ... nicht so lange wie üblich tätig	Ja	72	67	0,34
	Nein	28	33	
4b ... weniger geschafft ...	Ja	74	73	0,82
	Nein	26	27	
4c ... nur bestimmte Dinge ...	Ja	72	65	0,12
	Nein	28	35	
4d ... Schwierigkeiten bei Ausführung	Ja	70	61	0,03
	Nein	30	39	
5a ... nicht so lange wie üblich tätig	Ja	32	30	0,78
	Nein	68	70	
5b ... weniger geschafft ...	Ja	37	31	0,23
	Nein	63	69	
5c ... nicht so sorgfältig wie üblich	Ja	35	28	0,18
	Nein	65	72	
8 Inwieweit haben die Schmerzen behindert	Überhaupt nicht	10	10	0,76
	Ein bisschen	24	29	
	Mässig	26	23	
	Ziemlich	23	24	
	Sehr	16	13	

\* Mit eindimensionalem Chi<sup>2</sup>-Test geschätzte Wahrscheinlichkeit dafür, dass die Antwortverteilungen für SF-36 und SF-36m gleich sind.

**Tabelle 4** Relative Häufigkeiten für die Antwortkategorien bei den acht modifizierten Items des SF-36 und SF-36m vor bzw. bei Beginn des Klinikaufenthalts (N=98-108)

„Körperliche Rollenfunktion“ der Unterschied zwischen SF-36 und SF-36m signifikant ( $p=0,03$ , t-Test), wobei der SF-36m im Mittel um ca. sieben Punkte günstigere Werte lieferte als der SF-36. Auch die multivariate Varianzanalyse der acht Skalen mit Messwiederholung zeigt einen signifikanten Unterschied zwischen beiden Versionen des SF-36 an ( $p=0,04$ ).

#### Ergebnisse zur Äquivalenz von Alltags- und Klinikkontext

Abbildung 3 zeigt ähnlich zu Abbildung 2 die acht Skalen des SF-36, wobei diesmal nicht Original- und modifizierte Version gegenübergestellt sind, sondern  $t_0$  und  $t_1$ : also einmal die Bearbeitung von SF-36 oder SF-36m zuhause und zum anderen das Ausfüllen der Bogen zu Beginn des Klinikaufenthalts.

Abbildung 3 und eine entsprechende MANOVA zeigen einen kleinen, aber konsistenten und signifikanten Unterschied ( $p<0,01$ ) zwischen den SF-36-Werten, die bei Beantwortung der Bogen zuhause resultieren, und denjenigen Werten, die bei Bearbeitung in der Klinik zustande kommen: Zuhause werden bei fast allen Skalen leicht negativere Werte produziert als in der Klinik, wobei die Effektstärken immerhin Werte von bis zu 0,28 erreichen.

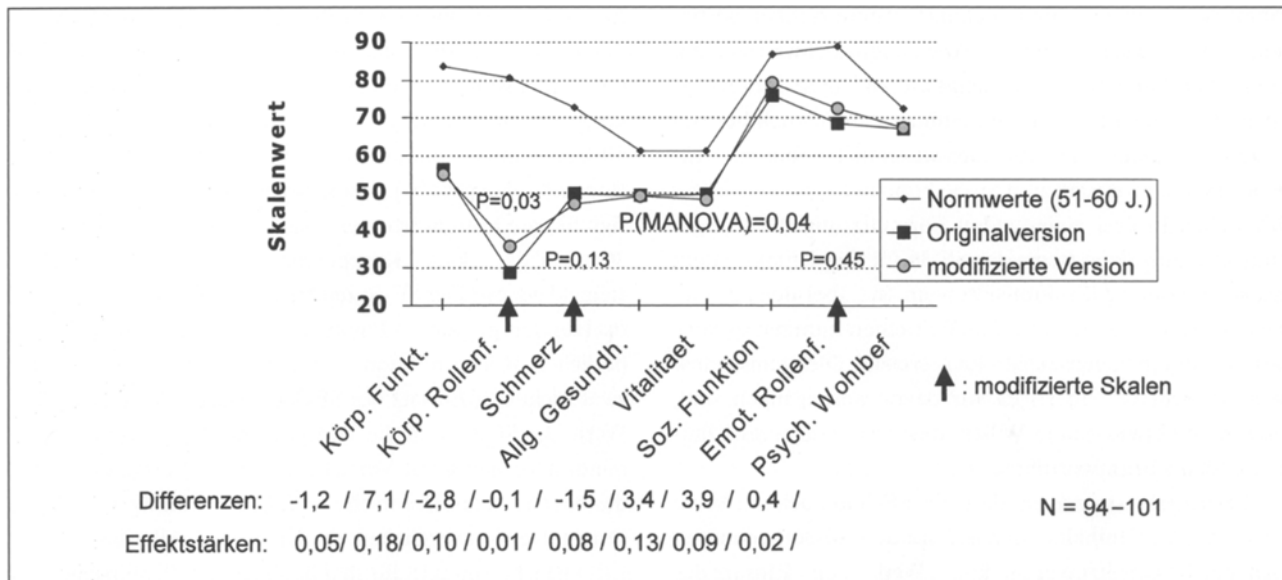


Abbildung 2 Altersentsprechende Normwerte und Skalenmittelwerte für den SF-36 und den SF-36m

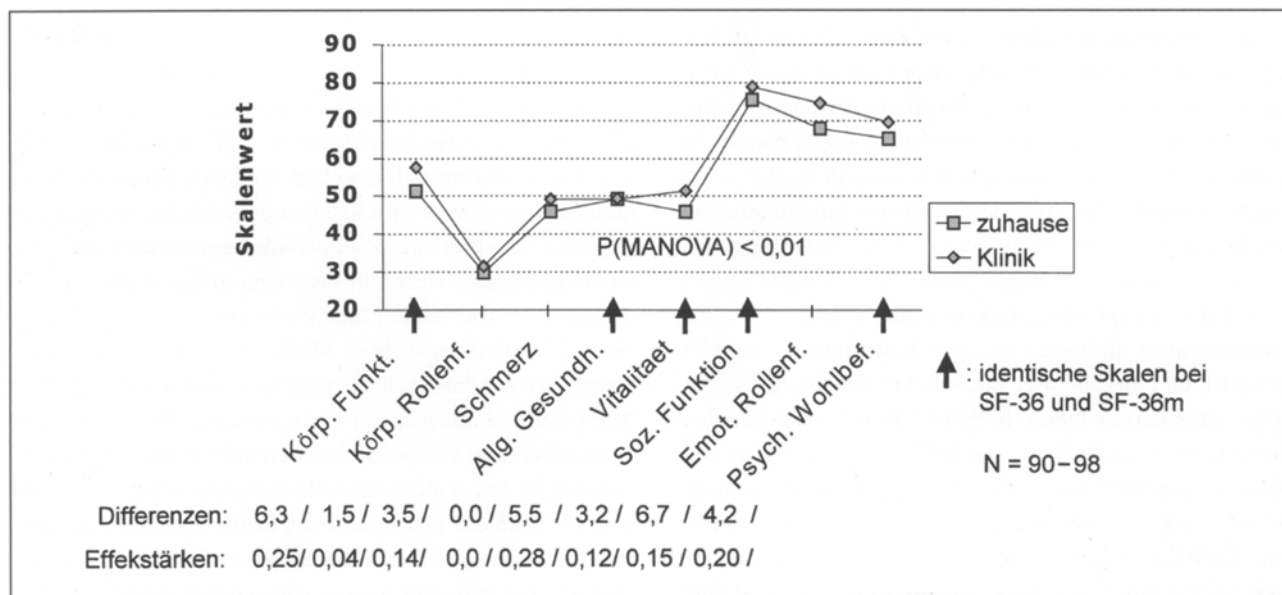


Abbildung 3 Zuhause ( $t_0$ ) und in der Klinik ( $t_1$ ) beobachtete Skalenmittelwerte für den SF-36 bzw. den SF-36m

## Diskussion

Der Einsatz des SF-36 am Ende eines Klinikaufenthalts ist aufgrund der sprachlichen Einbettung einiger Items problematisch, was bei der Durchführung einer Studie ebensolche Schwierigkeiten bereitet wie bei der Auswertung der Ergebnisse. Bei der Anwendung des Fragebogens reagieren etliche Patienten berechtigterweise mit Unmut, wenn ihnen of-

fensichtlich unpassende Fragen gestellt werden. Bei der Auswertung kommt es zu vielen fehlenden Werten bei vier Skalen, wie von Zwingmann et al. dokumentiert wurde, die Prozentsätze an fehlenden Werten um die 20% berichteten (vgl. Tab. 3 und<sup>10</sup>). Dieses Problem ist unter Anwendern und potentiellen Anwendern des SF-36 in Kliniken wohl bekannt, wie auf verschiedenen Kongressen immer wieder deutlich wird, ohne dass unseres Wissens neben der

Publikation von Zwingmann et al.<sup>10</sup> weitere Artikel existieren, in denen das Ausmass der Anwendungsschwierigkeiten konkret dokumentiert und nachgewiesen wurde. Allerdings fallen die fehlenden Werte im ambulanten Bereich bei Befragungen zuhause zum Teil auch wesentlich höher aus<sup>11</sup> als in der Handanweisung zum SF-36 berichtet.

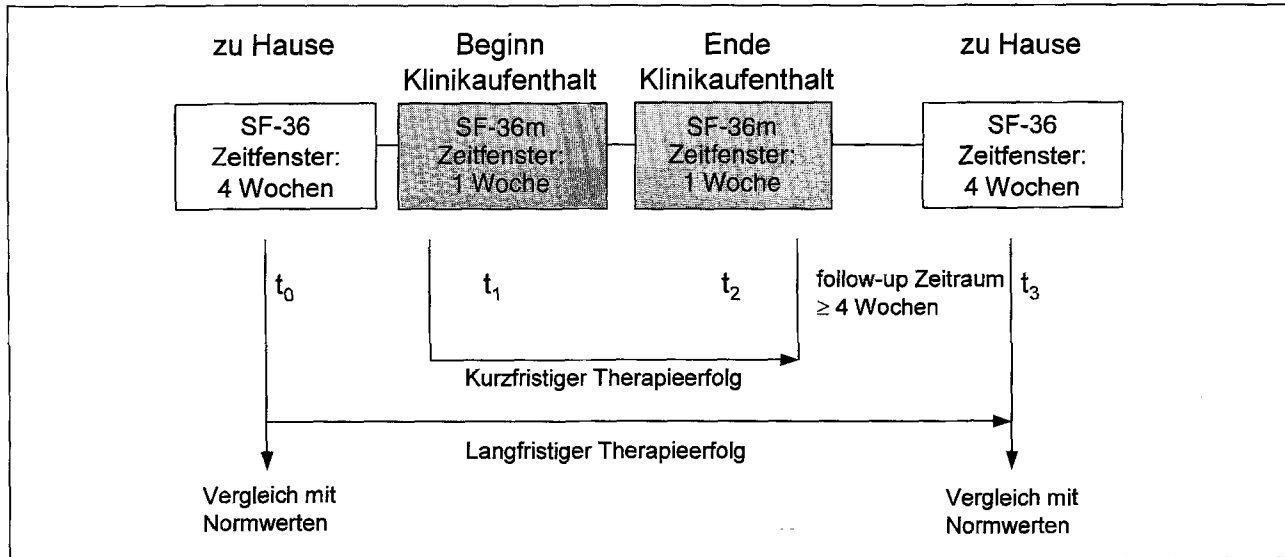
Wir haben in drei einleitenden Formulierungen zu insgesamt acht einzelnen Items des SF-36 für den Klinikkontext kritische Kontext-Konkretisierungen wie „bei der Arbeit“ und „zu Hause“ entfernt und damit einen minimal modifizierten Fragebogen, den SF-36m, erstellt. Die Items selbst wurden nicht verändert, da wir davon ausgegangen sind, dass sie mit etwas gutem Willen auch am Ende eines Klinikaufenthalts beantwortbar sind.

Unsere Ergebnisse zeigen, dass der SF-36m auch am Ende eines Klinikaufenthalts ohne die für den SF-36 bekannten Probleme eingesetzt werden kann. Weder beim Einsatz des Bogens noch bei der Auswertung treten nennenswerte Schwierigkeiten auf. Hingewiesen sei noch darauf, dass die Anzahl fehlender Werte bei den Fragebogen nicht nur vom Fragebogen selbst sondern sicher auch von der Art der Instruktion abhängig ist. Die positiven Ergebnisse zur Akzeptanz des SF-36m sind allerdings dahingehend etwas einzuschränken, dass die gewählte Vorgehensweise rein methodisch gesehen suboptimal ist (vgl. Tab. 3). Die historische Kontrolle durch die Arbeit von Zwingmann et al.<sup>10</sup> ist dadurch begrenzt, dass die betrachteten Patientengruppen deutlich verschieden voneinander sind. Unsere Stichproben enthielten vermutlich einen wesentlich höheren Anteil an schwerwiegend chronifizierten und älteren Patienten. Zwingmann et al. hatten dagegen Patienten, die zu 74% männlich und im Durchschnitt 44 Jahre alt waren. Die Autoren analysierten lediglich Patienten mit unspezifischen chronischen Rückenschmerzen, bei denen es keine Hinweise auf entzündliche Genese oder akuten Bandscheibenvorfall gab. Im idealen methodischen Design hätten wir zur Prüfung der Akzeptanz des SF-36m am Ende des Klinikaufenthalts randomisiert entweder den SF-36 oder den SF-36m eingesetzt. Die hier berichteten Ergebnisse wurden jedoch im Rahmen von Qualitätssicherungsprojekten gewonnen, bei denen wir es unseren Partnern in den Kliniken nicht hätten zumuten können, den am Ende eines Klinikaufenthalts bekanntermassen problematischen SF-36 aus „rein akademisch-methodischem Interesse“ auch mit einzusetzen, wo wir doch mit dem SF-36m eine vermutlich kliniktauglichere Version zur Verfügung hatten. So bleiben weitere Erfahrungen mit dem SF-36m in anderen Stichproben abzuwarten, ehe definitiv davon ausgegangen werden kann, dass er die beschriebenen Probleme auch unter anderen als den in unserer Studie realisierten Umständen beseitigt.

Etwas weniger positiv sind die Ergebnisse hinsichtlich der Äquivalenz von SF-36 und SF-36m. Item 4d, Schwierigkeiten bei der Ausführung alltäglicher Tätigkeiten aufgrund der körperlichen Gesundheit, wurde in der Originalversion signifikant häufiger mit ja beantwortet als beim modifizierten Bogen (70% vs. 61%). Dementsprechend fanden sich auch signifikante Mittelwertunterschiede zwischen SF-36 und SF-36m bei der Skala „Körperliche Rollenfunktion“, zu der Item 4d gehört: Der SF-36 zeichnete auf einer Skala von Null bis Hundert ein um 7,1 Punkte ungünstigeres Bild der „Körperlichen Rollenfunktion“ als der SF-36m. Dies verbietet zwar nicht den Einsatz des SF-36m, führt jedoch dazu, dass Werte des SF-36m für die „Körperliche Rollenfunktion“ zumindest vorläufig mit Vorsicht und unter Berücksichtigung des beobachteten Unterschieds mit den Normwerten für die Originalversion verglichen werden sollten. Allerdings ist die Effektstärke von 0,18 für den beobachteten Unterschied als gering einzustufen, was deutlich macht, dass es sich um einen lediglich geringfügigen Unterschied zwischen den beiden Versionen handelt, der vielleicht auch aufgrund des multiplen statistischen Testens „signifikant“ wurde. Daher wollen wir auch nicht den Versuch unternehmen, die für uns überraschend aufgetretenen signifikanten Unterschiede zwischen den beiden Versionen inhaltlich zu deuten.

Beim Vergleich der Bearbeitung von SF-36 zuhause und in der Klinik zeigten sich ebenfalls entgegen unseren Erwartungen signifikante ( $p < 0,01$ ) und vergleichsweise etwas grössere Unterschiede. In der Klinik resultierten konsistent etwas günstigere Werte für die Lebensqualität als zuhause, wobei die Effektstärken hier Werte von bis zu 0,28 für die Skala „Vitalität“ erreichten. Methodisch einzuschränken ist diese Aussage dahingehend, dass diese Unterschiede natürlich prinzipiell auch durch einen zeitlichen Trend in Richtung Genesung verursacht worden sein könnten. Dies erscheint bei einem mittleren zeitlichen Abstand des Re-Tests von 13 Tagen und dem gegebenen Kontext jedoch nicht sehr wahrscheinlich. Wahrscheinlicher erscheint, dass Patienten in der Klinik tatsächlich ihren eigenen Gesundheitszustand im Vergleich mit den ebenfalls beeinträchtigten anderen Patienten günstiger beurteilen als zuhause, wo sie sich vermutlich überwiegend mit Gesunden vergleichen. Dies impliziert ganz wesentlich, dass in einer Klinik erhobene SF-36-Werte nicht ohne weiteres mit zuhause erhobenen Werten verglichen werden dürfen bzw. mindestens mit eher geringen, aber konsistenten Verzerrungen zugunsten einer positiveren Beurteilung der Lebensqualität in der Klinik zu rechnen ist.

Damit werden auch die kleinen, aber signifikanten Unterschiede zwischen SF-36 und SF-36m relativiert: Selbst wenn SF-36 und SF-36m vollständig äquivalent wären, würde die Beurteilung der Lebensqualität in der Klinik etwas günstiger



**Abbildung 4** Ein Untersuchungsdesign, bei dem langfristige Therapieeffekte durch Einsatz des SF-36 zuhause und kurzfristige Therapieeffekte durch Einsatz des SF-36m in der Klinik erfasst werden

ausfallen als zuhause, und SF-36-Werte aus dem heimischen Kontext dürften nicht – oder nur annäherungsweise – mit den in Kliniken erhobenen Werten verglichen werden. Wenn unsere Interpretation für diese beobachteten Unterschiede korrekt ist, so dürfte es sich dabei aber nicht um ein Problem handeln, das spezifisch für den SF-36 ist, sondern um ein allgemeines Problem, wenn man Patienten ihre Lebensqualität einmal in einer Klinik und einmal zuhause einschätzen lässt – unabhängig vom eingesetzten Fragebogen.

Wie kann man unter diesen Umständen sowohl die Veränderungen der Lebensqualität während eines Klinikaufenthalts als auch längerfristige Veränderungen zuverlässig abbilden? Abbildung 4 zeigt am Beispiel des SF-36 ein Untersuchungsdesign, bei dem sowohl längerfristige Veränderungen der Lebensqualität über den Einsatz des SF-36 zuhause als auch kurzfristige Veränderungen der Lebensqualität über den Einsatz des SF-36m in der Klinik zuverlässig erfassbar sind.

Dadurch, dass in Abbildung 4 sowohl die Beurteilung des kurzfristigen Reha-Erfolgs als auch die Beurteilung des langfristigen Reha-Erfolgs im gleichen Kontext erfolgt – in der Klinik bzw. zuhause –, lassen sich mit diesem Design die jeweiligen Therapie-Erfolge unverzerrt abbilden. Verzichtet man auf die Erhebung von Daten zu  $t_0$  zuhause, so erfolgt, wenn unsere Begründung für die beobachteten Unterschiede zwischen  $t_0$  und  $t_1$  richtig ist, bei der Beurteilung des langfristigen Therapieerfolgs – den Differenzen zwischen  $t_1$  und  $t_3$  – eine *Unterschätzung des Therapieerfolgs*: In der Klinik

beurteilen die Patienten ihre Lebensqualität durch den impliziten Vergleich mit anderen Patienten günstiger als zuhause. Mit dem in Abbildung 4 vorgeschlagenen Design wird auch deutlich, dass der SF-36m den SF-36 nicht ersetzen kann und soll, sondern dass es sich um eine minimal modifizierte Version des SF-36 handelt, die speziell beim Einsatz in Kliniken bei weitgehender Vergleichbarkeit mit dem SF-36 dessen bekannte Probleme in diesem Kontext vermeidet. Im Übrigen sei noch erwähnt, dass auch verschiedene Applikationsformen des SF-36, z.B. als Interview gegenüber dem üblichen Fragebogen zum Selbstausfüllen<sup>14</sup> bzw. die Akutversion mit einem Zeitfenster von einer Woche gegenüber den üblichen vier Wochen Retrospektionszeit<sup>15</sup> zu unterschiedlichen Ergebnissen zu führen scheinen, so dass auch bei vom Standard abweichenden Applikationsformen Vorsicht beim Vergleich mit den publizierten Normen oder anderen Applikationsformen geboten ist.

Kurz zusammengefasst lassen sich folgende Schlussfolgerungen ziehen:

- Eine minimal modifizierte Version des SF-36, die wir SF-36m genannt haben, ist im Gegensatz zur Originalversion auch am Ende von Klinikaufenthalten problemlos einsetzbar.
- Dieser SF-36m scheint weitgehend äquivalent zum Originalfragebogen zu sein, wobei einschränkend zu bemerken ist, dass jedoch möglicherweise die Skala „Körperliche Rollenfunktion“ im Vergleich zum Original leicht überschätzt wird.
- Bei Einsatz des SF-36 in Kliniken kommt es offensichtlich zu einer leichten bis moderaten Überschätzung der Lebensqualität verglichen mit der üblichen Applikation im häuslichen Kontext.

### Danksagung

Wir danken für die ausgezeichnete Zusammenarbeit bei der Datenerhebung C. Henke und A. Scheffel aus der Vogtlandklinik, Bad Elster, und E. Lischka und N. Lischka von der Malteser Klinik Dr. v. Weckbecker, Bad Brückenau. Für

schnelle und kompetente Reaktionen auf unsere e-mails mit verschiedenen Problemen und Fragen danken wir ausserdem M. Bullinger, B. Gandek, M. Kosinski, Ch. Rietz und Ch. Zwingmann. Abschliessend danken wir zwei anonymen Gutachtern für prompte und wertvolle Hinweise.

### Zusammenfassung

**Fragestellung:** Der Einsatz des 36 Item Short-Form Health Survey (SF-36) in Kliniken ist problematisch, da einige Formulierungen in dem Fragebogen explizit auf den häuslichen bzw. beruflichen Kontext bezogen und damit in einer Klinik sehr unpassend sind. Um diesem bereits bekannten Problem zu begegnen, entfernten wir aus drei einleitenden Formulierungen die für einen Klinikkontext klar unpassenden Ausdrücke – insgesamt 23 Wörter – und erstellten so eine modifizierte Version des SF-36, den SF-36m. In der vorliegenden Arbeit werden in erster Linie zwei Fragen untersucht: Inwieweit beseitigt die angedeutete Modifikation die beschriebenen Probleme bei Applikation des SF-36 in Kliniken, und inwieweit sind SF-36 und SF-36m noch äquivalent?

**Methoden:** Der SF-36m wurde an mehr als 300 Patienten in zwei Kliniken untersucht. In einer der Kliniken wurde mit einem randomisierten Design vor und zu Beginn einer Rehabilitationsmassnahme die Äquivalenz von SF-36 und SF-36m an 104 Patienten geprüft.

**Ergebnisse:** Die für den SF-36 beschriebenen hohen Anzahlen an fehlenden Werten wurden bei Einsatz des SF-36m am Ende von Klinikaufenthalten nicht mehr beobachtet. Es wurden lediglich kleine Unterschiede zwischen SF-36 und SF-36m mit Effektstärken  $< 0,20$  beobachtet, wobei jedoch die Skala „Körperliche Rollenfunktion“ und speziell Item 4d in der modifizierten Version unerwartet signifikant günstigere Wertannahmen ( $p < 0,10$ ). Zusätzlich zeigten sich jedoch grössere und signifikante Unterschiede abhängig davon, ob die Fragebogen zuhause oder in der Klinik ausgefüllt wurden.

**Schlussfolgerungen:** Der SF-36m ist im Gegensatz zum Original-SF-36 am Ende von Klinikaufenthalten problemlos einsetzbar. Dabei ist der SF-36m weitgehend äquivalent zur Originalversion. Allerdings scheint die Bearbeitung von SF-36 und SF-36m in Kliniken zu einer Überschätzung der Lebensqualität verglichen mit der Bearbeitung zuhause zu führen.

### Résumé

**Une version du SF-36 allemand convenable pour un contexte clinique et une comparaison psychométrique avec le questionnaire original**

**Objectifs:** L'utilisation du 36 Item Short-Form Health Survey (SF-36) dans des cliniques est problématique, parce que plusieurs formulations du questionnaire se réfèrent explicitement à un contexte domestique et professionnel, et pour cette raison sont très déplacés dans une clinique. Afin de remédier ce problème déjà bien connu, nous avons éliminé dans trois introductions les expressions inadéquates dans le contexte clinique – en totalité 23 mots – et nous avons généré une version modifiée du SF-36, le SF-36m. Dans la recherche décrite ici, nous avons avant tout analysé deux questions: A quel degré est-ce que la modification a éliminé les problèmes indiqués lors de l'application du SF-36 dans des cliniques, et à quel degré le SF-36 et le SF-36m restent-ils équivalents?

**Méthodes:** Le SF-36m a été analysé sur plus de 300 patients dans deux cliniques. Dans une des cliniques, on a procédé avec un design randomisé pour évaluer l'équivalence du SF-36 et du SF-36m avant et au début d'une intervention de réhabilitation sur 104 patients.

**Résultats:** Nous n'avons plus observé le grand nombre de réponses manquantes du SF-36 dans les cas où nous avons utilisé le SF-36m à la fin du séjour hospitalier. On a seulement observé des différences minuscules entre le SF-36 et le SF-36m (effet de masse  $< 0,20$ ), mais la dimension «Körperliche Rollenfunktion» (les limitations liées à l'état physique) et surtout l'item 4d dans la version modifiée à pris des valeurs qui étaient significatives à un degré inespéré ( $p < 0,10$ ). En plus, des différences plus ou moins grandes et significatives se sont montrées en fonction du lieu où les questionnaires ont été complétés à la maison ou dans la clinique.

**Conclusions:** Contrairement à l'original SF-36, le SF-36m peut être utilisé à la fin d'un séjour clinique sans problèmes. En même temps, le SF-36m est largement équivalent à la version originale. Mais l'utilisation du SF-36 et du SF-36m dans un contexte clinique semble résulter dans une évaluation plus positive de la qualité de vie comparé à l'utilisation à domicile.

## Literaturverzeichnis

- 1 *Wolinsky FD, Stump TE.* A measurement model of the Medical Outcomes Study 36-Item Short-Form Health Survey in a clinical sample of disadvantaged, older, black, and white men and women. *Med Care* 1996; **34**: 537–48.
- 2 *Ware JE, Snow KK, Kosinski M, Gandek B.* SF-36 Health Survey manual and interpretation guide. Boston, M.A.: New England Medical Center, The Health Institute, 1993.
- 3 *Hunt SM, McKenna SP, McEwen J, Williams J, Papp E.* The Nottingham Health Profile: subjective health and medical consultations. *Soc Sci Med* 1981; **15A**: 221–9.
- 4 *Kohlmann T, Bullinger M, Kirchberger-Blumstein I.* Die deutsche Version des Nottingham Health Profile (NHP): Übersetzungsmethodik und psychometrische Validierung. *Soz Präventivmed* 1997; **42**: 175–85.
- 5 *Kaplan RM, Anderson JP, Ganiats T.* The Quality of Wellbeing Scale: rationale for a single quality of life index. In: Walker S, Rosser M, eds. *Quality of life assessment: key issues in the 1990's.* Dordrecht: Kluwer Academic Press, 1993: 65–94.
- 6 *Bergner M.* Development, use and testing of the Sickness Impact Profile. In: Walker S, Rosser M, eds. *Quality of life assessment: key issues in the 1990's.* Dordrecht: Kluwer Academic Press, 1993: 95–110.
- 7 *Bullinger M.* German translation and psychometric testing of the SF-36 health survey: preliminary results from the IQOLA project. *Soc Sci Med* 1995; **41**: 1359–66.
- 8 *Gandek B, Ware JE, Aaronson NK, et al.* Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. *J Clin Epidemiol* 1998; **51**: 1171–8.
- 9 *Keller SD, Ware JE Jr, Bentler PM, et al.* Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. *International Quality of Life Assessment. J Clin Epidemiol* 1998; **51**: 1179–88.
- 10 *Zwingmann C, Metzger D, Jäckel WH.* Short Form-36 Health Survey (SF-36): psychometrische Analysen der deutschen Version bei Rehabilitanden mit chronischen Rückenschmerzen. *Diagnostica* 1998; **44**: 209–19.
- 11 *Walach H, Güthlin C.* Die psychometrischen Eigenschaften des deutschen MOS-SF-36-Health-Surveys – eine Analyse von Akupunkturpatienten. *ZGesundheitswiss* (im Druck).
- 12 *Bullinger M, Kirchberger I.* SF-36 Fragebogen zum Gesundheitszustand. Handanweisung. Göttingen: Hogrefe, 1998.
- 13 *Jöreskog KG, Sörbom D.* LISREL VII: a guide to the program and applications. Chicago, IL: SPSS Inc., 1988.
- 14 *Lyons RA, Wareham K, Lucas M, Price D, Williams J, Hutchings HA.* SF-36 scores vary by method of administration: implications for study design. *J Public Health Med* 1999; **21**: 41–5.
- 15 *Keller SD, Bayliss MS, Ware JE Jr, Hsu MA, Damiano AM, Goss TF.* Comparison of responses to SF-36 Health Survey questions with one-week and four-week recall periods. *Health Serv Res* 1997; **32**: 367–84.

## Korrespondenzadresse

**Dipl.-Psych. Dr. Horst Müller**  
**Forschungsinstitut für Balneologie**  
**und Kurortwissenschaft**  
**Lindenstrasse 5**  
**D-08645 Bad Elster**

**Tel.: ++49-37437/557-22**  
**Fax: ++49-37437/557-77**  
**e-mail: horst.mueller@fbk.sms.sachsen.de**