

Thomas Kohlmann¹, Monika Bullinger², Inge Kirchberger-Blumstein³

¹ Institut für Sozialmedizin, Medizinische Universität Lübeck

² Institut für Medizinische Psychologie, Universität Hamburg

³ Institut für Medizinische Psychologie, Universität München

Die deutsche Version des Nottingham Health Profile (NHP): Übersetzungsmethodik und psychometrische Validierung

Zusammenfassung

Das Nottingham Health Profile wurde Ende der 70er Jahre in Grossbritannien entwickelt und hat seither als Instrument zur Patientenselbstbeurteilung der subjektiven Gesundheit im englischen Sprachraum, seit Mitte der 80er Jahre zunehmend auch in anderen Ländern, Verbreitung erfahren. Im folgenden wird die deutschsprachige Übersetzung des Nottingham Health Profiles sowie die Prüfung seiner psychometrischen Eigenschaften an 10 Stichproben von insgesamt über 1000 gesunden und erkrankten Personen beschrieben. Die Prüfung der Gütekriterien (Praktikabilität, Reliabilität, Validität, Sensitivität) verlief befriedigend. Insgesamt erscheint das Nottingham Health Profile als praktikables, zuverlässiges und gültiges Instrument zur Messung der subjektiven Gesundheit, das sowohl in klinischen als auch in epidemiologischen Studien eingesetzt werden kann.

Seit etwa Mitte der 70er Jahre ist ein Wandel im Verständnis der Zielkriterien medizinischen Handelns zu beobachten: Zur angemessenen Beurteilung von Krankheitsfolgen und Behandlungsergebnissen erschien es notwendig, neben den objektivierbaren medizinischen Kriterien (z.B. klinische Symptomatik, Überlebenszeit) auch die „subjektive Gesundheit“ bzw. die „gesundheitsbezogene Lebensqualität“ in die Betrachtung einzubeziehen. Damit erweiterte sich einerseits – entsprechend der Definition der WHO – der Gesundheitsbegriff um psychosoziale Komponenten, andererseits

wurde der eigenständige Wert der Selbstausskunft von Patienten über ihre gesundheitliche Verfassung anerkannt.

Im Zuge dieser Entwicklung wurde speziell im anglo-amerikanischen Sprachraum eine Reihe von Instrumenten zur Messung der subjektiven Gesundheit und ihrer verschiedenen Dimensionen entwickelt. Hierzu gehören z.B. das *Sickness Impact Profile*¹, der *Quality of Well-Being Index*², der *Psychological General Well-Being Index*³, der *Medical Outcome Study Fragebogen*⁴ und das *Nottingham Health Profile*^{5,6}. Diese Messverfahren stellten gewissermassen

die erste Generation von Instrumenten zur *krankheitsübergreifenden* Messung der subjektiven Gesundheit dar.

Seit ihrer Entwicklung sind diese Messverfahren vielfach sowohl in der klinischen als auch in der epidemiologischen Forschung eingesetzt worden. Die methodische Kontrolle und Validierung der Übersetzung und Adaptation von Originalinstrumenten in andere Sprachen erfolgte dabei auf z.T. sehr unterschiedlichem Niveau. Erst in jüngster Zeit wurden Richtlinien zur Übersetzung, psychometrischen Prüfung und Normierung von Messinstrumenten vorgeschlagen⁷⁻¹⁰.

Das *Nottingham Health Profile (NHP)* gehört zu den am häufigsten verwendeten Instrumenten zur Erfassung des subjektiven Gesundheitszustands, es kann heute als eines der Standardverfahren zur Messung der gesundheitsbezogenen Lebensqualität bezeichnet werden. Neben der Originalversion sind autorisierte, nach einem standardisierten Protokoll¹¹ übersetzte und getestete Versionen des NHP derzeit in sieben Ländern verfügbar (Frankreich, Spanien, Schweden, Italien, Niederlande, Dänemark, Finnland)¹², weitere nationale Fassungen werden zur Zeit vorbereitet. Die vorliegende

Arbeit beschreibt die Übersetzung und Validierung der autorisierten deutschsprachigen Version des NHP.

Material und Methoden

Methodisches Vorgehen bei der Übersetzung

Das NHP ist ein Selbstausfüll-Fragebogen, dessen 38 Einzelfragen („Items“) sechs Dimensionen der subjektiven Gesundheit zugeordnet sind (Tabelle 1). Die Items sind als Aussagen formuliert, zu denen die Befragten angeben sollen, ob diese „im Moment“ für sie zutreffen („ja“) oder nicht zutreffen („nein“). Die Bearbeitungszeit beträgt in der Regel weniger als 10 Minuten. In einem früher noch im NHP enthaltenen zweiten Teil war nach übergreifenden gesundheitsbedingten Einschränkungen in sieben Lebensbereichen gefragt worden, die Anwendung dieses Teils wird jedoch nicht mehr empfohlen.

Die Übersetzung fand in mehreren, aufeinander aufbauenden Schritten statt:

– Zunächst wurde von einer der Autorinnen (M.B.) eine durch Rückübersetzung kontrollierte Erstübersetzung des englischen Originals angefertigt. Ein Vergleich mit zwei bereits vorhandenen deutschen Übersetzungen

ergab eine hohe Übereinstimmung. Die Erstübersetzung wurde im Vorfeld der weiteren Schritte in vier Stichproben verwendet (s.u.: Stichprobe 1–4).

– Drei deutsch-muttersprachige Personen wurden unabhängig voneinander um eine schriftliche Übersetzung gebeten (Medizin-Soziologe, Ärztin, Krankenschwester). Im Vergleich dieser Übersetzungen untereinander und mit der Erstübersetzung ergaben sich geringfügige Abweichungen, die jedoch im allgemeinen als sprachlich äquivalente Ausdrucksformen zu bewerten waren und nur bei sechs Items zu einer leichten Veränderung der Erstübersetzung führten.

– Die jetzt vorliegende modifizierte Fassung des Fragebogens wurde einer Gruppe von vier nur deutschsprachigen Personen aus nicht-akademischen Berufen vorgelegt. Die Laienrunde sollte die Verständlichkeit und alltagsprachliche Angemessenheit der Übersetzung beurteilen und ggf. Veränderungsvorschläge machen. Die Ergebnisse der Laienrunde bekräftigten die bereits vorgenommenen Veränderungen der Erstübersetzung, sie deuteten im übrigen auf eine gute alltagsprachliche Akzeptanz hin.

– In eingehenden Diskussionen mit den beteiligten Übersetzern wurden die Unterschiede zwischen den Übersetzungsvarianten und die Vorschläge aus dem Laienpanel

besprochen. Eine Rückübersetzung der Endversion durch eine englisch-muttersprachige Kollegin ergab nur stilistische, jedoch in keinem Fall inhaltlich relevante Abweichungen vom englischen Original.

Untersuchte Stichproben und Vergleichskriterien

Mit der psychometrischen Prüfung im Rahmen der Felderprobung in mehreren Stichproben sollten die Praktikabilität (Zeitbedarf, „missing values“), die Reliabilität (innere Konsistenz, Test-Retest) und die kriterienbezogene Validität des NHP-Fragebogens sowie dessen Veränderungssensitivität untersucht werden. Es handelt sich bei den Stichproben der Felderprobung sowohl um „gesunde“ Personen (Studenten, Bevölkerung) als auch um Patienten in ambulanter und stationärer (insbesondere rehabilitationsmedizinischer) Behandlung. In vier der Stichproben wurden Wiederholungsmessungen durchgeführt (Tabelle 2).

Zur Überprüfung der kriterienbezogenen Validität wurden in den untersuchten Kollektiven verschiedene Vergleichskriterien erhoben. Dies waren zum einen inhaltlich und methodisch verwandte „generische“ Verfahren zur Messung der Lebensqualität (SF 36, ALLTAG), zum anderen stärker spezialisierte Fragebögen, mit denen Einzelkomponenten der subjektiven Gesundheit (Funktionskapazität, Schmerzintensität, Depressivität) oder Störungen des Allgemeinbefindens bzw. psychovegetative Symptome (Beschwerdenliste) erfasst wurden (Tabelle 3).

Auswertung

Zur psychometrischen Prüfung des NHP wurden einerseits Methoden aus dem Bereich der klassischen Testanalyse (innere Konsistenz, Test-Retest-Reliabilität, Korrela-

Dimension	Anzahl Items	Beispielitem
Energy	3	I'm tired all the time
Pain	8	I have pain at night
Emotional reactions	9	Things are getting me down
Sleep	5	I take tablets to help me sleep
Social isolation	5	I feel lonely
Physical mobility	8	I can only walk about indoors

Tabelle 1. Dimensionen, Anzahl der Items und Originalbeispiele von Items des Nottingham Health Profile.

Nr.	Stichprobe	Diagnose	Jahr	N	Geschlecht (% weiblich)	Alter (\bar{x} , s)
1	Medizinstudenten	-	1992 ²	375	41	23,1, 2,4
2	Rehabilitationspatientinnen ¹	Mamma-Karzinom	1991 ²	139 ³	100	58,1, 9,4
3	Rehabilitationspatienten	tiefe Rückenschmerzen	1991 ²	75	35	45,4, 12,3
4a	Rehabilitationspatienten	chronische Polyarthrit	1991 ²	60	77	51,5, 12,1
4b		unspezifische Rückenschmerzen	1991 ²	33	52	53,4, 8,0
4c		Cox-/Gonarthrose	1991 ²	29	62	59,2, 7,7
5	Einwohner einer nord-deutschen Stadt	-	1991	308	51	46,8, 13,9
6	Rehabilitationspatienten	chronische Polyarthrit	1991	62	73	60,3, 11,8
7	Rehabilitationspatienten ¹	unspezifische Rückenschmerzen	1992	31 ³	65	49,6, 8,3
8a	Rehabilitationspatienten ¹	internistische Krankheiten	1992	39 ³	49	51,6, 11,5
8b		muskulo-skeletale Krankheiten	1992	37 ³	46	50,3, 10,1
9	Ambulante Patienten	Rückenschmerzen	1994	268	51	49,7, 8,3
10	Patienten ¹	arterielle Verschlusskrankheit	1993	308 ³	32	65,5, 10,5

¹ Stichproben mit Längsschnitterhebung zu zwei Messzeitpunkten.

² Die Befragung erfolgte mit der Erstübersetzung (siehe Text).

³ Anzahl der Fälle mit Daten zum ersten Messzeitpunkt.

Tabelle 2. Untersuchte Stichproben, Diagnosen, Untersuchungsjahr, Fallzahl, Geschlecht und Alter.

Variable	Abkürzung	Autoren	Stichprobe
Lebensqualität	SF-36	dt.: Bullinger et al. 1995 ²¹	1, 9, 10
Lebensqualität	ALLTAG	Bullinger et al. 1993 ²²	9
Beschwerden	BL	von Zerssen 1976 ²³	3, 4a-c, 6, 7, 8a-b
Depressivität	CES-D	dt.: Hautzinger u. Bailer 1993 ²⁴	4a-c, 6, 7, 8a-b
Funktionskapazität	FFbH-R	Kohlmann u. Raspe 1994 ²⁵	3, 4a-c, 6, 7, 8a-b
Funktionskapazität	HAQ	Fries et al. 1980 ²⁶	6, 7, 8a-b
Schmerzintensität	NRS	(numerische Ratingskala, 1-10)	3, 4a-c, 6, 7, 8a-b

Tabelle 3. Vergleichskriterien zur Untersuchung der kriterienbezogenen Validität.

tion mit externen Kriterien) verwendet, zum anderen wurde die Skalenstruktur des Instruments auf der Grundlage eines „multi-trait/multi-item“-Ansatzes (MTMI) überprüft¹³. In der MTMI-Analyse wird geprüft, in welchem Maße die einer (Sub-)Skala zugeordneten Items tatsächlich höher mit der eigenen Skala korrelieren als mit anderen Skalen („Konstruktvalidität“). Das von uns verwendete Programm NEWMAP berechnet nicht nur die zur Be-

urteilung der Skaleneigenschaften nach dem MTMI-Ansatz erforderlichen Indizes, es werden darüber hinaus weitere Angaben z.B. zu Boden- bzw. Deckeneffekten ausgegeben. Soweit geeignete Parameter zur Kennzeichnung von Subgruppen mit unterschiedlichem Schweregrad der Erkrankung bzw. der Symptombelastung zur Verfügung standen, wurden die NHP-Skalenwerte dieser Subgruppen verglichen, um dadurch weitere Validitätshinweise zu erhalten. Eine

Überprüfung der Sensitivität des NHP gegenüber Veränderungen im Zeitverlauf konnte bei den Patientengruppen erfolgen, die in unseren Längsschnittstudien vor und nach einer therapeutischen Intervention befragt wurden. Hierzu wurden Mittelwertsunterschiede zwischen zwei Messzeitpunkten mit parametrischen (t-Test) und nichtparametrischen Verfahren (Wilcoxon-Test) untersucht. Die „relative Effizienz“ der NHP-Subskalen bei der Veränderungsmes-

sung wurde durch Bildung von Quotienten aus t-Werten (abhängige Stichproben) für die NHP-Skalen und für Vergleichskriterien bestimmt¹⁴. Für die Auswertung benutzten wir neben dem bereits genannten Programm NEWMAP die Statistikprogramme SAS und SPSS.

Ergebnisse

Deskriptive Kennwerte und Reliabilität

In Tabelle 4 sind verschiedene deskriptive Kennwerte der NHP-Subskalen in den untersuchten Stichproben gegenübergestellt.

Gemessen an den Anteilen fehlender Werte, die in den meisten Stichproben im Bereich von 0 bis 3% lagen, ist der NHP-Fragebogen als ausreichend praktikables Instrument zu betrachten. Ausfälle aufgrund der nur zweistufigen Antwortvorgabe wurden nur selten beobachtet.

Zwischen den gesunden Vergleichskollektiven und den Patientenstichproben zeigten sich erwartungskonforme Unterschiede. In den gesunden Vergleichskollektiven der Stichproben 1 und 5 (bei der NHP-Dimension *Soziale Isolation* auch in den Patientengruppen) sind „Bodeneffekte“ zu verzeichnen. Dies bestätigt den bereits in der Literatur beschriebenen Be-

fund¹⁵, nach dem das NHP nur bedingt für eine Verwendung in überwiegend gesunden Populationen geeignet ist.

Erwartungsgemäss ergaben sich bei den Patienten mit Krankheiten der Bewegungsorgane höhere Mittelwerte insbesondere in den Subskalen *Schmerz* und *Physische Mobilität*, als am stärksten eingeschränkte Gruppen stellen sich die überwiegend älteren Patienten mit chronischer Polyarthrit in Stichprobe 6 und die Rückenschmerzpatienten dar. Die Dimensionen *Energieverlust*, *Emotionale Reaktion* und *Schlaf* waren in den erkrankten Stichproben ähnlich ausgeprägt.

Subskala	Kennwert	Stichprobe												
		1	2	3	4a	4b	4c	5	6	7	8a	8b	9	10
Energieverlust	Mittelwert ¹	8,5	37,3	35,7	41,1	42,4	34,5	17,3	53,8	32,3	36,8	26,7	55,4	41,3
	Standardabweichung	17,2	37,8	37,5	37,0	37,5	33,9	30,2	32,7	32,8	39,6	31,1	39,7	38,4
	Anteil fehlender Werte (%) ²	0,5	10,7	5,3	0,0	0,0	0,0	3,9	0,0	0,0	0,0	5,4	1,5	0,7
	Skalenwert 0 (%) ³	72,8	41,6	43,7	35,0	33,3	37,9	70,3	16,1	38,7	46,2	45,7	25,9	20,6
	Skalenwert 100 (%) ⁴	3,0	16,8	15,5	16,7	18,2	10,3	6,1	19,4	9,7	17,9	8,6	32,5	36,6
	Innere Konsistenz ⁵	0,60	0,69	0,70	0,64	0,65	0,49	0,72	0,48	0,54	0,76	0,48	0,71	0,69
	Test-Retest-Korrelation ⁶	nu ⁷	0,68	nu	nu	nu	nu	nu	nu	0,72	0,40	0,51	nu	0,75
Schmerz	Mittelwert ¹	2,8	22,4	37,6	51,9	58,3	50,4	13,3	63,7	35,3	23,7	38,2	62,5	46,1
	Standardabweichung	7,9	27,2	34,9	30,0	29,9	28,1	24,2	25,7	27,4	31,0	31,0	31,9	28,3
	Anteil fehlender Werte (%) ²	0,5	6,5	4,0	0,0	0,0	0,0	2,9	1,6	0,0	0,0	2,7	1,5	0,7
	Skalenwert 0 (%) ³	94,9	40,0	13,9	10,0	6,1	6,9	66,9	1,6	25,8	43,6	16,7	5,7	11,4
	Skalenwert 100 (%) ⁴	0,0	2,3	18,1	5,0	6,1	6,9	0,0	6,6	0,0	5,1	5,6	23,6	11,6
	Innere Konsistenz ⁵	0,77	0,82	0,88	0,80	0,80	0,74	0,86	0,75	0,73	0,88	0,81	0,80	0,79
	Test-Retest-Korrelation ⁶	nu	0,85	nu	nu	nu	nu	nu	nu	0,82	0,72	0,65	nu	0,78
Emotionale Reaktion	Mittelwert ¹	7,9	19,4	16,0	18,8	17,5	18,0	13,3	19,1	17,9	18,8	11,7	32,8	21,6
	Standardabweichung	13,5	21,1	22,1	20,6	23,4	22,3	20,2	21,2	16,2	22,5	17,6	30,6	27,7
	Anteil fehlender Werte (%) ²	0,5	8,6	2,7	1,7	0,0	0,0	2,9	1,6	0,0	0,0	2,7	1,5	0,7
	Skalenwert 0 (%) ³	65,1	32,0	47,9	33,9	45,5	31,0	56,5	36,1	25,8	43,6	52,8	25,9	41,7
	Skalenwert 100 (%) ⁴	0,0	0,0	0,0	0,0	0,0	3,4	0,3	0,0	0,0	0,0	0,0	3,8	4,2
	Innere Konsistenz ⁵	0,67	0,70	0,77	0,69	0,82	0,78	0,80	0,69	0,51	0,78	0,74	0,81	0,84
	Test-Retest-Korrelation ⁶	nu	0,71	nu	nu	nu	nu	nu	nu	0,49	0,33	0,64	nu	0,70

Tabelle 4. Statistische Kennwerte (Verteilungsparameter, Anteil fehlender Werte, Boden- und Deckeneffekte, innere Konsistenz und Test-Retest-Korrelation) der NHP-Subskalen in den untersuchten Stichproben.

Subskala	Kennwert	Stichprobe												
		1	2	3	4a	4b	4c	5	6	7	8a	8b	9	10
Schlaf	Mittelwert	5,2	34,8	29,6	40,0	36,4	48,3	15,3	42,9	33,6	31,3	26,0	41,3	34,5
	Standardabweichung	10,4	30,7	30,0	28,3	28,5	30,5	23,8	32,4	29,4	31,1	28,6	30,9	31,8
	Anteil fehlender Werte (%)	0,5	6,4	0,0	0,0	0,0	0,0	2,3	0,0	0,0	0,0	0,0	1,5	0,7
	Skalenwert 0 (%)	68,3	30,5	33,3	11,7	12,1	10,3	61,1	21,0	25,8	33,3	40,5	19,8	18,1
	Skalenwert 100 (%)	0,0	3,1	5,3	5,0	6,1	6,9	0,7	3,2	0,0	2,6	2,7	5,3	5,6
	Innere Konsistenz	0,63	0,72	0,76	0,67	0,66	0,72	0,71	0,76	0,73	0,79	0,72	0,76	0,79
	Test-Retest-Korrelation	nu	0,70	nu	nu	nu	nu	nu	nu	0,71	0,56	0,74	nu	0,74
Soziale Isolation	Mittelwert	13,4	11,6	6,7	7,3	12,1	9,0	4,4	7,1	7,1	8,7	4,3	15,3	15,1
	Standardabweichung	20,1	19,7	16,2	16,5	22,3	20,4	15,0	13,6	15,1	17,7	10,7	23,9	26,4
	Anteil fehlender Werte (%)	0,5	4,3	0,0	0,0	0,0	0,0	1,9	0,0	0,0	0,0	0,0	1,5	0,7
	Skalenwert 0 (%)	75,7	65,7	80,0	75,0	72,7	72,4	88,4	74,2	77,4	71,8	83,8	60,5	64,1
	Skalenwert 100 (%)	0,5	0,0	1,3	0,0	0,0	3,4	1,0	0,0	0,0	0,0	0,0	2,3	5,2
	Innere Konsistenz	0,66	0,56	0,67	0,65	0,72	0,75	0,79	0,13	0,57	0,61	0,24	0,68	0,80
	Test-Retest-Korrelation	nu	0,52	nu	nu	nu	nu	nu	nu	0,64	0,77	0,78	nu	0,70
Physische Mobilität	Mittelwert	3,7	18,6	28,0	31,0	34,1	37,9	8,4	54,4	19,4	18,6	28,1	37,5	33,3
	Standardabweichung	7,6	21,6	18,4	20,8	17,5	15,5	14,7	21,4	13,6	20,5	21,8	25,3	25,6
	Anteil fehlender Werte (%)	0,5	7,1	1,3	0,0	0,0	0,0	1,9	0,0	0,0	0,0	2,7	1,5	0,7
	Skalenwert 0 (%)	90,1	36,9	9,5	10,0	3,0	10,3	68,9	1,6	22,6	41,0	19,4	12,5	10,5
	Skalenwert 100 (%)	0,0	1,5	1,4	1,7	0,0	13,8	0,0	1,6	0,0	0,0	0,0	1,1	5,2
	Innere Konsistenz	0,50	0,75	0,61	0,63	0,53	0,32	0,68	0,60	0,32	0,67	0,68	0,79	0,80
	Test-Retest-Korrelation	nu	0,78	nu	nu	nu	nu	nu	nu	0,89	0,73	0,68	nu	0,75

¹ Zur Rohwertberechnung wurde die Anzahl der „ja“-Antworten auf den Bereich von 0 bis 100 normiert (0 = alle Items verneint, 100 = alle Items bejaht). Hohe Werte bedeuten also ungünstige Ausprägungen. Eine Gewichtung der einzelnen Items wurde nicht vorgenommen.

² Mit Ausnahme der Subskala „Energieverlust“ wurde bei der Skalenberechnung ein Antwortausfall bei höchstens einem Item zugelassen. Fehlende Angaben wurden als „nein“ gewertet.

³ „Bodeneffekt“.

⁴ „Deckeneffekt“.

⁵ Cronbachs Koeffizient Alpha.

⁶ Spearmans Rangkorrelation zwischen den Skalenwerten zum ersten und zweiten Messzeitpunkt. Die Zeitabstände zwischen den Messungen betragen ca. 1 Woche (Stichproben 7 und 10) und 4 Wochen (Stichproben 2, 8a und 8b).

⁷ nu: nicht untersucht.

Tabelle 4 (Fortsetzung).

Abgesehen von wenigen „Ausreißerwerten“ erreichten die Kennzahlen zur Beurteilung der inneren Konsistenz (Cronbachs Alpha) zufriedenstellende Werte. Sie liegen je nach Skala im Bereich von etwa 0,65 bis 0,85. Die Schätzung der Reliabilität auf der Grundlage der Test-Retest-Korrelation bestätigt im wesentlichen die Befunde aus der Analyse der inneren Konsistenz.

Konstrukt- und kriterienbezogene Validität

Die Überprüfung der inneren Struktur des NHP erfolgte mit Hilfe des weiter oben erläuterten MTMI-Ansatzes. In den Analysen wurden zwei gesunde Kollektive (Studenten, Bevölkerungsstichprobe) und drei Diagnosegruppen (Brustkrebs, chronische Polyarthrit, Rückenschmerzen) unter-

sucht. Um ausreichend präzise Schätzungen der Korrelationskoeffizienten zu erhalten, wurden bei der chronischen Polyarthrit die Stichproben 4a und 6, bei den Rückenschmerzen die Stichproben 3, 4b und 7 zu einer grösseren Stichprobe zusammengefasst. Nach dem Ergebnis der MTMI-Auswertungen konnte die für den NHP angenommene innere Struktur weitgehend bestätigt werden

	Stich- probe 1	Stich- probe 5	Stich- probe 2	Stich- probe 4a/6	Stich- probe 3/4 b/7	Stich- probe 9	Stich- probe 10
Energieverlust	100,0%	93,3%	93,3%	80,0%	86,7%	93,3%	80,0%
Schmerz	100,0%	97,5%	90,0%	97,5%	97,5%	94,3%	97,1%
Emotionale Reaktion	95,4%	93,3%	88,9%	93,3%	88,9%	97,5%	92,5%
Schlaf	96,7%	100,0%	100,0%	100,0%	88,0%	96,7%	90,0%
Soziale Isolation	96,0%	100,0%	88,0%	88,0%	96,0%	92,0%	88,0%
Physische Mobilität	80,0%	75,0%	82,5%	95,0%	75,0%	93,3%	93,0%

¹ Gezeigt wird der prozentuale Anteil von Item-Subskalen-Korrelationen, bei denen die Items höher mit der ihnen zugeordneten Skala als mit jeder „Fremdskala“ korrelieren. Die Stichproben werden nach der Art der Gesundheitsstörung (ggf. zusammengefasst) aufgeführt: gesunde Kollektive (Stichprobe 1 und 5), Brustkrebs (Stichprobe 2), chronische Polyarthritis (Stichprobe 4 a und 6), Rückenschmerzen (Stichprobe 3, 4 b und 7 bzw. 9), arterielle Verschlusskrankheit (Stichprobe 10).

Tabelle 5. Ergebnisse der MTMI-Analyse¹.

(Tabelle 5). In der Regel war bei über 90% der betrachteten Korrelationen eine erwartungskonforme Zuordnung der Items zu den entsprechenden Subskalen festzustellen.

Zur Beurteilung der kriterienbezogenen Validität sind in der Tabelle 6 zunächst die Korrelationen der NHP-Skalen mit den Subskalen des SF-36 und des ALLTAG-Fragebogens in verschiedenen Stichproben dargestellt. Wie sich zeigt, bestehen substantielle Korrelationen zwischen den auf vergleichbare Dimensionen bezogenen Skalen (z.B. physische Funktion, psychische Verfassung, Energie, Schmerz). Im allgemeinen sind die auf inhaltlich verwandte Konstrukte bezogenen Subskalen untereinander höher korreliert als mit Subskalen, die unterschiedliche Konstrukte messen sollen.

Bei der Darstellung der Korrelationen mit den spezialisierten Vergleichskriterien beschränken wir uns auf die Betrachtung der Stichproben 3, 4a–c, 6, 7 und 8b (Patienten mit muskulo-skeletalen Erkrankungen). Konstruktähnliche Kriterien stehen für die Bereiche körperliche Funktion (FFbH-R, HAQ), Schmerz (NRS), emotionale Verfassung (CES-D) und – mit Einschränkungen – für Energieverlust/Vitalität (BL) zur Verfü-

gung (zu den NHP-Skalen *Schlaf* und *Soziale Isolation* wurden in diesen Stichproben keine Vergleichskriterien erhoben). Wiederum bestehen im Sinne der konvergenten/diskriminanten Validität die jeweils höchsten Assoziationen mit den auf inhaltlich verwandte Konstrukte bezogenen Kriterien. In der Tendenz ergeben sich stärkere Zusammenhänge als mit den Subskalen der generischen Lebensqualitäts-Fragebögen (unterer Teil der Tabelle 6).

Vergleiche von Subgruppen

In vier der untersuchten Stichproben wurden Vergleiche der NHP-Skalenwerte nach einer Bildung von Subgruppen angestellt. Bei den Patientinnen mit Mamma-Karzinom (Stichprobe 2) wurden Fälle mit und ohne Rezidiv gegenübergestellt, in der Stichprobe von Rehabilitationspatienten mit Rückenschmerzen im Bereich der LWS (Stichprobe 3) wurden anhand einer Fremdbeurteilung durch den Arzt (Steinbrocker-Index) Patienten mit geringer und höherer Funktionseinschränkung unterschieden, in der Bevölkerungsstichprobe (Stichprobe 5) erfolgte ein Vergleich von Befragten mit und ohne aktuell bestehende Schmerzen (erfragt anhand einer Liste von 11

Körperregionen, vgl.¹⁶) und in der Gruppe der AVK-Patienten (Stichprobe 10) untersuchten wir den Kontrast zwischen den Fontaine-Stadien II und III/IV. Die Ergebnisse dieser Gruppenvergleiche sind in Tabelle 7 dargestellt.

Beim Vergleich der Brustkrebspatientinnen mit und ohne Rezidiv zeigte sich, dass Patientinnen mit Rezidiv im NHP in allen Dimensionen höhere Werte, d.h. eine stärker beeinträchtigte Verfassung angaben. Diese Unterschiede konnten bei vier der NHP-Subskalen statistisch gesichert werden.

Die Gruppe der vom Arzt als funktionell stärker beeinträchtigt eingestuften Rückenschmerzpatienten unterschied sich von den weniger eingeschränkten Patienten insbesondere hinsichtlich der NHP-Skalen *Schmerz* und *Physische Mobilität*, statistisch bedeutsame Gruppenunterschiede ergaben sich ebenso bei den Skalen *Energie* und *Schlaf*.

Befragte in der Bevölkerungsstichprobe mit Schmerzen in einer oder in mehreren Körperregionen waren deutlich stärker in ihrer subjektiven Gesundheit eingeschränkt, diese Unterschiede blieben auch nach statistischer Kontrolle von Alter und Geschlecht (hierarchische multiple Regression) erhalten.

Instrument	(Sub-)Skala	Subskala des NHP					
		Energieverlust	Schmerz	Emotionale Reaktion	Schlaf	Soziale Isolation	Physische Mobilität
SF-36 ¹	Körperl. Funktionsfähigkeit	0,16–0,53	0,35–0,62	0,08–0,44	0,18–0,40	0,06–0,38	0,42–0,78
	Soziale Funktionsfähigkeit	0,21–0,51	0,22–0,40	0,37–0,53	0,33–0,34	0,39–0,47	0,18–0,43
	Körperl. Rolleneinschränkung	0,25–0,53	0,24–0,43	0,19–0,46	0,26–0,34	0,16–0,34	0,24–0,55
	Emotionale Rolleneinschränkung	0,24–0,42	0,18–0,46	0,34–0,50	0,28–0,43	0,31–0,39	0,14–0,36
	Psychisches Wohlbefinden	0,26–0,58	0,20–0,46	0,51–0,74	0,35–0,52	0,41–0,54	0,19–0,46
	Vitalität	0,50–0,69	0,19–0,48	0,38–0,67	0,31–0,54	0,26–0,47	0,21–0,53
	Schmerz	0,17–0,53	0,38–0,67	0,19–0,47	0,25–0,44	0,06–0,37	0,35–0,61
	Allgem. Gesundheitswahrnehmung	0,28–0,48	0,08–0,42	0,24–0,50	0,23–0,43	0,21–0,33	0,00–0,46
ALLTAG ²	Körper	0,60	0,47	0,56	0,57	0,40	0,40
	Psyche	0,52	0,31	0,46	0,46	0,48	0,26
	Alltag	0,56	0,45	0,46	0,46	0,32	0,54
	Sozialleben	0,41	0,26	0,36	0,30	0,54	0,23
	Lebensfreude	0,52	0,37	0,40	0,40	0,48	0,32
	Medizinische Versorgung	0,19	0,16	0,14	0,15	0,12	0,05
BL ³	Allgemeinbeschwerden	0,59–0,61	0,07–0,49	0,52–0,68	0,13–0,63	0,13–0,44	–0,01–0,35
FFbH-R ³	Funktionskapazität ⁶	0,16–0,41	0,43–0,73	–0,20–0,40	–0,09–0,22	0,03–0,31	0,53–0,87
HAQ ⁴	Funktionsbehinderung	0,15–0,43	0,47–0,68	–0,11–0,19	–0,09–0,21	–0,04–0,22	0,79–0,84
NRS ³	Schmerzintensität	0,29–0,38	0,57–0,70	0,02–0,26	–0,02–0,35	–0,10–0,28	0,39–0,76
CES-D ⁵	Depressivität	0,38–0,60	0,24–0,55	0,59–0,72	0,29–0,61	0,38–0,67	0,17–0,27

¹ Stichproben 1, 9 und 10.

² Stichprobe 9.

³ Stichproben 3, 4a–c, 6, 7 und 8b.

⁴ Stichproben 6, 7 und 8b.

⁵ Stichproben 4a–c, 6, 7 und 8b.

⁶ Die FFbH-Werte wurden vor der Berechnung umgepolt, so dass hohe Werte einer niedrigen Funktionskapazität entsprechen.

Tabelle 6. Wertebereich der Produkt-Moment-Korrelation zwischen den NHP-Subskalen und anderen Instrumenten zur Messung der Lebensqualität sowie mit Vergleichskriterien bei Patienten mit muskulo-skeletalen Krankheiten.

Bei den AVK-Patienten waren diejenigen mit höherem Fontaine-Stadium (III und IV) in allen NHP-Dimensionen statistisch signifikant beeinträchtigt als die im niedrigeren Fontaine-Stadium.

Veränderungssensitivität

Die Sensitivität des NHP gegenüber Veränderungen nach therapeutischen Interventionen konnte in den zu zwei Zeitpunkten befrag-

ten Patientenkollektiven analysiert werden. Bei den Patientinnen mit Mamma-Karzinom (Stichprobe 2) zeigten sich im Vergleich der NHP-Werte vor und nach der vierwöchigen Rehabilitation signifikante Verbesserungen ($p < 0,05$) in den Dimensionen *Energie*, *Emotionale Reaktion*, *Soziale Isolation* und *Physische Mobilität*, bei den Patienten mit arterieller Verschlusskrankheit (Stichprobe 10) ergaben sich ähnliche Effekte, diese waren

allerdings weniger stark ausgeprägt. Interventionseffekte waren auch bei den Rückenschmerzpatienten zu finden, die nach ihrer Therapie in der Münchner Schmerzzambulanz Verbesserungen der NHP-Profile zeigten.

Eine vergleichende Beurteilung der Änderungssensitivität durch Berechnung der „relativen Effizienz“ haben wir in der Stichprobe von ebenfalls vor und nach einem vierwöchigen Rehabilita-

Stichprobe	Gruppierung	Subskala des NHP					
		Energieverlust	Schmerz	Emotionale Reaktion	Schlaf	Soziale Isolation	Physische Mobilität
Mamma-Karzinom	kein Rezidiv	20,4	19,2	12,3	29,7	8,2	12,4
	Rezidiv	41,6*	26,8	35,1*	36,2	19,7*	27,1*
Tiefe Rückenschmerzen	Steinbrocker <2	24,0	38,0	11,5	22,3	6,9	18,3
	Steinbrocker ≥2	46,3*	72,6*	21,9	36,3*	7,9	36,1*
Bevölkerung	keine Schmerzen	8,1	1,3	8,3	7,7	3,9	0,9
	Schmerzen	26,8*	25,5*	18,3*	22,8*	5,0	15,9*
Arterielle Verschlusskrankheit	Stadium II	32,2	36,6	14,7	25,1	9,6	25,8
	Stadium III/IV	58,8*	66,3*	36,4*	54,2*	27,4*	49,3*

* $p < 0,05$ im Kruskal-Wallis-Test.

Tabelle 7. Mittelwertsvergleiche der NHP-Skalen bei Untergruppen mit unterschiedlicher Krankheitsausprägung in verschiedenen Stichproben.

NHP-Skala	Vergleichskriterium	t-Wert NHP-Skala	t-Wert Vergleichskriterium	Relative Effizienz
Energieverlust	BL	3,25	5,17	0,63
Schmerz	NRS	2,58	3,40	0,76
Emotionale Reaktion	CES-D	3,94	5,04	0,78
Physische Mobilität	FFbH-R	3,50	2,76	1,27
	HAQ	3,50	2,84	1,23

Stichprobe 8a und 8b. Die relative Effizienz ist gleich dem Quotienten der t-Werte für verbundene Stichproben. Die t-Werte berechnen sich aus der Differenz der Mittelwerte zu zwei Messzeitpunkten dividiert durch den Standardfehler der Differenz.

Tabelle 8. Analyse der Veränderungssensitivität: Relative Effizienz der NHP-Subskalen gegenüber Vergleichskriterien¹.

tionsaufenthalts befragten Patienten der Stichprobe 8a (internistische Krankheiten) und 8b (muskulo-skeletale Krankheiten) durchgeführt. Hierzu wurden die in den Vergleichskriterien Beschwerdenliste (BL), Schmerzintensität (NRS), Depressivität (CES-D) und Funktionseinschränkung (FFbH-R, HAQ) beobachtbaren Veränderungen zu den in den NHP-Werten aufgetretenen Veränderungen in Beziehung gesetzt (Tabelle 8). Eine relative Effizienz von 1 oder darüber besagt,

dass Veränderungen auf den NHP-Skalen mindestens genau so gut abgebildet wurden, wie bei den Vergleichskriterien – Werte der relativen Effizienz von kleiner als 1 sind in umgekehrter Weise zu interpretieren. Da für die Subskalen *Schlaf* und *Soziale Isolation* keine direkten Vergleichsvariablen erhoben wurden, sind diese in der Tabelle nicht aufgeführt. Die NHP-Skalen *Energie*, *Schmerz* und *Emotionale Reaktion* erreichten mit Werten zwischen 0,63 und 0,78 eine etwas geringere Sensi-

vität als ihre Vergleichskriterien. Auffällig ist zum einen der vergleichsweise geringe Abfall der Energie-Skala (drei Items) gegenüber der mit 24 Items eher umfangreichen BL-Skala, zum zweiten die hohe Sensitivität der aus nur einem Item bestehenden Schmerzintensitätsskala (NRS). Gegenüber den beiden Instrumenten zur Erfassung funktioneller Einschränkungen (FFbH-R, HAQ) erwies sich die NHP-Skala *Physische Mobilität* als geringfügig sensitiver.

Diskussion

Mit dieser Arbeit haben wir die autorisierte deutschsprachige Adaptation des NHP nachgezeichnet und eine differenzierte Beschreibung der im Verlauf der Validierung an verschiedenen Stichproben erhobenen Befunde vorgelegt. Die Ergebnisse unserer Analysen zeigten, dass in den einbezogenen Stichproben in einem Bereich von jungen gesunden Probanden bis hin zu klinisch in relevantem Umfang beeinträchtigten Patienten befriedigende bis gute psychometrische Eigenschaften darstellbar sind. Die Masszahlen zur Beurteilung der Reliabilität lassen das Instrument als für Gruppenvergleiche hinreichend zuverlässig erscheinen, eine Verwendung des NHP in der Individualdiagnostik ist jedoch nur mit Einschränkungen möglich. Unsere Befunde belegen eine zufriedenstellende Konstrukt- und kriterienbezogene Validität des Fragebogens.

Aus dem einheitlichen Format der Fragen und Antwortkategorien ergibt sich eine sehr gute Praktikabilität und Akzeptanz, die eine Anwendung auch in Kollektiven ermöglicht, die einer differenzierten Befragung weniger gut zugänglich sind. Die gemeinsame Applikation des NHP mit anderen Fragebögen bereitet im allgemeinen keine Schwierigkeiten.

Der Fragebogen erscheint aufgrund seiner Messcharakteristik bei Patienten mit mittel- bis schwergradiger Beeinträchtigung deutlich besser einsetzbar als in Kollektiven mit nur geringfügigen Gesundheitseinschränkungen: Personen aus der zuletzt genannten Gruppe erreichen auf den NHP-Subskalen häufig Werte, die am unteren Skalenende (keine Beeinträchtigung) konzentriert sind. Die Differenzierungsfähigkeit des Instruments kann deshalb bei solchen Befragten eingeschränkt sein, daneben können sich „Bodeneffekte“ in der

Veränderungsmessung bemerkbar machen. Als Screeninginstrument z.B. in bevölkerungsbezogenen Studien ist das NHP dann geeignet, wenn mittel- bis schwergradige Beeinträchtigungen der subjektiven Gesundheit identifiziert werden sollen.

Trotz seines Status als „generisches Messverfahren“ ist vor der Anwendung des NHP in einer konkreten Studie zu beachten, dass mit dem Fragebogen ein begrenzter Ausschnitt aus den vielfältigen Dimensionen der gesundheitsbezogenen Lebensqualität erfasst wird und dass selbst die im NHP repräsentierten Dimensionen mit einer erkennbaren Akzentsetzung erhoben werden. Während z.B. Aspekte der Mobilitätseinschränkung, des Schmerzes und der Schlafstörungen in ausreichender Breite vertreten sind, werden die Bereiche der Vitalitätseinschränkung oder der beeinträchtigten Sozialbeziehungen in geringerem Umfang bzw. mit einer Akzentuierung eher gravierender Probleme einbezogen.

Die autorisierte deutsche Übersetzung des NHP wurde oder wird gegenwärtig in Deutschland, Österreich und der Schweiz in einer Reihe von Studien eingesetzt, deren Spektrum von der Analyse der Komorbidität psychiatrischer Patienten bis hin zur Evaluation der Behandlung auf einer Intensivstation reicht. Besonders umfangreiche Erfahrungen liegen bislang bei Patienten mit muskulo-skeletalen Krankheiten vor^{17,18}. Laufende Vorhaben verwenden das NHP in grossen Stichproben u. a. bei der Bewertungsanalyse abdominaler Transplantationsprogramme¹⁹ oder im Rahmen der Entwicklung eines standardisierten Instrumentariums zur Messung des erwerbsbezogenen Leistungsvermögens in der Rehabilitation²⁰. Die aus diesen Studien bisher verfügbaren Ergebnisse zeigen, dass der Fragebogen in diesen Kollektiven mit ähnlichem Erfolg wie in den hier dar-

gestellten Studien einsetzbar ist. Das standardisierte und methodisch nach einheitlichen Gesichtspunkten kontrollierte Vorgehen bei der Übersetzung macht das NHP besonders für internationale Studien interessant, die Bekanntheit und Verfügbarkeit des NHP in anderen Ländern dürfte zudem die Kommunikation von Forschungsergebnissen im internationalen Rahmen erleichtern.

Summary

The German version of the Nottingham Health Profile (NHP): Method of translation and psychometric validation

The Nottingham Health Profile was developed in the late 70s in Great Britain and has since then been widely used in anglo-american countries as well as in other countries. We report about the translation process and the examination of psychometric properties of the German version of the Nottingham Health Profile in 10 samples of over 1000 healthy and ill persons. Results of psychometric assessment (feasibility, reliability, validity, sensitivity) were satisfactory. This suggests that the German translation of the Nottingham Health Profile is a reliable and valid instrument for patient-based reports of subjective health which can be used in clinical research as well as in epidemiologic studies.

Résumé

La version française du Nottingham Health Profile (NHP): Méthode de traduction et validation psychométrique

Le Nottingham Health Profile a été mis au point en Grande-Bretagne vers la fin des années 70 et a été largement utilisé dans les pays anglophones comme instrument de mesure de la santé perçue par le patient lui-même. Il est également de plus en plus utilisé dans d'autres pays depuis le milieu des années 80. L'article décrit la traduction allemande du Nottingham Health Profile ainsi que l'analyse de ses propriétés psychométriques à l'aide de 10 échantillons de plus de 1000 sujets, en bonne santé ou malades. Le contrôle des critères de qualité (praticabilité, fiabilité, validité, sensibilité) s'est déroulé de façon satisfaisante. Au total, le Nottingham Health Profile apparaît comme un instrument praticable, fiable et valide de mesure de la santé perçue et peut être utilisé aussi bien dans des essais cliniques que dans des enquêtes épidémiologiques.

Literaturverzeichnis

1 Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981; 19:787–805.
 2 Kaplan RM, Bush JW. Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychol* 1982; 1:61–80.
 3 DuPuy HJ. The Psychological General Well-being Index. In: Wenger NK, Mattson ME, Furberg CD et

al. eds. Assessment of quality of life in clinical trials of cardiovascular therapies. New York: Le Jacq, 1984:170–183.
 4 Stewart AL, Ware JE, eds. Measuring functioning and well-being. The Medical Outcomes Study Approach. Durham: Duke University Press, 1992.
 5 Hunt SM, McEwen J. The development of a subjective health indicator. *Sociology of Health and Illness* 1980; 2:231–246.
 6 Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived

health status: a validation study. *J Epidemiology and Community Health* 1980; 34:281–286.
 7 Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-national instruments for minimum requirements to optimal models. *Quality of Life Research* 1993; 2:451–459.
 8 Sartorius N. A WHO-method for the assessment of health-related quality of life (WHO-QOL). In: Walker S, Rosser M. eds. Quality of life assessment: key issues in the 1990's. Dordrecht: Kluwer, 1993: 201.
 9 Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I, McKenna SP. Cross-cultural adaptation of health measures. *Health Policy* 1991; 19: 33–44.
 10 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology* 1993; 46:1417–1432.
 11 Hunt SM. A methodology for the translation of health measures. Manchester: Galen Research & Consultancy, 1988.
 12 Bucquet D, ed. European Guide to the Nottingham Health Profile. Montpellier, 1992.
 13 Hays RD, Hayashi T, Carson S, Ware JE. User's Guide for the Multitrait Analysis Program (MAP). Santa Monica: Rand Corporation, 1988.
 14 Deyo RA, Diehr P, Patrick D. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Controlled Clinical Trials* 1991; 12:142S–158S.
 15 Brazier JE, Harper R, Jones NMB et al. Validating the SF-36 health survey questionnaire: a new outcome measure for primary care. *Br Med J* 1992; 305:160–164.
 16 Kohlmann T. Schmerzen in der Lübecker Bevölkerung. *Der Schmerz* 1991; 5:208–213.
 17 Kohlmann T, Herlyn K, Siegrist J. „Lebensqualität“ bei Erkrankun-

- gen der Bewegungsorgane. *Psychomed* 1994; 6:6–11.
- 18 *Kohlmann T, Raspe H, Markus H, Peschel U.* Zur Messung der subjektiven Gesundheit in der Rheumatologie: Die deutsche Version des „Nottingham Health Profile“ (NHP). *Zeitschrift für Rheumatologie (Suppl. 2)* 1992; 51:49.
 - 19 *Dörning H, Grobe T, Schwartz FW.* Veränderung der Lebensqualität durch Nierentransplantation. *Gesundheitswesen* 1995; 57:549–550.
 - 20 *Deck R, Raspe H.* Zur Messung des erwerbsbezogenen Leistungsvermögens bei rheumatischen Erkrankungen: Funktionsbehinderung, Schmerz, schmerzbezogene Kognitionen und Affekte. *Versicherungsmedizin* 1992; 44:214–221.
 - 21 *Bullinger M, Kirchberger I, Ware JE.* Der deutsche SF-36 Health Survey. Übersetzung und psychometrische Testung eines krankheitsübergreifenden Instruments zur Erfassung der gesundheitsbezogenen Lebensqualität. *Zeitschrift für Gesundheitswissenschaften* 1995; 3:21–36.
 - 22 *Bullinger M, Kirchberger I, v. Steinbüchel N.* Der Fragebogen Alltagsleben – ein Verfahren zur Erfassung der gesundheitsbezogenen Lebensqualität. *Zeitschrift für Med Psychol* 1993; 2: 121–131.
 - 23 *Von Zerssen D.* Die Beschwerdenliste. Manual. Weinheim: Beltz Test Gesellschaft, 1976.
 - 24 *Hautzinger M, Bailer M.* Allgemeine Depressionsskala (ADS). Weinheim: Beltz Test Gesellschaft, 1993.
 - 25 *Kohlmann T, Raspe H.* Die patientennahe Diagnostik von Funktionseinschränkungen im Alltag. *Psychomed* 1994; 6:21–27.
 - 26 *Fries JF, Spitz PW, Kraines RG, Holman HR.* Measurement of patient outcome in arthritis. *Arthritis and Rheumatism* 1980; 23: 137–145.

Danksagung

Wir danken unseren Kolleginnen und Kollegen Thomas Abel, Ruth Deck, Marlies Engster, Beverly Geist, Karen Herlyn, Regine Klockgether und Heiner Raspe für die Unterstützung bei der Übersetzung des Fragebogens. Aus der Zusammenarbeit mit der European Group for Quality of Life and Health Measurement, insbesondere mit Sonja Hunt und Steve McKenna, erhielten wir wertvolle Hinweise zur Methodik der Übersetzung. Für hilfreiche und sehr konstruktive Kommentare danken wir dem anonymen Gutachter.

Korrespondenzadresse

Dr. Thomas Kohlmann
 Institut für Sozialmedizin
 Medizinische Universität Lübeck
 St. Jürgen-Ring 66
 D-23564 Lübeck