

**Untersuchung von Response-Shift-Einflüssen auf die Erfassung  
der gesundheitsbezogenen Lebensqualität  
bei Prostatakrebspatienten in der onkologischen Rehabilitation**

**Inaugural-Dissertation  
zur  
Erlangung der Doktorwürde  
des Fachbereichs Psychologie  
der Universität Koblenz-Landau**

vorgelegt von  
Matthias Jelitte  
aus Bad Driburg

Kürnach  
im April 2010

# Danksagung

Ich bedanke mich sehr bei Herrn Prof. Dr. Reinhold Jäger, der sich bereit erklärt hat, mich als Doktorand zu betreuen und die vorliegende Dissertation zu begutachten. Die fachliche Begleitung und Diskussion war für mich eine große Unterstützung bei der Erstellung der Arbeit.

Herzlich danken möchte ich Herrn Prof. Dr. Hermann Faller, der diese Arbeit ebenfalls begutachtet hat. Ich möchte mich bei ihm außerdem bedanken für die persönliche Betreuung, die für mich stets eine hilfreiche Förderung und positive Motivation war.

Diese Arbeit ist nur möglich gewesen, weil mich Menschen mit Krebserkrankungen unterstützt haben. Ich bedanke mich mit großem Respekt dafür, dass sie mir durch die Teilnahme an den Befragungen geholfen haben. In diesem Zusammenhang möchte ich auch Herrn Dr. Thomas Schulte, dem ärztlichen Leiter der Klinik Bad Oexen, danken, ohne dessen Engagement und Offenheit für meine Fragestellungen zum Response-Shift meine Arbeit ebenfalls nicht in dieser Form möglich gewesen wäre.

Besonders bedanken möchte ich mich bei Herrn Dipl.-Psych. Michael Schuler. Die gemeinsamen Besprechungen und Diskussionen über die statistischen Hintergründe dieser Arbeit waren für mich immer sehr gewinnbringend.

Auch Frau Dipl.-Psych. Andrea Reusch und meinem ehemaligen Kollegen Dr. Wilmar Igl habe ich zu danken. Mit ihnen durfte ich ein Projekt zur Veränderungsmessung durchführen. Das war für mich der Einstieg in die Response-Shift-Thematik.

Ansonsten danke ich meinen Kolleginnen und Kollegen am Institut für Psychotherapie und Medizinische Psychologie der Universität Würzburg, die mir formell und informell Rückmeldungen und Anregungen gegeben haben.

Danken möchte ich auch meinen Eltern, Georg und Angelika Jelitte. Die Hilfe, die ich von ihnen bekommen habe, war für mich eine große Entlastung.

Ich danke meiner Frau Susanna Kurz. Sie hat mich während der Erstellung der Arbeit immer unterstützt und mich in vielen Situationen von anderen Aufgaben entlastet. Dann danke ich ihr noch für das Korrekturlesen und die inhaltlichen Anregungen. Außerdem danke ich unseren Kindern Ruben und Neven. Es ist wunderbar, dass ich durch meine Familie jeden Tag erfahre, was wichtiger ist als eine Promotion.

# Zusammenfassung

In der onkologischen Rehabilitation ist gesundheitsbezogene Lebensqualität (GLQ) ein wichtiges Outcome-Kriterium zur Bewertung von Behandlungsmaßnahmen. Mittlerweile steht eine Vielzahl an Instrumenten zur Erfassung der GLQ bei Menschen mit Krebserkrankungen zur Verfügung. Ein häufig eingesetztes Verfahren im Rahmen der onkologischen Rehabilitation in Deutschland ist der EORTC-QLQ-C30, ein krebspezifisches Lebensqualitätsinstrument. Um Veränderungen in der GLQ bei Patienten mit onkologischer Rehabilitation zu erfassen, werden in der Regel zu Beginn und am Ende der Rehabilitation beziehungsweise zu einem Katamnesezeitpunkt Messungen der GLQ vorgenommen. Aus der Differenz der Skalenwerte werden Aussagen über Veränderungen abgeleitet.

Kritisch ist dieser Ansatz der Veränderungsmessung anzusehen, wenn sich die Bewertungsgrundlagen der befragten Person für die Einschätzung der Fragen zur GLQ zu beiden Messzeitpunkten (MZPen) unterscheiden. Im Extremfall ist davon auszugehen, dass zwar zu beiden MZPen Angaben von Personen vorliegen, jedoch die Grundlagen für die subjektive Beurteilung der Fragebogenitems nicht miteinander übereinstimmen. Die Veränderung des Bewertungshintergrunds eines internen Konstrukts, wie beispielsweise der GLQ, aufgrund eines kritischen Ereignisses, wird nach Sprangers und Schwartz (1999) als *Response-Shift* bezeichnet. Drei verschiedene Response-Shift-Prozesse werden unterschieden: Die *Rekalibrierung* stellt eine Veränderung der internen subjektiven Metrik zur Beurteilung eines Konstrukts beziehungsweise von dessen Komponenten dar. Die *Repriorisierung* bedeutet eine quantitative Änderung der Relevanz einzelner Komponenten für ein Konstrukt. Die *Neukonzeptualisierung* umfasst die qualitative Neudefinition eines Konstrukts. Response-Shift tritt vor allem dann auf, wenn sich Menschen mit bedrohlichen oder kritischen Lebensereignissen auseinandersetzen (Sprangers 2002). Die Mitteilung einer Diagnose oder der Beginn einer Behandlung im Rahmen einer schwerwiegenden, chronischen oder auch lebensbedrohlichen Erkrankung kann solch ein kritisches Ereignis sein. Weiterhin definieren Sprangers und Schwartz (1999) in ihrem Modell verschiedene Einflussfaktoren auf einen Response-Shift wie Persönlichkeitsmerkmale und Bewältigungsstrategien.

Es existieren verschiedene Strategien zur Erfassung von Response-Shift. Zu unterscheiden sind individuelle, Design- und statistische Ansätze. Individuelle Ansätze sind darauf ausgerichtet, die für eine Person relevanten Dimensionen oder Komponenten der Lebensqualität

mittels Exploration oder Fragebogen zu erfassen. Die persönliche Zusammensetzung der Lebensqualität sowie die subjektive Bedeutung dieser Komponenten für die Lebensqualität werden als Ausgangspunkt zur Erfassung von Response-Shift verwendet. Zu einem späteren Messzeitpunkt wird diese Befragung wiederholt. Ändern sich die persönlich relevanten Lebensqualitätsdimensionen oder deren Bedeutung für die Lebensqualität der Person, wird dies als Response-Shift bezeichnet. Beispiele für dieses Vorgehen sind die Schedule for Evaluation of Individual Quality of Life (SEIQoL, O'Boyle et al. 2000) und ein Ansatz mit Hilfe des Patient Generated Index (Ruta et al. 1994). Am häufigsten wird in der Literatur auf den so genannten *Then-Test* zurückgegriffen, der zu den Design-Ansätzen gezählt wird. Dabei wird zum zweiten MZP eine *retrospektive* Einschätzung der GLQ zum MZP 1 erhoben. Die Differenz zwischen dem tatsächlichen Prä-Wert und dem retrospektiv erfassten Prä-Wert wird als Hinweis auf eine Rekalibrierung angesehen. Kritisch an diesem Vorgehen ist neben dem ökonomischen Aufwand die Verzerrung der Retrospektivbeurteilung durch fehlerbehaftete Erinnerungen oder implizite Veränderungstheorien (Norman 2003). Im Bereich der statistischen Ansätze hat Oort (2004) eine Strategie zur Response-Shift-Erfassung mittels Konfirmatorischer Faktorenanalyse (KFA) vorgeschlagen. Der Autor geht von einem *Linear Latent Variable Model* (LLVM) aus und definiert die Änderung eines gesamten Messmodells zwischen zwei MZPen als Neukonzeptualisierung. Die Beibehaltung des Messmodells bei gleichzeitiger Veränderung der Faktorladungen der einzelnen manifesten Variablen auf dem latenten Konstrukt operationalisiert eine Repriorisierung. Sind lediglich die Restriktionen der Intercepts eines Indikators zwischen beiden MZPen aufzuheben, wird dies als uniforme Rekalibrierung bezeichnet. Die Invarianz von Messfehlern eines Indikators stellt eine *non-uniforme* Rekalibrierung dar. Diese Form der Rekalibrierung kann so stark ausgeprägt sein, dass sie die Intercepts des Indikators beeinflusst und diese ebenfalls freizusetzen sind. Die aus diesen statistischen Definitionen abgeleiteten Invarianzhypothesen sind mit Hilfe der KFA und Modifikationsindizes überprüfbar. Dieses statistische Vorgehen zur Response-Shift-Erfassung wurde in der vorliegenden Arbeit angewendet.

Es wurde untersucht, ob Response-Shift in einer Stichprobe von  $N = 212$  Prostatakrebspatienten (PCa-Patienten) in einem Zeitraum von drei Monaten nach Beginn der Anschlussheilbehandlung in einer onkologischen Rehabilitationsklinik auftritt und, wenn dies der Fall sein sollte, welche spezifischen Response-Shift-Prozesse stattfinden (Fragestellung 1, Hauptfragestellung). Weiter wurde der Frage nach der Bedeutung der Veränderung der Globalen Lebensqualität im genannten Zeitraum für einen Response-Shift nachgegangen (Fragestellung

2, Nebenfragestellung). Außerdem wurden vier potenzielle Moderatorvariablen aus dem Response-Shift-Modell nach Sprangers und Schwartz (1999) bezüglich ihres Einflusses auf einen Response-Shift überprüft (Fragestellung 3, Nebenfragestellung). Diese sind Benefit Finding, Selbstwirksamkeit, positive und negative Affektivität. Die GLQ wurde mit Hilfe des EORTC-QLQ-C30 erfasst. In das Messmodell der GLQ wurden die Indikatorvariablen Physische, Emotionale, Kognitive, Soziale und Rollenfunktionsfähigkeit aufgenommen.

Die Auswertungen zur Beantwortung der Fragestellung 1 wurden anhand der von Oort (2005a) beschriebenen Modelltestung vorgenommen. In einem ersten Schritt wurde die Passung eines Basismodells getestet, das sich aus den Messmodellen der GLQ zu beiden MZPen zusammensetzt. Das theoretisch abgeleitete Modell zeigte in der vorliegenden Stichprobe insgesamt eine gute Passung mit der empirischen Kovarianzmatrix ( $\chi^2/df = 1.50$ , RMSEA = 0.05, CFI = 0.98, TLI = 0.97). Anschließend wurden alle für die Überprüfung eines Response-Shifts notwendigen Parameterrestriktionen vorgenommen: Die Ausprägungen der Faktorladungen, Intercepts und Messfehlervarianzen der manifesten Variablen wurden restringiert, indem die Werte dieser Parameter einer manifesten Variablen zu beiden MZPen gleichgesetzt wurden. Das so berechnete ‚Nullmodell‘ zeigte eine inakzeptable Passung mit den empirischen Daten der Stichprobe ( $\chi^2/df = 5.30$ , RMSEA = 0.14, CFI = 0.76, TLI = 0.71). Aufgrund dieses Ergebnisses war davon auszugehen, dass in der vorliegenden Stichprobe zwischen beiden MZPen ein Response-Shift stattgefunden hat. In einem weiteren Schritt wurde mit Hilfe von Modifikationsindizes ein Modell, ausgehend vom ‚Nullmodell‘, abgeleitet, in dem alle Parameter freigesetzt wurden, auf denen ein Response-Shift stattgefunden hat. Die Passung dieses ‚Response-Shift-Modells‘ war gut ( $\chi^2/df = 1.31$ , RMSEA = 0.04, CFI = 0.99, TLI = 0.98). In der vorliegenden Stichprobe sind im ‚Response-Shift-Modell‘ die Messfehlervarianzen aller Indikatoren und die Intercepts der Skalen Physische, Kognitive und Rollenfunktionsfähigkeit freizusetzen. Daraus folgt, dass in der vorliegenden Stichprobe ein Response-Shift in Form einer non-uniformen Rekalibrierung aufgetreten ist. Eine reine uniforme Rekalibrierung, eine Repriorisierung und eine Neukonzeptualisierung traten in der Gesamtstichprobe nicht auf.

Die GLQ verbesserte sich zwischen den beiden MZPen statistisch signifikant (Mittelwert der GLQ zum MZP 1 = 0.00, zum MZP 2 = 0.56;  $p < .001$ ). Nach der von Oort (2005a) beschriebenen Vorgehensweise zur Berechnung des Response-Shift-Anteils an beobachteten Effekten konnte auf der Skala Rollenfunktionsfähigkeit ein Response-Shift-Effekt in der Höhe

eines mittleren Effektes nach Cohen (1988) nachgewiesen werden. Auf der Skala Physische Funktionsfähigkeit war der nachgewiesene Response-Shift-Effekt klein. Dies gilt ebenso für die Skala Kognitive Funktionsfähigkeit. Auf allen drei Skalen zeigten sich kleine Effekte durch die Veränderung der GLQ. Der Response-Shift auf den Skalen Physische und Rollen-funktionsfähigkeit führte zu einer ‚Überschätzung‘ der Veränderung durch die Verbesserung der GLQ. Auf der Skala Kognitive Funktionsfähigkeit hingegen ‚überdeckte‘ der Response-Shift eine Veränderung durch die GLQ.

Die Ergebnisse der Fragestellung 2 wurden ermittelt, indem die Gesamtstichprobe von 212 PCa-Patienten anhand des Medians der Veränderung auf der EORTC-QLQ-C30-Skala *Globale Lebensqualität* in zwei gleichgroße Subgruppen unterteilt wurde. Für jede Gruppe wurde eine separate Response-Shift-Ableitung durchgeführt, um die gruppenspezifischen Response-Shift-Prozesse zu identifizieren und zu vergleichen. Als wesentliches Ergebnis konnte gezeigt werden, dass die Gruppe mit deutlichen Veränderungen in der Globalen Lebensqualität mehr Parameterfreisetzungen, mehr Response-Shift-Prozesse und auf der Skala Rollen-funktionsfähigkeit einen ausgeprägteren Response-Shift-Effekt aufwies als die Vergleichsgruppe mit geringer Veränderung der Globalen Lebensqualität.

Die Beantwortung der Fragestellung 3 erbrachte Hinweise auf den Einfluss potenzieller Moderatorvariablen. Zur Überprüfung der Moderatorvariablen Benefit Finding, Selbstwirksamkeit, positive und negative Affektivität wurden ebenfalls Subgruppenanalysen durchgeführt, die am Median des jeweiligen Moderators gebildet wurden. Bei Personen mit einem höheren Ausmaß an Benefit Finding, Selbstwirksamkeitserwartung und positiver Affektivität lassen sich mehr Parameterfreisetzungen und Response-Shift-Prozesse nachweisen als in den entsprechenden Vergleichsgruppen. Jedoch weisen die berechneten Response-Shift-Effekte bei diesen drei Moderatorvariablen eher keine Unterschiede zwischen den Subgruppen auf. Die Analyse der negativen Affektivität als Moderatorvariablen zeigte in beiden Gruppen gleich viele Response-Shift-Prozesse. In der Gruppe der Personen mit einem höheren Ausmaß an negativer Affektivität wurde auf der Skala Rollen-funktionsfähigkeit ein größerer Response-Shift-Effekt im Ausmaß eines kleinen Effekts als in der Vergleichsgruppe berechnet. Hierzu ist anzumerken, dass in der vorliegenden Studie keine Extremgruppen miteinander verglichen wurden, also die Unterschiede zwischen den verglichenen Subgruppen inhaltlich teilweise eher gering waren.

Die Ergebnisse der Arbeit weisen auf das Auftreten von Response-Shift bei PCa-Patienten in der onkologischen Rehabilitation im Rahmen einer Anschlussheilbehandlung hin. Dieses Ergebnis sollte hinsichtlich der Durchführung zukünftiger Evaluationsstudien Berücksichtigung finden, da Response-Shift die Kalkulation von Effekten beeinflussen kann. Da Response-Shift zu einer Über- oder Unterschätzung von Veränderungen bedingt durch die tatsächliche Veränderung der GLQ beitragen kann, sind die bisherigen Evaluationsergebnisse in diesem Versorgungsbereich aus methodischer Sicht kritisch zu hinterfragen. Die Response-Shift-Erfassung nach der Methode von Oort (2004) bietet einen statistischen Ansatz zu Untersuchung von weiteren, spezifischen Einflussfaktoren auf Response-Shift. Diese können entweder durch Gruppenvergleiche geprüft oder als Variablen in das Gesamtmodell aufgenommen werden. Mit der vorgestellten Methodik ist es möglich, spezifische Interventionen während der onkologischen Rehabilitation von PCa-Patienten genauer zu untersuchen. Hierzu liegen bisher kaum empirische Daten vor (Hergert et al. 2009). Ein Ergebnis der Nebenfragestellungen ist, dass das Ausmaß der Veränderung der Lebensqualität mit Response-Shift in Beziehung zu stehen scheint. Daher sollte dieser auch bei unterschiedlichen Strategien der Ergebnismessung in der onkologischen Rehabilitation, zum Beispiel bei der Zielorientierten Ergebnismessung (ZOE, Gerdes 1998), geprüft werden. Auch bei Auswertungen nach veränderungshomogenen Personengruppen ist ein Response-Shift zu erwarten, wenn Gruppen oder Skalen mit deutlichen Veränderungen selektiv ausgewertet werden. Fragen zur Allgemeinen Modellentwicklung von Response-Shift und Lebensqualität sollten weiter verfolgt werden. Die Ergebnisse der Nebenfragestellungen liefern einige Hinweise auf den Einfluss von Moderatorvariablen. Hier zeigt sich jedoch noch enormer Forschungsbedarf. Der statistische Ansatz der Messung von Response-Shift und Veränderungen durch die GLQ ist ebenfalls im Rahmen der Erfassung von Änderungssensitivität von Lebensqualitätsinstrumenten einsetzbar. Die Anwendung der KFA-Methode zur Response-Shift-Erfassung in anderen Forschungskontexten mit Veränderungsmessungen erscheint empfehlenswert. Vor allem die Psychotherapieforschung bietet gute Möglichkeiten, methodische und inhaltliche Aspekte des Phänomens zu untersuchen.

Einige methodische Einschränkungen liegen der Studie zu Grunde. Die Moderatorvariablen wurden erst zum MZP 2 erhoben. Obwohl der hohe Rücklauf nicht dafür spricht, kann aufgrund der zusätzlichen Fragebogenbelastung der befragten Personen ein Selektionseffekt provoziert worden sein, der die Generalisierbarkeit der Ergebnisse in Frage stellt. Daran gebunden ist der Hinweis, dass auch die Angaben zu soziodemographischen Merkmalen der

Studienteilnehmer erst zum MZP 2 erfasst wurden und somit nur eine eingeschränkte Dropoutanalyse durchgeführt wurde, die jedoch auf keine besonderen Unterschiede zwischen den Teilnehmern der Auswertungen und den Nichtteilnehmern hinwies. Außerdem konnte der Einfluss des objektiven Erkrankungszustands nicht kontrolliert werden. Als zusätzliche Erklärung für die non-uniforme Rekalibrierung in der Gesamtstichprobe kann eine Regression zur Mitte nicht ausgeschlossen werden. Jedoch ist nicht anzunehmen, dass diese für das Ausmaß aller Parameterveränderungen verantwortlich zu machen ist.



# Inhaltsverzeichnis

<b>1. Einleitung</b> .....	<b>1</b>
1.1 Hintergrund und Fragestellung.....	1
1.2 Überblick über die vorliegende Arbeit.....	5
<b>A. Stand der Forschung</b> .....	<b>9</b>
<b>2. Lebensqualität und Gesundheit</b> .....	<b>10</b>
2.1 Konzepte und Modelle der gesundheitsbezogenen Lebensqualität .....	12
2.2 Gesundheitsbezogene Lebensqualität und Krebs .....	14
2.3 Modelle der gesundheitsbezogenen Lebensqualität bei Krebserkrankungen .....	16
2.4 Messung gesundheitsbezogener Lebensqualität bei Krebspatienten.....	22
2.4.1 Generische Instrumente zur Messung gesundheitsbezogener Lebensqualität .....	23
2.4.2 Krebspezifische und krankheitsspezifische Instrumente zur Messung gesundheitsbezogener Lebensqualität.....	24
2.5 Prostatakarzinom und gesundheitsbezogene Lebensqualität.....	26
2.5.1 Krebserkrankungen allgemein .....	27
2.5.2 Prostatakarzinom .....	29
2.5.3 Empirische Befunde zu Prostatakrebs und Lebensqualität .....	31
2.6. Onkologische Rehabilitation bei Patienten mit Prostatakrebs.....	39
2.7 Zusammenfassung .....	42
<b>3. Response-Shift</b> .....	<b>44</b>
3.1 Response-Shift – Definition und Modelle .....	45
3.1.1 Response-Shift nach Howard und Kollegen .....	46
3.1.2 Alpha, Beta und Gamma Change nach Golembiewski.....	47
3.1.3 Response-Shift nach Sprangers und Schwartz.....	48
3.1.4 Erweiterung des Modells nach Sprangers durch Rapkin und Schwartz .....	57
3.1.5 Buffer-Modell nach Lepore und Eton.....	58
3.1.6 Dem Response-Shift nahestehende Konzepte und Konstrukte.....	61
3.2 Methoden zur Response-Shift-Erfassung .....	74
3.2.1 Individuell-orientierte Ansätze .....	74
3.2.2 Design-Ansätze .....	78
3.2.3 Statistische Ansätze .....	80
3.2.4 Exkurs: Veränderungsmessung und Response-Shift .....	99
3.2.5 Bewertung der Methoden zur Response-Shift-Erfassung .....	107
3.3 Empirische Befunde zum Response-Shift .....	111
3.3.1 Studien zum Vergleich methodischer Ansätze zur Erfassung von Response-Shift.....	113
3.3.2 Response-Shift bei Krebserkrankungen.....	115

3.4 Zusammenfassung .....	119
<b>4. Response-Shift und Veränderungsmessung von gesundheitsbezogener Lebensqualität bei Prostatakrebspatienten in der onkologischen Rehabilitation .....</b>	<b>121</b>
<b>5. Fragestellungen der vorliegenden Arbeit.....</b>	<b>130</b>
<b>B. Empirischer Teil.....</b>	<b>144</b>
<b>6. Methoden .....</b>	<b>145</b>
6.1 Allgemeines Vorgehen .....	145
6.1.1 Verwendung statistischer Programme .....	146
6.1.2 Bildung verschiedener Subgruppen und Ausreißeranalyse .....	146
6.1.3 Beschreibung der Stichprobe zur Modellableitung.....	148
6.1.4 Beschreibung der Stichprobe zur Beantwortung der Hauptfragestellung.....	149
6.1.5 Ersetzen fehlender Werte .....	151
6.2 Beschreibung der Messinstrumente.....	152
6.2.1 Prüfung auf Normalverteilung .....	157
6.2.2 Skalenüberprüfung.....	158
6.3 Vorgehensweise der Modellprüfung .....	159
6.3.1 Allgemeine Voraussetzungen zur Ableitung eines Messmodells .....	160
6.3.2 Prüfung eines Messmodells .....	162
6.3.3 Testung von Diskrepanzfunktion und Goodness-of-Fit-Indizes .....	163
6.3.4 Ableitung eines Messmodells .....	166
<b>7. Ergebnisse .....</b>	<b>170</b>
7.1 Ergebnisse zur Fragestellung 1 .....	170
7.1.1 Prüfung des Basismodells .....	170
7.1.2 Prüfung des ‚Null-Modells‘ .....	173
7.1.3 Ableitung eines Response-Shift-Modells .....	175
7.1.4 Response-Shift und Effektstärkeschätzungen in der Gesamtstichprobe .....	182
7.1.5 Zusammenfassung der Ergebnisse der Fragestellung 1 .....	185
7.2 Ergebnisse zur Fragestellung 2.....	187
7.2.1 Dichotomisierung der Gesamtstichprobe.....	188
7.2.2 Stichprobenbeschreibung.....	190
7.2.3 Überprüfung der Modellpassung in beiden Gruppen.....	192
7.2.4 Beziehung zwischen der Veränderung der Globalen Lebensqualität und Response-Shift .....	195
7.2.5 Zusammenfassung der Ergebnisse zur Fragestellung 2 .....	202
7.3 Ergebnisse zur Fragestellung 3 .....	203
7.3.1 Beschreibung der Substichproben zur Auswertung der Fragestellung 3 .....	204
7.3.2 Überprüfung der Modellpassung in den einzelnen Substichproben .....	211

7.3.3 Beziehung zwischen Moderatorvariablen und Response-Shift.....	212
7.3.4. Zusammenfassung der Ergebnisse der Fragestellung 3 .....	229
<b>8. Diskussion .....</b>	<b>231</b>
8.1 Response-Shift bei PCa-Patienten in der onkologischen Rehabilitation – Fragestellung 1 ...	233
8.1.1 Kurzbeschreibung der Ergebnisse der Fragestellung 1 .....	234
8.1.2 Diskussion der Ergebnisse der Fragestellung 1 .....	238
8.2 Einfluss der Veränderung der Lebensqualität auf Response-Shift – Fragestellung 2 .....	253
8.2.1 Kurzbeschreibung der Ergebnisse der Fragestellung 2.....	253
8.2.2 Diskussion der Ergebnisse der Fragestellung 2 .....	255
8.3 Einfluss von Benefit Finding, Selbstwirksamkeit, positiver und negativer Affektivität auf Response-Shift – Fragestellung 3 .....	259
8.3.1 Benefit Finding .....	259
8.3.2 Selbstwirksamkeit.....	261
8.3.3 Positive Affektivität.....	262
8.3.4 Negative Affektivität .....	264
8.4 Methodische Einschränkungen der Fragestellungen 2 und 3 .....	266
8.5 Kritische Betrachtung allgemeiner methodischer Aspekte der Studie .....	267
<b>9. Perspektiven.....</b>	<b>273</b>
9.1 Response-Shift im Rahmen von Anpassungsprozessen .....	273
9.2 Response Shift als methodisches Konzept .....	279
<b>Literaturverzeichnis.....</b>	<b>286</b>
<b>Tabellenverzeichnis .....</b>	<b>308</b>
<b>Abbildungsverzeichnis .....</b>	<b>311</b>
<b>Anhangsverzeichnis.....</b>	<b>312</b>

# Abkürzungsverzeichnis

Abkürzung	Erläuterung
<i>Statistische Begriffe</i>	
KFA	Konfirmatorische Faktorenanalyse
x, X	Messwert eines Individuums, Vektor einer beobachteten Variable
i	Untersuchung, Individuum
j	Variable, Phänomen
k	Beobachtung
M, Mean	Mittelwert
SD	Standardabweichung
Var	Varianz
$M_{\text{Diff}}$	Differenz der Mittelwerte
Min	Minimum
Max	Maximum
Md	Median
n, N	Stichprobenumfang (N = Gesamtstichprobe, n = Teilstichprobe)
DO	Dropout
p	Probability, Wahrscheinlichkeit
*	$p < 0.05$ (zweiseitig)
**	$p < 0.01$ (zweiseitig)
n.s.	nicht statistisch signifikant
CI	Konfidenzintervall
t, T	Messzeitpunkt, wahrer Wert (engl. true value, true score), Test
IN	Intercept
e, E	Messfehler
FL	Faktorladung
$\chi^2$	Chi-Quadrat
df	degrees of freedom (Freiheitsgrade)
RMSEA	Root Mean Square Error of Approximation
CFI	Comparative Fit Index
TLI	Tucker-Lewis-Index
AIC	Akaike Information Criterion
ECVI	Expected Cross Validation Index
GRI	Guyatt's Responsiveness Index
ES, SES	Standardized Effect Size
SRM	Standardized Response Mean
d	Effektstärke
r	Pearson-Korrelations-Koeffizient
COV	Kovarianz
RCT	Randomized Control Trial
KTT	Klassische Testtheorie
IRT	Item-Response-Theorie
LLVM	Linear Latent Variable Model
<i>Fragebögen</i>	
EORTC-QLQ-C30	European Organization for Research and Treatment of Cancer – Quality of Life Questionnaire – Core 30
PF	Physische Funktionsfähigkeit
RF	Rollenfunktionsfähigkeit
SF	Soziale Funktionsfähigkeit
EF	Emotionale Funktionsfähigkeit
CF	Kognitive Funktionsfähigkeit
PANAS	Positive and Negative Affect Schedule

Fortsetzung nächste Seite

<b>Abkürzung</b>	<b>Erläuterung</b>
PA	Positive Affektivität
NA	Negative Affektivität
SW	Selbstwirksamkeit, Allgemeine Selbstwirksamkeitserwartung
BF	Benefit Finding, Benefit Finding Skala
Soz-Dat	Soziodemographische Daten
<b>Medizinische Begriffe</b>	
AHB	Anschlussheilbehandlung
PCa	Prostatakarzinom
RP	Radikale Prostatektomie
<b>Griechisches Alphabet</b>	
alpha, $\alpha$	Mittelwert des latenten Konstrukts, Interne Konsistenz (Cronbachs $\alpha$ )
beta, $\beta$	Mittelwert eines spezifischen Faktors
delta, $\delta$	Messfehlervarianz
epsilon, $\epsilon$	Wert / Vektor eines (latenten) Fehlerwertes eines Faktors
eta, $\eta$	Wert / Vektor eines (latenten) spezifischen (uniquen) Wertes eines Faktors
kappa, $\kappa$	Mittelwert eines latenten Faktors
lambda, $\lambda$	Faktorladung, Regressionskoeffizient
my, $\mu$	Mittelwert der manifesten Variable
sigma, $\sigma$	Varianz
tau, $t$	Intercept, Achsenschnittpunkt
xi, $\xi$	Wert / Vektor eines latenten Konstrukts
Gamma, $\Gamma$	Regressionskoeffizient
Phi, $\Phi$	Varianz des latenten Faktors, Varianz-Kovarianzmatrix der latenten Konstrukte (COV ( $\xi, \xi'$ ))
Psi, $\Psi$	Varianz-Kovarianzmatrix der spezifischen Faktoren (COV ( $\eta, \eta'$ ))
Sigma, $\Sigma$	Gesamte Varianz-Kovarianzmatrix
Theta, $\Theta$	Varianz-Kovarianzmatrix der Messfehler (COV ( $\epsilon, \epsilon'$ ))
<b>Sonstiges</b>	
MZP	Messzeitpunkt
GLQ	Gesundheitsbezogene Lebensqualität
LQ	Lebensqualität

**Anmerkungen zu den Abkürzungen:** Auflistung erfolgt nach inhaltlichen Kriterien. Abkürzungen wurden in der vorliegenden Arbeit nicht immer einheitlich verwendet, da in unterschiedlichem Kontext z. T. gleiche Abkürzungen eingesetzt werden, z. B.  $\alpha$  für den Mittelwert eines latenten Konstrukts und als Maß der Internen Konsistenz.

#### **Anmerkung zum Ausdruck und zur Rechtschreibung**

Um das Lesen der Arbeit nicht durch ständige Anführungszeichen für die Kennzeichnung der Skalenbezeichnungen unnötig kompliziert zu machen, sind die Adjektive der Skalennamen als Teil des Namens großgeschrieben (z.B. Physische Funktionsfähigkeit). Die englischsprachigen Begriffe sind in der Regel kleingeschrieben, jedoch wurden diejenigen, die auch in der deutschen Fachsprache einen gewissen Bekanntheitsgrad als feststehende Konzepte haben, großgeschrieben (z.B. Missing für fehlende Daten).

# 1. Einleitung

## 1.1 Hintergrund und Fragestellung

Der rehabilitativen Versorgung von Patienten mit einer onkologischen Grunderkrankung kommt innerhalb des Deutschen Rehabilitationswesens eine Besonderheit zu: Die Deutsche Rentenversicherung ist bei der Bewilligung von Maßnahmen der onkologischen Rehabilitation in erster Linie an der Verbesserung der Lebensqualität der betroffenen Personen orientiert und nicht am Erfolgskriterium der Besserung oder Wiederherstellung der Erwerbsfähigkeit (Koch 2000). Lebensqualität, oder häufig synonym verwendet die *gesundheitsbezogene Lebensqualität* (health-related quality of life, GLQ), stellt somit ein wichtiges Kriterium für die Entscheidung über die Bewilligung einer Maßnahme der onkologischen Rehabilitation dar und ist eines der wesentlichsten Evaluationskriterien für Behandlungsmaßnahmen und -verläufe in diesem Versorgungsbereich.

Die European Organization for Research and Treatment of Cancer (EORTC) hat es sich seit ihrer Gründung im Jahre 1962 unter anderem zur Aufgabe gemacht, Instrumente zur Erfassung der GLQ bei Krebspatienten zu entwickeln und deren Anwendung in der Praxis zu fördern. Mittlerweile existiert eine Vielfalt an Instrumenten zur Erfassung der Lebensqualität, die grob in drei verschiedene Kategorien eingeteilt werden: Die *generischen Lebensqualitätsinstrumente* sind unabhängig von einer konkreten Grunderkrankung einer Person anwendbar und ermöglichen einen Vergleich der Lebensqualität von Personen mit unterschiedlichen Grunderkrankungen oder von Erkrankten mit Gesunden. Für die Behandlung von Krebspatienten existieren weiterhin die sogenannten *krebsspezifischen Lebensqualitätsinstrumente*, die bei Patienten mit unterschiedlichen Krebserkrankungen anwendbar sind, jedoch nicht bei Betroffenen anderer Grunderkrankungen. Mit Verfahren aus dieser Instrumentengruppe können Vergleiche der Lebensqualität bei Personen mit unterschiedlichen Krebserkrankungen angestellt werden. Als dritte Möglichkeit gibt es *krankheitsspezifische Lebensqualitätsinstrumente*, die sich konkret auf charakteristische Beeinträchtigungen und Belastungen bei einzelnen Krebserkrankungen beziehen. Die Auswahl eines bestimmten Fragebogens aus einer der drei Kategorien ist abhängig vom jeweiligen Untersuchungsziel.

Hinsichtlich der Evaluation von Behandlungsmaßnahmen hat sich der Grundsatz etabliert, die Anwendung eines krankheitsspezifischen mit entweder einem krebsspezifischen oder

einem generischen Lebensqualitätsinstrumentes zu kombinieren. Auf diese Art ist einerseits eine Aussage über die Veränderung der für eine einzelne onkologische Grunderkrankung typischen Krankheits- beziehungsweise Behandlungsfolgen möglich und andererseits eine Einschätzung der Ergebnisse im Vergleich mit Personen mit anderen Erkrankungen oder der gesunden Allgemeinbevölkerung. Die verschiedenen Vergleiche können zu unterschiedlichen Bewertungen führen. So kann ein kleiner Effekt als Resultat einer Evaluation anhand eines krankheitsspezifischen Lebensqualitätsinstrumentes ein befriedigendes Ergebnis für einen Behandler sein. Wenn dann jedoch bei einem krebsspezifischen Instrument heraus kommt, dass der ermittelte Effekt deutlich kleiner ist als bei anderen Krebserkrankungen, relativiert dieses Ergebnis möglicherweise die zuvor vorgenommene positive Bewertung. Je nach Vergleichsmaßstab verändert sich die Bewertung. Ein Aspekt der für die vorliegende Arbeit von besonderer Relevanz ist.

Vielfach wurde diskutiert, auf welche Weise GLQ am besten zu erfassen ist. Es gab Ansätze, diese von außenstehenden Personen quasi objektiv beurteilen zu lassen. Der Arzt oder die Pflegekraft sollte die Lebensqualität einer kranken Person bewerten. Die Unklarheit über die zugrunde liegenden Bewertungskriterien führte oftmals zu dem Ergebnis, dass Nichtbetroffene oder Behandler deutlich negativere Einschätzungen lieferten als die Betroffenen selbst. Diese „paradoxen Befunde“ nach Herschbach (2002) konnten nicht nur für den Vergleich von Lebensqualitätsbeurteilungen zwischen Gesunden und Kranken gemacht werden. Es fiel insgesamt auf, dass in vielen Studien zum Thema selbst massiv Erkrankte ihre Lebensqualität eher positiv beurteilten, oftmals sogar besser als Personen mit gut behandelbaren Akuterkrankungen. Studienergebnisse solcher Art legen zwei unterschiedliche Schlüsse nahe: Zum Einen scheinen Menschen in der Lage zu sein, selbst in für die eigene Gesundheit und die weitere Lebensperspektive sehr kritischen Situationen eine positive Perspektive zu entwickeln. Ein Befund der aus unterschiedlichen Forschungszweigen Bestätigung findet, insbesondere der Bewältigungsforschung mit ihren vielfältigen Konzepten zum Umgang mit Lebensbedingungen, in denen sich die betroffenen Personen mit widrigen Gegebenheiten auseinandersetzen müssen und dennoch ein inneres Gleichgewicht im Sinne einer positiven Haltung nicht verlieren. Zum Anderen deuten die Ergebnisse darauf hin, dass GLQ ein subjektives Konstrukt ist, dessen Bewertung von außen als kaum möglich anzusehen ist. Deswegen ist der zentrale Zugang zur Erfassung der GLQ die Selbstbeurteilung des Betroffenen.

Insbesondere in der onkologischen Rehabilitation zeigt sich somit die Bedeutung der Patientenorientierung und des patientennahen Outcome-Kriteriums für die Bewertung einer Behandlungsmaßnahme sowie eines Behandlungserfolgs. Dieser Bezug auf die subjektive Wahrnehmung und das subjektive Erleben des Patienten in der onkologischen Rehabilitation gewichtet stark das individuelle Bewertungssystem des Einzelnen. Der Nachteil daran ist wiederum, dass von außen nicht wirklich erkennbar ist, wie das individuelle Bewertungssystem für das Konstrukt GLQ gestaltet ist. Zwar haben sich mittlerweile einige Kerndimensionen der GLQ in der Wissenschaft etabliert. Hierzu zählt beispielweise die Beurteilung der Lebensqualität anhand der vier Dimensionen *Körperliche Gesundheit*, *Psychische Befindlichkeit*, *Soziale Beziehungen* und *Alltagskompetenzen*. Ob diese Dimensionen jedoch bei jeder einzelnen Person in der immer gleichen Weise für die Beurteilung der GLQ herangezogen werden, ist ein viel diskutiertes Thema in der Untersuchung zu Veränderungen der Lebensqualität.

Ein Konzept, das diesen Sachverhalt zu beschreiben versucht, ist der sogenannte *Response-Shift*. Dieser wird allgemein als Veränderung des subjektiven Bewertungshintergrunds aufgrund eines besonderen Ereignisses verstanden und ist seit ca. 10 Jahren in der Lebensqualitätsforschung ein zunehmend häufiger beschriebenes und untersuchtes Phänomen. Die Bedeutung von Response-Shift für die wissenschaftliche Untersuchung von Lebensqualität im Rahmen der onkologischen Rehabilitation in Deutschland im Speziellen und der Evaluationsforschung im Allgemeinen ist darin zu sehen, dass es die Validität einer quantitativen Aussage über eine Differenz zwischen zwei Messwerten in Frage stellt, da im Extremfall möglicherweise zwei qualitativ unterschiedliche Zustände miteinander in Beziehung gesetzt werden. Für die GLQ wird das Phänomen als relevant angesehen, da es sich bei dieser um ein sogenanntes ‚latentes Konstrukt‘ handelt, also eine Variable, die nicht direkt erfassbar ist. Weiterhin wird GLQ überwiegend durch Selbsteinschätzungen anhand persönlicher Bewertungen und nicht aufgrund objektivierbarer Leistungen erfasst. Wenn Response-Shift für die Erfassung der Lebensqualität ein Einflussfaktor sein sollte, sind die bisherige Art der Veränderungsmessung und der Effektkalkulationen in Frage zu stellen, was weiterhin bedeuten würde, dass die bisher berichteten Effekte – beispielsweise von Maßnahmen der onkologischen Rehabilitation, die anhand von Mittelwertsdifferenzen von Lebensqualitätsinstrumenten erfasst und festgestellt wurden – möglicherweise fehlerhaft sind.

Mittlerweile liegen zwar zahlreiche Untersuchungen zum Response-Shift vor, die den Einfluss des Phänomens auf die Veränderungsmessung bestätigen. Neben der Frage der ge-



nauen inhaltlichen Bedeutung von Response-Shift für die Erfassung der Lebensqualität ist jedoch auch die nach der am besten geeigneten Erfassungsmethode bisher ungeklärt. Es existieren verschiedene Ansätze beziehungsweise Gruppen von Strategien, wobei der prominenteste Ansatz die Verwendung des so genannten retrospektiven Prätest (engl. Thentest) ist, bei dem anstelle des tatsächlichen Prätests zum Zeitpunkt der Postmessung eine retrospektive Einschätzung des Ausgangszustands erhoben und mit dem Postwert in Beziehung gesetzt wird. Aus der Differenz dieser beiden Einschätzungen wird dann wiederum ein Maß für die Veränderung ermittelt, von dem angenommen wird, dass dieser nicht durch einen Response-Shift beeinflusst ist. Das Vorgehen ist jedoch nicht unumstritten, so dass sich weitere Ansätze zur Response-Shift-Erfassung entwickelt haben. Das statistische Vorgehen der Invarianztestung von Parametern in Messmodellen über mehrere Messzeitpunkte anhand der Konfirmatorischen Faktorenanalyse ist eine weitere Möglichkeit, um Response-Shift zu erfassen. Der besondere Vorteil dieser Strategie ist darin zusehen, dass es möglich ist verschiedene Response-Shift-Prozesse zu differenzieren, ohne dass zusätzlicher Messaufwand entsteht. Ausreichend ist eine herkömmliche Prä-Post-Messung. Der wesentlichste Nachteil ist die Tatsache, dass mittels der KFA nur Aussagen über Gruppen, nicht jedoch über individuelle Verläufe abgeleitet werden können.

International kommt Response-Shift seit der ersten ausführlichen Darstellung und Diskussion aufgrund des Werkes „Adaption to Changing Health – Response Shift in Quality-of-Life Research“ von Sprangers und Schwartz (2000) eine zunehmende Bedeutung in der Lebensqualitätsforschung zu. Eine erste Metaanalyse zum Thema wurde 2006 von Schwartz und Kollegen publiziert. Die Aussage der Metaanalyse lassen zwar keine eindeutigen Aussagen über einen Response-Shift zu, allerdings ist die Qualität der verwendeten Studien sehr heterogen. In Deutschland gibt es dagegen bisher nur sehr wenige Studien zum Thema. Auf dem Gebiet der Rehabilitationsforschung sind es kaum mehr als eine handvoll Arbeiten, die sich explizit mit Response-Shift beschäftigen. Die dabei angewandte Strategie der Veränderungsmessung ist das hierzulande als quasi-indirekte Veränderungsmessung bezeichnete Vorgehen unter Verwendung des retrospektiven Prätests. Studien zur Response-Shift-Erfassung mittels Konfirmatorischer Faktorenanalyse liegen in Deutschland bisher nicht vor. Diese Lücke versucht die vorliegende Arbeit zu schließen.

Ziel der Arbeit ist die Untersuchung der Fragestellung, ob Response-Shift als Einflussfaktor auf die Erfassung der GLQ bei einer Stichprobe von PCa-Patienten in der onkologi-

schen Rehabilitation auftritt, und wenn ja, in welcher Form. Als Untersuchungsansatz wird die Invarianztestung mit Hilfe der Konfirmatorischen Faktorenanalyse gewählt. Weiterhin werden bei Identifikation von Response-Shift in der untersuchten Stichprobe Einflussfaktoren auf das Phänomen untersucht. Diese sind aus einem Modell nach Sprangers und Schwartz (2000) abgeleitet und umfassen das Ausmaß der Veränderung der Lebensqualität sowie spezifische Einflussfaktoren wie Benefit Finding, Selbstwirksamkeit, positive und negative Affektivität.

## **1.2 Überblick über die vorliegende Arbeit**

Der erste Teil *A. Stand der Forschung* umfasst die Darstellung des theoretischen Hintergrunds mit den Schwerpunkten GLQ bei Krebserkrankungen und Response-Shift, Ableitung eines spezifischen Wirkmodells von Response-Shift unter Einbezug der untersuchten Variablen sowie die Beschreibung der Fragestellungen der vorliegenden Arbeit.

Zu Beginn werden die Grundlagen zum Konstrukt Lebensqualität und Gesundheit vermittelt (Kapitel 2). Dieses Kapitel ist unterteilt in verschiedene Abschnitte, wobei zuerst allgemein auf Konzepte und Modelle der GLQ eingegangen wird (Kapitel 2.1). Anschließend wird die Thematik für Krebserkrankungen spezifiziert (Kapitel 2.2). Im weiteren Verlauf der Arbeit werden explizit Modelle der GLQ bei Krebserkrankungen beschrieben, wobei eines der in der Literatur dargestellten Modelle die Basis für das in der vorliegenden Arbeit verwendeten Modells liefert (Kapitel 2.3). Die Messung der GLQ allgemein sowie bei Personen mit Krebserkrankungen ist Thema des darauf folgenden Kapitels (2.4), wobei schwerpunktmäßig auf die Einteilung in generische, krebs- und krankheitsspezifische Fragebögen eingegangen wird. Die Stichprobe der vorliegenden Arbeit bildet eine Gruppe von Männern mit Prostatakrebs, so dass auch auf diese besondere Krebserkrankung genauer eingegangen wird (Kapitel 2.5). Neben der allgemeinen Beschreibung dieser Form der Krebserkrankung werden empirische Befunde zur Lebensqualität bei Personen mit Prostatakrebs und deren Behandlung beschrieben. Es folgt eine kurze praktische Darstellung der Grundlagen, Ziele und des Ablaufs der onkologischen Rehabilitation im stationären Kontext der Rehabilitation, da in diesem Rahmen die Untersuchung der Lebensqualität der vorliegenden Stichprobe stattgefunden hat (Kapitel 2.6). Das Kapitel zum Thema Lebensqualität und Gesundheit endet mit einer kurzen Zusammenfassung (Kapitel 2.7).

Response-Shift ist das Thema des dritten Kapitels. Die Darstellung beginnt mit der Vorstellung der bisher publizierten Modelle und Konzepte zum Response-Shift (Kapitel 3.1). Dabei wird ebenfalls auf die Entwicklung der Forschung zum Thema Bezug genommen. Der vorliegenden Arbeit liegt ein Modell von Response-Shift nach Sprangers und Schwartz (2000) zugrunde. Es liefert die Basis einerseits für die Differenzierung von Response-Shift in Rekalibrierung, Repriorisierung und Neukonzeptualisierung. Diesen Definitionen wird im weiteren Verlauf der Arbeit gefolgt. Andererseits entwickeln die Autorinnen ein allgemeines Konzept zur Integration von Response-Shift in den Verlauf der Auseinandersetzung mit einem Ereignis (im vorliegenden Fall einer Erkrankung) und dessen Einfluss auf die wahrgenommene Lebensqualität. Dieses allgemeine Modell liefert auch die Grundlagen für die Ableitung der Fragestellungen 2 und 3, die sich auf Einflussfaktoren auf Response-Shift-Prozesse beziehen. Weiterhin enthält das Kapitel eine Beschreibung von Konzepten beziehungsweise Konstrukten, die für die inhaltliche Beschreibung und Erklärung von Response-Shift bedeutsam sind. Dazu zählen Selbstregulationsprozesse, Coping und Benefit Finding. Der Abschnitt endet mit der Darstellung eines integrativen Modells der Anpassung an chronische Erkrankungen, das die zuvor beschriebenen Konzepte und Response-Shift gemeinsam berücksichtigt. Im anschließenden Abschnitt werden die Methoden zur Response-Shift-Erfassung vorgestellt und bewertet (Kapitel 3.2). Das Kapitel enthält Darstellungen über individuell-orientierte Erfassungsstrategien, Design- und statistische Ansätze. Die zuletzt genannte Vorgehensweise wird ausführlich vor dem Hintergrund der Invarianzmessung beschrieben, die für die in der vorliegenden Arbeit verwendete Methode der Konfirmatorischen Faktorenanalyse relevant ist. In diesem Kontext wird auch das Linear Latent Variable Model nach Oort (2004) vorgestellt, das die (mess-)theoretische Grundlage für die Response-Shift-Erfassung mittels KFA liefert. Aufgrund der praktischen Relevanz des Response-Shifts für die Ansätze der Veränderungsmessung wird auf dieses Thema ausführlich in Form eines Exkurses eingegangen. In diesem Zusammenhang werden Methoden der Veränderungsmessung in Form der direkten, indirekten und quasi-indirekten Veränderungsmessung beschrieben. Anschließend wird ein Bezug hergestellt zwischen der klinischen Bedeutsamkeit von Effekten und Response-Shift, da Response-Shift herkömmlich kalkulierte Effekte beeinflussen kann. Aufgrund dessen wird zusätzlich noch Änderungssensitivität als wesentliches Gütekriterium im Rahmen von Evaluationsstudien vorgestellt und in den Response-Shift-Kontext gesetzt. Im anschließenden Kapitel (3.3) werden die bisher publizierten Befunde zum Response-Shift dargestellt. Es wird einerseits auf Studien, die verschiedene Response-Shift-Erfassungsstrategien vergleichen, sowie andererseits auf empirische Arbeiten zum Response-Shift bei Menschen mit Krebserkrankun-

gen eingegangen. Mittlerweile liegen zu vielen verschiedenen Grunderkrankungen Response-Shift-Untersuchungen vor. Studien an Personen mit onkologischen Grunderkrankungen wurden jedoch vergleichsweise häufig durchgeführt. Eine Zusammenfassung zum Thema Response-Shift (Kapitel 3.4) rundet das Kapitel ab.

Ein spezifisches Wirkmodell zur Beschreibung der in der vorliegenden Arbeit untersuchten Variablen und Prozesse im Rahmen der onkologischen Rehabilitation bei Patienten mit Prostatakarzinom wird als Zusammenfassung der vorher dargestellten Inhalte präsentiert (Kapitel 4).

Es folgt die Ableitung der Fragestellungen (Kapitel 5), wodurch der theoretische Teil der Arbeit beschlossen wird. Die Beantwortung der Fragestellungen erfordern ein gestuftes Vorgehen: Zuerst wird geprüft, ob in der vorliegenden Untersuchungsstichprobe ein Response-Shift aufgetreten ist und, falls ja, in welcher Form. Anschließend wird bei entsprechender Beantwortung der gerade beschriebenen Fragestellung untersucht, inwieweit sich die Differenz der Lebensqualitätsveränderung auf einen Response-Shift auswirkt. Abschließend soll der Einfluss der Variablen Benefit Finding, Selbstwirksamkeit sowie positive und negative Affektivität auf einen Response-Shift geprüft werden.

Der zweite Teil der Arbeit *B. Empirischer Teil* beinhaltet die Darstellung der methodischen Grundlagen und Durchführung der Untersuchung sowie die ausführliche Beschreibung der Ergebnisse. Diese werden im Anschluss diskutiert. Aus der Diskussion lassen sich Forschungsfragen als zukünftige Perspektiven aus der kritischen Ergebnisbetrachtung ableiten.

Es werden methodischen Grundlagen der vorliegenden Arbeit beschrieben (Kapitel 6). Dazu gehören die Beschreibung des allgemeinen Vorgehens der Datenerhebung, die Verwendung statistischer Verfahren, der Stichprobenziehung sowie das Ersetzen fehlender Werte (Kapitel 6.1). Im anschließenden Abschnitt werden die Messinstrumente zur Erfassung der GLQ sowie der Einflussfaktoren auf einen potenziellen Response-Shift vorgestellt (Kapitel 6.2). Darauf folgen die Darstellung der Vorgehensweise der Modellprüfung und die Ableitung eines Messmodells zur Bearbeitung der vorliegenden Fragestellungen.

Es folgt die Ergebnisdarstellung zu den verschiedenen Fragestellungen (Kapitel 7). Die Vorstellung der Ergebnisse beginnt mit der allgemeinen Überprüfung, ob ein Response-Shift

in der vorliegenden Stichprobe nachzuweisen ist (Kapitel 7.1). Das Vorgehen der Invarianztestung wird ausführlich anhand der Überprüfung der verschiedenen Modelle beschrieben. Abschließend wird der Einfluss des Response-Shifts auf die kalkulierten Effekte nach herkömmlicher Effektberechnung geprüft. Im folgenden Abschnitt werden die Ergebnisse zur Untersuchung des Ausmaßes der Lebensqualitätsveränderung auf den Response-Shift vorgestellt (Kapitel 7.2). Für die Gesamtstichprobe wurde für jede Person ein Differenzwert aufgrund der Prä-Post-Angaben auf der Skala *Globale Lebensqualität* des EORTC-QLQ-C30 berechnet und die Gesamtstichprobe anhand dieser Differenzwerte dichotomisiert. Für beide Gruppen (große und geringe Differenzwerte in der Veränderung der Globalen Lebensqualität) wurden separate Response-Shift-Ableitungen durchgeführt. Unterschiede zwischen den Subgruppen werden beschrieben. Das Vorgehen zur Untersuchung des Einflusses der Variablen Benefit Finding, Selbstwirksamkeit, positive und negative Affektivität entspricht dem gerade beschriebenen Zugang (Kapitel 7.3). Pro Einflussvariable wurde die Gesamtstichprobe in zwei Subgruppen entsprechend des Medians des Differenzwertes eingeteilt und separat auf Response-Shift getestet. Die Unterschiede zwischen den Gruppen werden ebenfalls dargestellt.

Die Ergebnisse der Fragestellungen werden anschließend diskutiert (Kapitel 8). Vor dem Hintergrund der Ergebnisse der Fragestellung 1 wird auf methodische Aspekte des Ansatzes nach Oort (2004) zur Response-Shift-Erfassung, die inhaltliche Interpretation des Response-Shifts in der untersuchten Stichprobe, die Bedeutung des Response-Shifts für die Effektkalkulation und die Veränderungsmessung sowie für die Wirksamkeitserfassung von Behandlungen in der onkologischen Rehabilitation von PCa-Patienten eingegangen (Kapitel 8.1). Aussagen zur Modellbildung nach Sprangers und Schwartz werden anhand der Ergebnisse der Fragestellungen 2 und 3 diskutiert (Kapitel 8.2, 8.3). Abschließend werden die methodischen Begrenzungen der vorliegenden Arbeit beschrieben (Kapitel 8.4, 8.5).

Über die Ergebnisse der vorliegenden Arbeit hinausgehende Forschungsfragen werden als Perspektiven für zukünftige Response-Shift-Studien vorgestellt (Kapitel 9). Das Kapitel ist unterteilt in inhaltliche (Kapitel 9.1) und methodische Aspekte (Kapitel 9.2).

## **A. Stand der Forschung**

## 2. Lebensqualität und Gesundheit

Die wissenschaftliche Auseinandersetzung mit dem Konzept Lebensqualität begann in den 1960er Jahren vorwiegend im Bereich der sozialwissenschaftlichen Wohlfahrtsforschung. Lebensqualität galt als Maß für die Übereinstimmung zwischen objektiven Lebensbedingungen und deren subjektiven Bewertungen durch die dort lebenden Personen (Glatzer & Zapf 1984). Der weitere Verlauf der Lebensqualitätsforschung lässt sich in vier verschiedene Phasen einteilen (Bullinger 2000): (1) Die erste Phase dauert an bis in die frühen 1980er Jahre und zeichnete sich durch Ansätze der Definition und Konzeptionalisierung des Konstrukts Lebensqualität aus. (2) Anschließend wurden bis in die frühen 1990er Jahre hinein methodische Ansätze zur Messung von Lebensqualität entwickelt und diskutiert. (3) Daraufhin fand das Konstrukt Eingang in verschiedenste klinische Bereiche der medizinischen Versorgung bis in die 2000er Jahre und etablierte sich als Kriterium zur Beurteilung von Behandlungsmaßnahmen. (4) Seit einigen Jahren wird Lebensqualität vor einem methodologischen Hintergrund thematisiert. Es geht dabei wiederum um die Grundlagen der Messbarkeit des Konstrukts. Außerdem ist Lebensqualität im Bereich der Qualitätssicherung klinischer Einrichtungen und der Verteilung von Ressourcen im Gesundheitswesen zu einem bedeutsamen Kriterium geworden.

Aktuell existiert keine einheitliche Definition für Lebensqualität. Referenzpunkte zur Bestimmung des Konstrukts bieten allgemeine Beschreibungen, deren Überprüfbarkeit bisweilen unklar bis unmöglich ist. Ausgangspunkt der Diskussion stellt häufig der Ansatz der World Health Organisation von 1948 dar, welcher Gesundheit nicht als Abwesenheit von Krankheit, sondern als körperliches, psychisches und soziales Wohlbefinden beschreibt. In einer aktualisierten Umschreibung des Konzepts Lebensqualität der WHO heißt es:

„Quality of life is defined as an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person’s physical health, psychological state, level of independence, social relationships, and their relationship to salient features of their environment” (WHOQOL Group 1993, S. 153).

Lebensqualität bezieht sich demnach auf das *subjektive Erleben* von Personen in ihrem gesellschaftlichen und kulturellen Kontext, wobei verschiedene *Komponenten* bedeutsam sind: die körperliche Gesundheit, das psychische Erleben, die Funktionsfähigkeit im Alltag, soziale Beziehungen und Umweltaspekte. Die medizinische Perspektive richtet ihren Fokus in erster Linie auf die ersten vier Aspekte der Lebensqualität und berücksichtigt weniger den kulturellen Rahmen. Jedoch ist mit Ferrans (2005) davon auszugehen, dass, wenn Personen krank werden, und dies gilt in besonderem Maße für chronische Erkrankungen, häufig alle möglichen Lebensbereiche davon betroffen sind.

Nach Spilker (1996) lassen sich drei verschiedene Ebenen der Lebensqualität voneinander abgrenzen und in einem Modell integrieren. Als übergeordnetes Maß für Lebensqualität steht in dem Modell die *allgemeine subjektive Befindlichkeit* einer Person. Dieses Allgemeinmaß ist an keine wesentlichen Inhalte gebunden und absolut unspezifisch. Auf einer darunter gelegenen zweiten Ebene lassen sich allgemeine *Komponenten der Lebensqualität* differenzieren. Dabei handelt es sich um verschiedene Bereiche der Lebenssituation von Personen wie beispielsweise die körperliche Gesundheit, die psychische Befindlichkeit, soziale und ökonomische Aspekte, aber auch Bereiche wie Religiosität und Spiritualität. Auf einer dritten Ebene findet eine *Konkretisierung der Komponenten* statt. Zur körperlichen Gesundheit zählen beispielsweise vorhandene Erkrankungen oder körperliche Einschränkungen, zur psychischen Befindlichkeit entsprechend Emotionalität, Verarbeitungs- und Bewältigungsprozesse etc. Auf dieser Ebene werden somit einzelne Merkmale der verschiedenen Komponenten beschrieben.

Diese kurz skizzierten Annahmen zur Beschreibung und Definition von Lebensqualität liefern den Rahmen für die Ableitung dessen, was unter GLQ zu verstehen ist. Darauf wird im Kapitel 2.1 eingegangen. Auf die Bedeutung von Krebserkrankungen für die GLQ wird im Kapitel 2.2 Bezug genommen. Anschließend werden im Kapitel 2.3 konkrete Modelle zur GLQ bei Menschen mit Krebserkrankungen vorgestellt. Im darauffolgenden Kapitel 2.4 werden Grundlagen der Messung gesundheitsbezogener Lebensqualität bei Krebspatienten überblicksartig präsentiert. Anschließend wird die Erkrankung Prostatakrebs beschrieben und in den Kontext der GLQ gesetzt. Daran schließt sich in Kapitel 2.6 eine Beschreibung der Rehabilitation bei PCa-Patienten an, bevor im Kapitel 2.7 nochmals die wesentlichen Informationen zusammengefasst werden.



## 2.1 Konzepte und Modelle der gesundheitsbezogenen Lebensqualität

Subjektive Gesundheit beziehungsweise GLQ ist ein Teilbereich der allgemeinen Lebensqualität, der sich auf Aspekte des Befindens und Handelns bezieht, die mit körperlichen Einschränkungen oder (chronischen) Krankheiten zusammenhängen (Bullinger 2000). Die Entwicklung der Forschung zur GLQ ist eng mit der Entwicklung medizinischer Standards und der Epidemiologie chronischer und lebensbedrohlicher Erkrankungen verbunden. Durch die Verbesserung der medizinischen Versorgung ist es gelungen, vielen Akuterkrankungen – wie beispielsweise Infektionskrankheiten – ihren lebensbedrohlichen Charakter zu nehmen sowie die Überlebenszeit bei vielen chronischen Krankheiten zum Teil deutlich zu steigern. Andererseits verschiebt sich durch eine verbesserte medizinische und hygienische Grundversorgung und eine damit verbundene ‚Alterung‘ der Gesellschaft das Krankheitspektrum stärker hin zu chronischen Erkrankungen (Rose 2003). Diese sind je nach individueller Lebenssituation und Krankheitsstadium beziehungsweise -verlauf mit zum Teil erheblichen Einschränkungen und Anforderungen für die betroffene Person verbunden. GLQ hat in den vergangenen Jahren in der Gesundheitsversorgung in Deutschland an Bedeutung zugenommen, da es bei vielen chronischen Erkrankungen nicht mehr um Heilung oder um eine entscheidende Lebensverlängerung geht, sondern darum, das subjektive Befinden des Patienten zu verbessern oder zu stabilisieren (ebd.).

Genauso wie sich die allgemeine Lebensqualität in ein Modell mit verschiedenen Ebenen untergliedern lässt, trifft dies auch auf die GLQ und ihre Einzelkomponenten zu. Für die Beschreibung des Konstrukts werden im englischen Sprachgebrauch häufig Auswirkungen und Konsequenzen von Behandlungsmaßnahmen durch das medizinische Versorgungssystem in die Definitionen integriert. Ferrans (2005) kategorisiert Modelle der GLQ in drei Gruppen, wobei sich diese vorwiegend nach drei Dimensionen einordnen lassen: die erste Dimension bildet die *Anzahl der Komponenten*, die als lebensqualitätsrelevant angesehen werden. Das Spektrum der Definitionen variiert von lediglich der funktionalen oder psychischen Komponente bis hin zu allen möglichen Lebensbereichen. Die zweite Unterscheidungsdimension bildet die *Perspektive* auf das Konzept. Die Eckpunkte dieser Dimension bilden einerseits Ansätze, die lediglich negativ- beziehungsweise defizitorientiert sind, und andererseits solche Modelle, die positiv beziehungsweise kompetenzorientiert sind. Bei Ersteren handelt es sich um Definitionen, die Einschränkungen und deren Auswirkungen auf die GLQ thematisieren, bei Zweiteren steht das gesundheitsbezogene Wohlbefinden im Vordergrund, das trotz Einbu-

ßen ausgeglichen beziehungsweise positiv erlebt werden kann. Die dritte Dimension bezieht sich auf die *Methode*, wie GLQ erfasst wird. Die Bandbreite möglicher Ansätze wird durch die Pole ‚objektive Beurteilbarkeit‘ und ‚subjektive Bewertung‘ aufgespannt. Wird GLQ anhand funktionaler Leistungsfähigkeit eingeschätzt, ist eine objektive Beurteilung eher gegeben, als wenn eine Person die *Bedeutung* dieser Leistungsfähigkeit für sich selbst bewerten soll.

Trotz der Mannigfaltigkeit an relevanten Aspekten und Lebensbereichen für die Definition der GLQ hat sich eine Operationalisierung bestehend aus vier Komponenten etabliert. Dies sind nach Ware (1984, sowie Maurischat et al. 2004, Cella et al. 1993) *Körperliche Gesundheit*, *Psychisches Befinden*, *Soziale Beziehungen* und *Alltagskompetenz*. Diese Einteilung hat im deutschen Forschungsbereich breiten Konsens gefunden. Unter *Körperlicher Gesundheit* werden Aspekte wie Mobilität, Geschicklichkeit, Beweglichkeit und Einflüsse durch Behandlungsmaßnahmen (Wirkung und Nebenwirkungen) sowie Veränderungen durch die Erkrankungen selbst (Symptome) zusammengefasst. *Psychische Befindlichkeit* beinhaltet Aspekte wie psychische Störungen (Angst, Depression), Bewältigungsfähigkeit, Selbstwirksamkeit und Kontrollüberzeugung. *Soziale Beziehungen* umfassen beispielsweise familiäre und intime Beziehungen, soziale Kontakte, Integration und das soziale Netzwerk. In der *Alltagskompetenz* wird die körperliche Leistungsfähigkeit in Bezug zu Anforderungen und Aufgaben im Alltag gesetzt. Dazu gehören einerseits basale Kompetenzen wie Essen, Waschen oder Ankleiden. Andererseits zählen dazu auch erweiterte Kompetenzen wie das Ausüben einer Erwerbstätigkeit, das Führen eines Haushalts oder die Bewältigung finanzieller Anforderungen. Nicht konkret enthalten sind in dieser Konzeption der GLQ Inhaltsbereiche wie die Spiritualität, Akzeptanz des eigenen Körpers oder Stigmatisierungen, die Spilker (1996) ebenfalls als wichtige Lebensqualitätskomponenten angesehen werden.

Das allgemeine Modell der GLQ bildet den Ausgangspunkt für die Diskussion des Konstrukts im Zusammenhang mit Krebserkrankungen. Im Weiteren wird dieser Zusammenhang genauer beschrieben, um die Ableitung eines Modells vorzubereiten, das für die Operationalisierung von GLQ bei Krebserkrankungen geeignet ist.

## 2.2 Gesundheitsbezogene Lebensqualität und Krebs

Durch die Integration der GLQ in die medizinische Versorgung von Patienten mit Krebserkrankungen vollzog sich ein Paradigmenwechsel, der deutlich die Selbstbestimmung und aktive Rolle des Patienten im Verlauf seiner Erkrankung betont (Aaronson et al. 1993, Bullinger 1994). Nicht mehr nur Mortalität und Morbidität beziehungsweise medizinische Parameter wie Tumorwachstum oder objektive Lebenszeitverlängerung sind Kriterien der Beurteilung von Behandlungen und Grundlagen für Behandlungsentscheidungen, sondern auch das patientenseitige Kriterium der GLQ. Dies trifft umso mehr zu, je schwerer die Krankheit ist und umso belastender die Behandlungen für den Patienten sind.

In der Onkologie ist das Konzept der GLQ seit etwa 25 Jahren ein Thema. Somit ist sie einer der ersten Bereiche der medizinischen Gesundheitsversorgung, in denen das Konstrukt Einzug gehalten hat (Bullinger 2000). Ein Grund dafür ist darin zu sehen, dass es bei Krebspatienten durch eine intensive medizinische Behandlung zu einer Verlängerung des Lebens kommen kann, die dann aber zumeist mit enormen Belastungen durch Nebenwirkungen der Therapie begleitet wird (Aaronson 1991). Schwarz und Kollegen (1995) sprechen in diesem Zusammenhang von einer *lebensqualitätsadaptierten* Therapie. Nach Aaronson und Kollegen (1986) ist das Kriterium Lebensqualität ein Parameter, der in die Kosten-Nutzen-Relationen von unterschiedlichen Behandlungen bei Krebserkrankungen mit einbezogen werden soll. In diesem Zusammenhang verweist Krischke (1995) auf vier Gründe, warum Forschung zur GLQ im Rahmen von Krebserkrankungen von Bedeutung ist:

1. Die medizinische Betreuung von Krebspatienten zielt auf eine Lebensverlängerung oder Leidensverminderung (oder beides). Diese verursacht möglicherweise eine Veränderung der Lebensperspektive der krebserkrankten Person oder geht mit dieser einher (durch die Krankheit bedingt). Die Lebensqualitätsforschung bietet Möglichkeiten, die Veränderung der subjektiven Lebensperspektive transparent und verstehbar zu machen.
2. Nebenwirkungen und chronische Folgeerscheinungen der Krebstherapien müssen erfasst, bewertet und behandelt werden.
3. Bei der großen Anzahl von Krebserkrankungen besteht ein öffentliches und individuelles Interesse an Informationen zur Bewältigung der Krankheit.

4. Es besteht ein Wissensdefizit darüber, welche Hilfeleistungen Krebspatienten generell und im Verlauf der Erkrankung und Behandlung wirklich benötigen.

Zwei Aspekte der Auseinandersetzung mit einer Krebserkrankung durch den Betroffenen werden in dieser Aufzählung besonders betont: Die *Behandlungsmöglichkeiten* von Krebserkrankungen und deren *Bewältigung*. Die Betroffenen setzen sich mit einem lebensbedrohlichen Ereignis, einer potenziell tödlichen Krankheit, die behandelt werden muss, auseinander, woraus resultiert, dass sie in besonderer Weise behandelt, betreut und unterstützt werden müssen. Standardmäßig werden in der Krebstherapie chirurgische oder chemische Maßnahmen beziehungsweise Bestrahlungen eingesetzt. Ziel allein ist nicht, aus den genannten Möglichkeiten die für die Patienten hinsichtlich ihrer spezifischen Erkrankung effizienteste Maßnahme auszuwählen, sondern darüber hinaus Angebote zu schaffen, die Befindlichkeit der Betroffenen positiv zu beeinflussen. Neben der behandlungsbezogenen Perspektive wird gleichzeitig der Bewältigungsbezug betont. Von besonderer Bedeutung ist, wie sich die Betroffenen einerseits mit der Krankheit Krebs bewusst und/oder unbewusst auf kognitiver, emotionaler und handlungsbezogener Ebene auseinandersetzen, und andererseits, welche Anforderungen an die Bewältigungsfähigkeit der Patienten im Rahmen der Therapie gestellt werden, die durch belastende Nebenwirkungen und Begleiterscheinungen gekennzeichnet sein können.

Studien zur Überprüfung der Lebensqualität bei Krebspatienten ergeben oftmals paradoxe Befunde (Albrecht & Devlieger 1999, Herschbach 2002). So zeigte sich beispielsweise der unerwartete Befund, dass schwerkranke Krebspatienten eine bessere Lebenszufriedenheit angaben als Gesunde oder nur leicht Erkrankte. Die Art der Therapie (radikal vs. schonend) scheint ebenfalls in keinem eindeutigen Verhältnis zur Lebensqualität der behandelten Patienten zu stehen. Sugarbaker und Kollegen (1981) kamen zu dem Ergebnis, dass Krebspatienten nach Amputation eine bessere Lebensqualität aufwiesen als organerhaltend behandelte Patienten mit der gleichen Erkrankung. Weitere Untersuchungen zeigen, dass Krebspatienten häufig eine gute Lebensqualität berichten und positive Erfahrung aus der eigenen Erkrankung ziehen (Stamatiadis-Smidt et al. 2006). In diesem Zusammenhang wird die Krankheitsbewältigung als mitentscheidender Prozess diskutiert. Einen Überblick über mehr als 100 Studien zu diesem Thema bietet Faller (2001). Die Funktion und Bedeutung einzelner Krankheitsbewältigungsprozesse wird aufgrund widersprüchlicher Befunde kontrovers diskutiert. Zwar werden in der Literatur vier Bewältigungsstile besonders hervorgehoben: Dazu zählen

Kampfgeist und Verleugnung, von denen angenommen wird, dass diese mit einer längeren Überlebenszeit in Beziehung stehen, und depressive Krankheitsverarbeitung und stoisches Akzeptieren, die mit einer kürzeren Überlebenszeit assoziiert sind. Insgesamt kommt Faller (2001, 2004) jedoch zu dem Schluss, dass eine Vorhersage über den Verlauf einer Krebserkrankung aufgrund spezifischer Bewältigungsstile nicht möglich ist. Hinsichtlich der subjektiven Lebensqualität stellt Faller (2001) fest, dass Hinweise dafür existieren, diese als einen unabhängigen Prädiktor für die Überlebenszeit von Krebspatienten zu bezeichnen, selbst wenn biomedizinische Parameter kontrolliert werden. Für Patientinnen mit metastasierten MammaCa nimmt der Autor beispielsweise an, dass in der subjektiven Beurteilung der eigenen Gesundheit die gesamte körperliche Situation der Betroffenen besser beschrieben wird, als dies spezifische biomedizinische Risikofaktoren, wie zum Beispiel die durch den Arzt beurteilte körperliche Leistungsfähigkeit nach dem Karnofsky-Index oder das Stadium beziehungsweise die Histologie des Tumors, können. Hierin zeigt sich die besondere Bedeutung der subjektiven Bewertung und des subjektiven Erlebens der Erkrankung.

Die paradoxen Befunde zu Krebserkrankungen und GLQ sind für die vorliegende Arbeit von großer Bedeutung. Eine Erklärung für dieses Phänomen ist der sogenannte Response-Shift, auf den im Kapitel 3 ausführlich eingegangen wird. Eine gute GLQ ist in der Behandlung von Krebspatienten zu einem entscheidenden Therapieziel geworden. Vorab soll die Frage nach krebsspezifischen Anpassungen des GLQ-Modells bei entsprechend Erkrankten thematisiert werden.

### **2.3 Modelle der gesundheitsbezogenen Lebensqualität bei Krebserkrankungen**

Die Konzeptualisierung gesundheitsbezogener Lebensqualität bei Krebserkrankungen orientiert sich stark an den allgemeinen Definitionen des Konstrukts. Kontrovers diskutiert wird die Frage, ob bei krebsspezifischen Festlegungen der relevanten Dimensionen der GLQ auch Krankheitssymptome und Nebenwirkungen von Therapien in die Modellbildung aufgenommen werden sollen. Je nach Fragestellung der Studien zeigen sich unterschiedliche Komponenten, die mit GLQ bei Krebspatienten in Beziehung gesetzt werden. Aaronson und Kollegen (1986) postulierten sechs verschiedene Dimensionen der Lebensqualität bei Krebspatienten. Zu diesen zählen Symptome der Krankheit und Nebenwirkungen der Behandlungen (1), der funktionale Status (2), die psychische Befindlichkeit (3), die sozialen Beziehungen (4), die Sexualität und das Körperbild (5) und die Zufriedenheit mit der medizinischen Behandlung (6).

Fayers und Hand (1997, 2002) publizierten ein Modell der GLQ bei Krebspatienten, das sie anhand der Rotterdam Symptom Checkliste (Haes et al. 1990) ableiteten. Dieses Instrument wird häufig in Lebensqualitätsstudien bei Krebserkrankungen eingesetzt und von den Autoren zu den krebspezifischen Lebensqualitätsfragebögen gezählt. Das Modell setzt sich aus vier Faktoren zusammen:

1. Faktor – psychische Befindlichkeit: umfasst verschiedene psychische Erlebensqualitäten wie Depressivität, Ängstlichkeit, Irritiertheit etc.
2. Faktor – Schmerzerleben: Kopf-, Rücken- und Muskelschmerz
3. Faktor – Übelkeit und Erbrechen
4. Faktor – Symptome: Appetitverlust, Fatigue, Mundtrockenheit etc.

Kritisch anzumerken an diesem Modell ist zum Einen die Vernachlässigung der Differenzierung der Funktionsfähigkeiten wie beispielsweise die Körperliche, Soziale oder Emotionale Funktionsfähigkeit. Die im Abschnitt über die Konzepte und Modelle der GLQ beschriebenen Dimensionen werden nicht übernommen. Zum Anderen ist unklar, welche Bedeutung die Faktoren zwei bis vier für die Konzeptualisierung der Lebensqualität von Krebspatienten besitzen. Um diesen Aspekt weiter auszuführen, soll kurz auf die Konstruktionsprinzipien der meisten Lebensqualitätsfragebögen in der Onkologie eingegangen werden. Die methodische Grundlage der Skalenkonstruktion dieser Instrumente ist ein psychometrisch-korrelativer Ansatz, der auch als induktive Konstruktion von Skalen bezeichnet wird (Amelang & Zielinski 1997). Bei diesem Ansatz werden diejenigen Items zu Skalen zusammengefasst, die miteinander hoch korrelieren und damit eine gemeinsame Dimension repräsentieren (Homogenität einer Skala). Im vorliegenden Fall handelt es sich bei den Dimensionen um die einzelnen Faktoren. In der Lebensqualitätsforschung werden die genannten vier Dimensionen Körperliche Gesundheit, Psychisches Befinden, Soziale Beziehungen und Alltagskompetenz als Kerndimensionen der GLQ bezeichnet. Diese Annahme setzt voraus, dass diese vier Dimensionen die GLQ abbilden. Im korrelativen Sinne handelt es sich bei der GLQ um einen Faktor höherer Ordnung, der aus der gemeinsamen Korrelation der vier Kerndimensionen ableitbar ist. Diese Ableitung geschieht wiederum aufgrund statistischer Zusammenhänge zwischen allen vier Faktoren. Im vorliegenden Fall geht es jedoch in erster Linie nicht um die Konstruktion eines Fragebogens, sondern um die korrelativen Grundlagen zur Bildung eines Modells der GLQ bei Krebspatienten. Die korrelative Grundlage beinhaltet die Annahme, dass im Falle einer Veränderung der GLQ sich auf den vier Kerndimensionen ebenfalls Ver-

änderungen ergeben müssen. Umgekehrt gilt, wenn sich auf einer Kerndimension eine Veränderung ergibt, sollte diese auch die GLQ verändern.

Fraglich ist diese Modellannahme jedoch hinsichtlich der Art der Beziehung zwischen der GLQ und den von Fayers und Hand (2002) beschriebenen Faktoren Zwei, Drei und Vier (Schmerzen, Symptome, Behandlungsnebenwirkungen). Besteht eine direkte Beziehung zwischen der Lebensqualität und den genannten Faktoren, so ist davon auszugehen, dass von der Ausprägung eines Faktors auf die GLQ geschlossen werden kann. So sollte gelten, dass wenn eine Person angibt, sehr an Schmerzen zu leiden, die GLQ deutlich negativ ausgeprägt ist. Andersherum besagt der korrelative Ansatz jedoch auch, dass wenn die GLQ deutlich negativ ausgeprägt ist, auf das Vorliegen von (starken) Schmerzen geschlossen werden kann. Die Regelmäßigkeit des zuletzt genannte Zusammenhangs ist jedoch nicht gegeben, wie Fayers und Hand (1997, 2002) selbst darzustellen versuchten. Anhand der Berechnung von Kreuztabellen auf Einzelitemebene konnten die Autoren nachweisen, dass ein zu erwartender deutlicher Zusammenhang zwischen der GLQ und den einzelnen Symptomen, Nebenwirkungen der Behandlungen und Schmerzerleben nicht gegeben war. Zwar ließ sich ein hoher Zusammenhang zwischen einer hohen positiven Lebensqualität und der Abwesenheit entsprechender Symptome, Behandlungsnebenwirkungen sowie Schmerzerleben nachweisen. Eine deutlich negativ ausgeprägte Lebensqualität ging jedoch nicht mit einer hohen Ausprägung der untersuchten Items einher. Es zeigte sich demnach eine ausgeprägte Asynchronität in der Beziehung zwischen GLQ und den genannten Variablen. Dieses Ergebnis veranlasste Fayers und Hand (1997, 2002) zwischen zwei Kategorien von Variablen in der Lebensqualitätsforschung zu unterscheiden: Indikatorvariablen und Kausalvariablen.

*Indikatorvariablen* sind solche Variablen, in denen sich eine latente Variable wie die GLQ ausdrückt beziehungsweise widerspiegelt. Dabei handelt es sich um Marker für die nicht direkt erfassbaren Komponenten der Lebensqualität. Diese stehen zu dieser in einem ungerichteten Verhältnis. Indikatorvariablen ein und derselben Lebensqualitätskomponente sind hoch miteinander korreliert und entsprechen in ihrer Ausprägung dem Ausmaß der zugrunde liegenden Komponente.

*Kausalvariablen* sind Variablen, die einen Einfluss auf eine Zielvariable ausüben, deren Ausprägung jedoch nicht widerspiegeln, sondern diese verursachen. Bezogen auf das latente Konstrukt Lebensqualität bei Krebspatienten unter Chemotherapie bedeutet das, wenn be-

stimmte Therapienebenwirkungen auftreten, dann beeinflussen diese die GLQ der entsprechenden Person. Der Umkehrschluss ist jedoch nicht zulässig: Wenn die Lebensqualität schlecht ist, dann ist das nicht gleichbedeutend mit dem Auftreten von Nebenwirkungen. Kausalvariablen sind nicht unbedingt mit dem latenten Konstrukt korreliert. Wenn sie aber auftreten, ist die Wahrscheinlichkeit sehr hoch, dass sie auf die Lebensqualität einen Einfluss haben.

Die unsystematische Kombination aus Kausal- und Indikatorvariablen ist nach Fayers (2004 sowie Fayers et al. 1997) ein wesentlicher Grund dafür, dass Modelle der GLQ bei Krebspatienten oft nicht repliziert werden können. Die Ausprägung der Kausalvariablen ist in hohem Maße vom Krankheitsverlauf, spezifischen Krebsleiden, spezifischen Krebstherapien und behandlungsbezogenen Nebenwirkungen abhängig. Dies alles sind mögliche Einflussfaktoren auf die Faktorenstruktur und entsprechende Itemkorrelationen, so dass es häufig schwer möglich ist, Aussagen über Lebensqualität bei verschiedenen Subgruppen selbst ein und derselben Krebserkrankung zu generalisieren. Aus diesem Grund wird in der vorliegenden Arbeit ein Modell der GLQ zugrundegelegt, das sich lediglich auf Indikatorvariablen im Sinne von Fayers und Hand (2002) bezieht.

Boehmer und Luszczynska (2006) übertrugen die Einteilung lebensqualitätsrelevanter Variablen in Kausal- und Indikatorvariablen auf einen Fragebogen der Europäischen Organisation für Forschung und Therapie von Krebserkrankungen (EORTC), den EORTC-QLQ-C30, einen krebspezifischen Fragebogen (ausführliche Darstellung des Instruments siehe Kapitel Methoden 6.2). Sie leiten ein Modell ab, das in Abbildung 1 zu sehen ist. Entsprechend der Konstruktion des Fragebogens EORTC-QLQ-C30 wurden fünf Funktionsskalen definiert. Dazu zählen die Komponenten Physische Funktion, Rollenfunktion, Kognitive Funktion, Soziale Funktion und Emotionale Funktion. Diese Skalen bilden die Ausprägung des latenten Konstrukts GLQ ab. Des Weiteren werden die drei Symptomskalen Fatigue, Schmerzen sowie Übelkeit und Erbrechen definiert. Außerdem enthält der Fragebogen sechs Ein-Item-Fragen zu weiteren Symptomen und Nebenwirkungen (Kurzatmigkeit, Schlafstörungen, Appetitmangel, Verstopfung, Durchfall) und Belastungen (finanzielle Probleme).



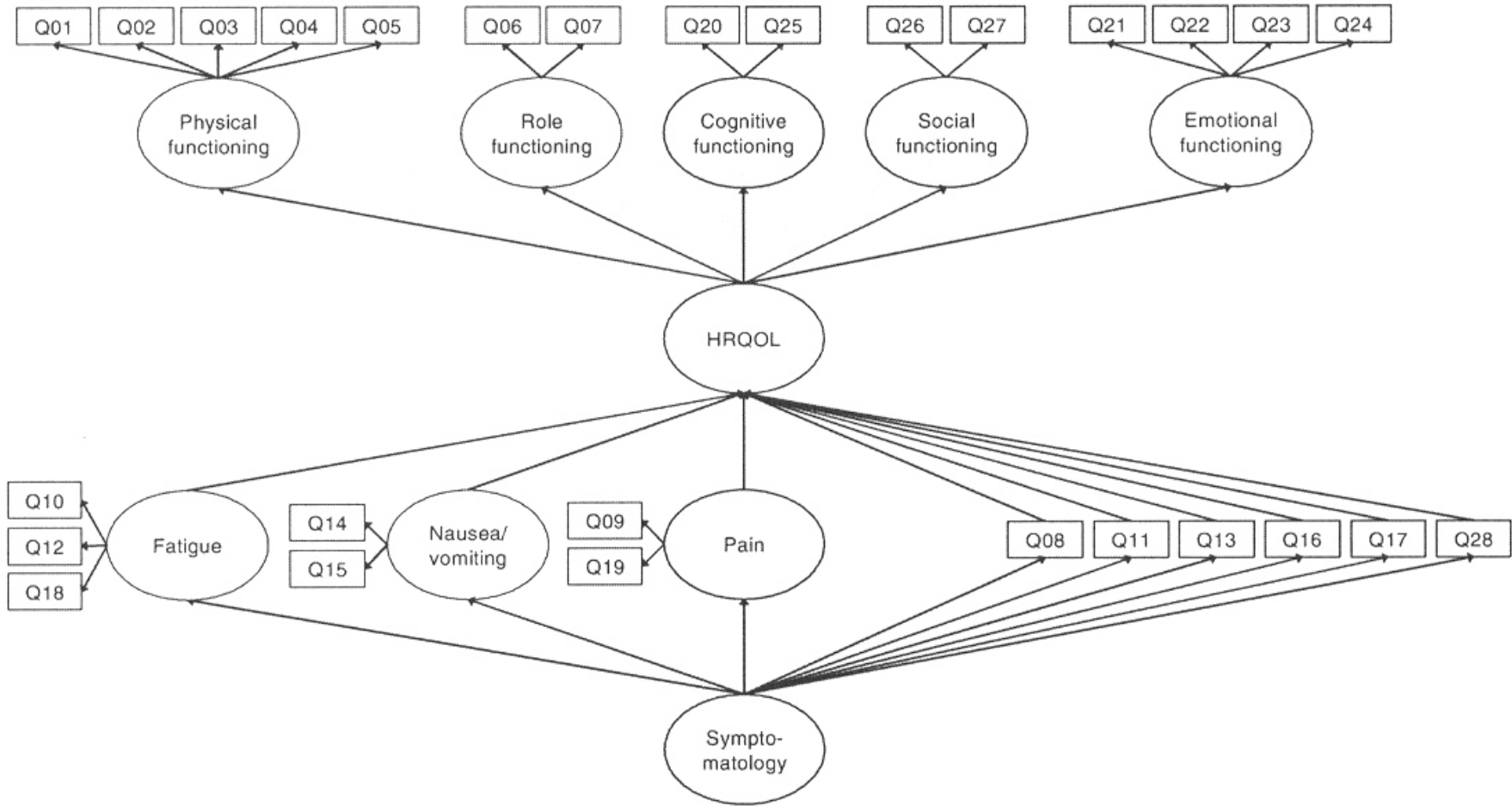


Abbildung 1: GLQ nach Boehmer und Luszczynska (2006) gemessen mit dem krebsspezifischen Lebensqualitätsfragebogen EORTC-QLQ-C30

Das Lebensqualitätsmodell der genannten Autorinnen stellt aufgrund der theoretischen Annahmen ein krebsspezifisches Modell zur GLQ dar und integriert die allgemeinen Dimensionen der Lebensqualität, die mittels generischer Lebensqualitätsfragebögen ermittelt wurden, und krebsspezifische Symptome und Behandlungsnebenwirkungen sowie finanzielle Belastungen. Nach Böhmer und Luszczynska (2006) sind die fünf Funktionsskalen Indikatorvariablen der GLQ bei Krebspatienten und die Symptomskalen sowie Items als Kausalvariablen zu bezeichnen. Die Symptomskalen und -items sowie das Item der finanziellen Belastung sind krebsspezifisch und werden von der Körperlichen Funktion als Indikatorvariable abgegrenzt.

Mittels Konfirmatorischer Faktorenanalyse untersuchten die Autorinnen 205 Personen mit Krebserkrankungen im Gastrointestinalbereich einen Monat nach chirurgischer Behandlung. Die Ergebnisse werten die Autorinnen als Bestätigung für die Unterteilung der Items beziehungsweise Skalen des Fragebogens in Indikator- und Kausalvariablen, obwohl ein alternativ getestetes Modell, in dem alle Items beziehungsweise Skalen als ‚Indikatoren‘ für GLQ definiert wurden, lediglich eine marginal schlechtere Passung aufwies. Die Differenz zwischen den Modellen ist statistisch nicht signifikant. Der Entscheidung von Boehmer und Luszczynska (2006) zur Annahme des Indikator- und Kausalmodells liegt ein Erklärungsansatz nach Bollen (1989) zugrunde, der die inhaltliche Plausibilität der Bedeutung einzelner Items beziehungsweise Skalen anhand eines so genannten *Thought Test* untersucht. Anhand dieses Tests werden Beziehungen zwischen einzelnen Items beziehungsweise Skalen und dem Kriterium *GLQ* auf inhaltliche und logische Konsistenz geprüft. Dies geschieht anhand der Stimmigkeit von einfachen wenn-dann-Aussagen. Die Autorinnen geben dafür ein Beispiel: „... changes of HRQOL have an impact on emotional functioning and thus a change in emotional functioning implies that HRQOL has changed“ (ebd., S.132). Diese Beziehung entspricht der zwischen einer Indikatorvariable und dem latenten Konstrukt. Zur Kausalvariable beschreiben Boehmer und Luszczynska (ebd., S. 132) Folgendes: „ ... HRQOL almost certainly changes if a patient suddenly suffers lengthy and severe vomiting. But if the level of HRQOL changes, it does not necessarily imply that the level of vomiting changes“. Somit geben inhaltliche Überlegungen den Ausschlag für die Akzeptanz des Modells.

Das Modell von Boehmer und Luszczynska (2006) bildet den Ausgangspunkt für die Untersuchung der Fragestellungen der vorliegenden Arbeit, wobei die Unterteilung in Indikator- und Kausalvariablen übernommen wird. Im Zusammenhang mit den Fragestellungen

nach dem Auftreten und den Einflussfaktoren auf einen potenziellen Response-Shift werden ausschließlich die *Indikatorvariablen* auf Response-Shift Einflüsse untersucht.

## **2.4 Messung gesundheitsbezogener Lebensqualität bei Krebspatienten**

Neben den mannigfaltigen Definitions- und Konzeptualisierungsversuchen zur GLQ bei Krebspatienten ist deren Messung ein weiteres wichtiges Thema. Allgemeine Akzeptanz besteht in der Annahme, dass GLQ ein Konstrukt ist, das nur unzureichend durch Fremdbeobachtung eingeschätzt werden kann. Zwar gibt es Hinweise darauf, dass je enger der Kontakt zwischen Zielperson und Beurteiler ist, die Übereinstimmung zwischen deren Einschätzungen der Lebensqualität der entsprechenden Person zunimmt (Aaronson 1991, Postulart et al. 2000). Jedoch gilt die individuelle GLQ als subjektives Maß, das durch die betroffene Person selbst eingeschätzt werden soll (Slevin et al. 1988, Fossa et al. 1990, Giannakopoulos 2005). Auch für PCa-Patienten ist eine mangelnde Übereinstimmung zwischen ärztlicher und persönlicher Bewertung der Lebensqualität durch den Betroffenen festgestellt worden (Litwin et al. 1998). Ausnahmen von der Strategie der Selbstbeurteilung sind bei Personen zu machen, die aus gesundheitlichen oder anderen Gründen nicht mehr in der Lage sind, selbst Auskunft über ihr eigenes gesundheitliches Wohlbefinden abzugeben (Sprangers & Aaronson 1992).

Für die Lebensqualitätsmessung bei Betroffenen mit Krebserkrankungen lassen sich verschiedene Zielkriterien unterscheiden (Krischke 1995). Dazu zählen beispielsweise die Verbesserung und Optimierung der Patientenbetreuung, die Bewertung der Therapienebenwirkungen (Toxizität) und Identifikation prognostischer Faktoren, die Beschreibung und Dokumentation des Verlaufs der Erkrankung und der Therapie, die Beurteilung des Rehabilitationsbedarfs von Krebspatienten, die Auswahl alternativer Behandlungsstrategien sowie eine angemessene Ressourcenallokation im Gesundheitswesen. Im klinischen Kontext wird GLQ in erster Linie im Bereich der Maßnahmenplanung auf struktureller Ebene und zur Evaluation von Behandlungsstrategien eingesetzt und weniger zur individuellen Therapieplanung. Für den letztgenannten Zweck stehen verschiedene belastungsorientierte Verfahren zur Verfügung, wie beispielsweise die Psychoonkologische Basisdokumentation (Keller & Mussel 2005) oder der Fragebogen zur Belastung von Krebskranken (FBK-R23, Herschbach et al. 2003), die in der Vergangenheit methodisch weiterentwickelt wurden. Talcott und Clark (2005) weisen jedoch darauf hin, dass auch die GLQ stärker in die Arzt-Patient-Interaktion und gemeinsame Therapieplanung mit einbezogen werden sollte.

Im Folgenden werden drei Gruppen von Fragebogeninstrumenten zur Erfassung gesundheitsbezogener Lebensqualität bei Krebserkrankten vorgestellt. Dabei handelt es sich um generische, krebsspezifische und krankheitsspezifische Instrumente. Diese drei Ausrichtungen von Fragebögen ermöglichen Untersuchungen auf verschiedenen Ebenen: a) Vergleich mit gesunden beziehungsweise nichtkrebskranken Personen aus der Allgemeinbevölkerung (generisch), mit Betroffenen anderer Krebserkrankungen (krebsspezifisch) und beispielsweise verschiedenen Behandlungsstrategien bei Personen, die an derselben Krebserkrankung leiden (krankheitsspezifisch).

#### 2.4.1 Generische Instrumente zur Messung gesundheitsbezogener Lebensqualität

Generische Instrumente erlauben den Vergleich zwischen Personen, die an verschiedenen Krankheiten leiden oder von erkrankten Personen mit der Allgemeinbevölkerung (Erickson 2005, Litwin & Talcott 2005). Es gibt verschiedene Instrumente, die zur Erfassung der GLQ in Studien zur Evaluation von Krebsbehandlungen eingesetzt werden. Beispiele für Instrumente dieser Gruppe sind das *Nottingham Health Profile* (NHP, Hunt et al. 1981) der *Medical Outcomes Study Short-Form 36* (SF-36, Ware & Sherbourne 1992) sowie der *Quality of Life Index* (QL Index, Spitzer et al. 1981). Die beiden erstgenannten Verfahren sind Selbstbeurteilungsinstrumente, der QL-Index nach Spitzer ist ein Fremdbeurteilungsverfahren. Einen kurzen Überblick über diese Instrumente bietet Tabelle 1 (angelehnt an Erickson 2005).

Tabelle 1: Generische Instrumente zur Erfassung der GLQ bei Krebskranken

	NHP	SF-36	QL Index
<b>Erfassung von</b>	Gesundheitsproblemen in der erwachsenen Allgemeinbevölkerung	Körperlicher und psychischer Befindlichkeit	Gesundheitsaspekten bei Personen mit chronischen Krankheiten durch den Arzt
<b>Quantitativer Itemumfang</b>	45 Items	36 Items	5 Items
<b>Inhaltsbereiche</b>	Emotionale Reaktionen, Energielosigkeit, Schmerzen, Physische Mobilität, Schlaf, Soziale Isolation sowie Probleme im Alltag (Beruf, Hausarbeit, soziales Leben, familiäres Leben, Sexualeben, Freizeit)	Körperliche Funktionsfähigkeit, körperliche Rollenfunktion, Schmerz, Allgemeine Gesundheitswahrnehmung, Vitalität, Soziale Funktionsfähigkeit, Emotionale Rollenfunktion, psychisches Wohlbefinden	Alltagsaktivitäten, Gesundheitswahrnehmung, Lebensausblick, körperliche Aktivität, Soziale Unterstützung
<b>Hauptgütekriterien</b>	angemessen bis gut	angemessen bis gut	angemessen
<b>Änderungssensitivität</b>	gegeben	gegeben	nicht bekannt

Anmerkungen: NHP = Nottingham Health Profile, SF-36 = Medical Outcomes Study Short-Form 36, QL Index = Quality of Life Index

Ziel dieser Instrumente im Rahmen von Studien mit onkologischen Patienten ist es, Aspekte oder Bereiche zu erfassen, die mit GLQ assoziiert werden oder diese auf allgemeiner Ebene abbilden. Hauptkritikpunkt an generischen Instrumenten ist der Mangel an Erfassung krankheitsspezifischer Symptome und Behandlungsnebenwirkungen.

#### 2.4.2 Krebspezifische und krankheitsspezifische Instrumente zur Messung gesundheitsbezogener Lebensqualität

Krebspezifische Lebensqualitätsinstrumente in der Onkologie ermöglichen Aussagen über GLQ bei Betroffenen unterschiedlicher Krebserkrankungen (Litwin & Talcott 2005). Diese umfassen Inhaltsbereiche, die für Krebserkrankungen allgemein als relevant angesehen werden. Beispiele sind der Functional Assessment of Cancer Therapy – General (Fact-G, Cella et al. 1993), der European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core 30 (EORTC QLQ-C30, Aaronson et al. 1993) sowie der Cancer Rehabilitation Evaluation System (CARES, Schag & Heinrich 1990). Die genannten Verfahren werden überblicksartig in Tabelle 2 dargestellt. Der EORTC-QLQ-C30 wird in Kapitel 6.2 ausführlich beschrieben.

Tabelle 2: Krebspezifische Instrumente zur Erfassung der GLQ bei Krebskranken

	<b>EORTC QLQ-C30</b>	<b>FACT-G</b>	<b>CARES</b>
<b>Erfassung von</b>	gesundheitsbezogener Lebensqualität bei Krebserkrankungen in klinischen Studien	funktionalem Status und psychischer Befindlichkeit bei Krebspatienten	psychosozialen und körperlichen Problemen
<b>Quantitativer Itemumfang</b>	30 Items	27 Items	zwischen 38 bis 142 Items (je nach Version)
<b>Inhaltsbereiche</b>	Körperliche Funktion, Rollenfunktion, Kognitive Funktion, Emotionale Funktion, Soziale Funktion, Allgemeine Lebensqualität, Fatigue, Übelkeit/Erbrechen, Schmerz, Kurzatmigkeit, Schlafstörungen, Appetitlosigkeit, Durchfall, finanzielle Schwierigkeiten	Körperliches Befinden, Funktionsfähigkeit, Verhältnis zu Freunden, Bekannten und Familie, Seelisches Wohlbefinden	Körperliche Funktion, Psychosoziale Funktion, Interaktion mit Ärzten, Partnerschaft, Sexualität
<b>Hauptgütekriterien</b>	angemessen bis gut	angemessen bis gut	angemessen bis gut
<b>Änderungssensitivität</b>	gegeben	gegeben	gegeben

Anmerkungen: EORTC-QLQ-C30 = European Organisation for Research und Treatment of Cancer Quality of Life Questionnaire – Core 30, Fact-G = Functional Assessment of Cancer Therapy – General, CARES = Cancer Rehabilitation Evaluation System

Krebsspezifische Lebensqualitätsinstrumente dienen der Erfassung allgemeiner Lebensqualitätsdimensionen vor dem Hintergrund krebsspezifischer Therapien und Behandlungen. Neben Dimensionen der GLQ werden teilweise Symptome und behandlungsbezogene Beschwerden und Beeinträchtigungen erfasst, wobei jedoch auf der Ebene allgemeiner Symptomatik erhoben wird. Somit ist zum Beispiel ein Vergleich von Behandlungsmaßnahmen bei verschiedenen Krebserkrankungen möglich.

Verfahren zur Erfassung der GLQ, die speziell für Männer mit Prostatakarzinom entwickelt wurden, werden als krankheitsspezifische Instrumente bezeichnet. Diese sind speziell auf die Bedürfnisse und Belastungen von PCa-Patienten abgestimmt. Sie sind zu unterscheiden von so genannten *symptomspezifischen* Fragebögen, die spezielle Belastungen bei (unterschiedlichen) Krebserkrankungen erfassen, wie beispielsweise Schmerzerleben bei Knochenmetastasen oder ähnliches. Tabelle 3 stellt eine Auswahl prostatakrebspezifischer Lebensqualitätsinstrumente vor (angelehnt an Litwin & Talcott 2005). Darin enthalten sind der European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Prostate Specific Module (EORTC – PSM, Borghede et al. 1997, deutsche Version Bestmann et al. 2006), der Functional Assessment of Cancer Therapy – Prostate (FACT-P, Esper et al. 1997) und der University of California, Los Angeles, Prostate Cancer Index (UCLA-PCI, Karakiewicz et al. 2003).

Tabelle 3: Krankheitsspezifische Instrumente zur Erfassung der GLQ bei PCa-Patienten

	<b>EORTC-PSM</b>	<b>FACT-P</b>	<b>UCLA-PCI</b>
<b>Erfassung von</b>	prostatakrebspezifischen Symptomen und Behandlungsfolgen	Symptomen und Problemen bei Prostatakrebs	Funktionsfähigkeiten und Beeinträchtigungen bei Prostatakrebs
<b>Quantitativer Itemumfang</b>	36 Items	12 Items	15 Items
<b>Inhaltsbereiche</b>	Probleme beim Urinieren, Inkontinenz, erektile Dysfunktion, sexuelle Probleme, Partnerprobleme, Schmerz, Hitzewallung, Ernährung, psychische Belastung	Fatigue, Allgemeine Lebensqualität, urinale Funktionsfähigkeit, Probleme beim Urinieren, Darmbeschwerden, sexuelle Funktionsfähigkeit, Geschlechtsidentität, Gewichtsreduktion	Urinale Funktionsfähigkeit, Beeinträchtigung bei Urinieren, Sexuelle Funktionsfähigkeit, Sexuelle Beeinträchtigung
<b>Hauptgütekriterien</b>	angemessen bis gut	angemessen bis gut	angemessen bis gut
<b>Änderungssensitivität</b>	nicht bekannt	gegeben	nicht bekannt

Anmerkungen: EORTC-PSM = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Prostate Specific Module, FACT-P = Functional Assessment of Cancer Therapy – Prostate, UCLA-PCI = University of California, Los Angeles, Prostate Cancer Index

In der vorliegenden Arbeit wurde der EORTC QLQ-C30 eingesetzt. Die Wahl des Instrumentes orientierte sich in erster Linie an der klinischen Praxis, da dieses Verfahren in Deutschland, wie in den meisten europäischen Ländern, und auch in der kooperierenden onkologischen Rehabilitationsklinik routinemäßig zur Verlaufsdiagnostik eingesetzt wird. Der EORTC-PSM fand daher keine Berücksichtigung. Die Verwendung des krebsspezifischen Fragebogens wird als valides Instrument zur Erfassung der GLQ bei PCa-Patienten angesehen, da sich einerseits der EORTC-PSM modular zusammensetzt aus den Indikatorvariablen der GLQ, die auch im EORTC-QLQ-C30 verwendet werden, sowie einer Reihe krankheitsspezifischer Symptome und Nebenwirkung (s.o.). Andererseits wird zur Modellierung der GLQ in der vorliegenden Arbeit dem Ansatz von Boehmer und Luszczynska (2006) gefolgt, die lediglich in den Indikatorvariablen die Komponenten der GLQ bei Krebspatienten sehen.

## **2.5 Prostatakarzinom und gesundheitsbezogene Lebensqualität**

Nach Vitinius und Rhode (2006) ist Lebensqualität als Kriterium einer psychologischen Dimension in uroonkologischen Studien in Deutschland bisher von nachgeordnetem Interesse. Die Autoren analysierten die Studienprotokolle der Arbeitsgemeinschaft Urologische Onkologie (AUO) der Deutschen Krebsgesellschaft über den Zeitraum von 1995 bis 2005 und kamen zu dem Schluss, dass Lebensqualität als Zielkriterium bei Studien zum PCa im Vergleich zu Studien mit Patienten mit Keimzelltumoren, Blasen- oder NierenzellCa verhältnismäßig häufig untersucht wurde. Allerdings wurde auch bei Studien zum PCa Lebensqualität eher als sekundäres Zielkriterium betrachtet. International ist die Anzahl der Studien zur Lebensqualität bei PCa-Patienten in den vergangenen Jahren stark gestiegen. GLQ hat sich nach Talcott und Clark (2005) zu einem Kernelement der medizinischen Behandlung von PCa-Patienten entwickelt.

Der Fokus der Studien richtet sich in den meisten Fällen entweder auf die Untersuchung unterschiedlicher Behandlungsstrategien, wobei insbesondere operative, medikamentöse oder strahlenbiologische Maßnahmen untersucht wurden, oder auf den Einfluss des Schweregrades der Erkrankung auf die Befindlichkeit der Patienten. Ein weiteres wichtiges Thema für die Durchführung von Untersuchungen zur Lebensqualität bei PCa-Patienten sind die Auswirkungen potenzieller Nebenwirkungen der entsprechenden Behandlungsstrategien. Bevor ein Überblick über die Ergebnisse zur Lebensqualität bei Prostatakrebs gegeben wird, werden einleitend Informationen zu Krebserkrankungen allgemein vorgestellt.

### 2.5.1 Krebserkrankungen allgemein

Aufgrund der demographischen Veränderungen in Deutschland sowie der verbesserten medizinischen Versorgung von Krebsbetroffenen zeichnen sich zwei Trends ab (Tschuschke 2006): Zum Einen wird die Erkrankungswahrscheinlichkeit aufgrund des steigenden durchschnittlichen Lebensalters zunehmen; zum Anderen werden sehr viel mehr Betroffene mit ihrer Krebserkrankung besser leben können. Aktuell wird die Zahl der jährlichen Neuerkrankungen an Krebs in Deutschland bei Männern auf ca. 218.250 und bei Frauen auf ca. 206.000 Fälle geschätzt. Für beide Geschlechter liegt das mittlere Erkrankungsalter etwa bei 69 Jahren. Das mittlere Sterbealter an Krebs liegt bei Männern im Alter von 71 Jahren und bei Frauen bei 76 Jahren. Die relative 5-Jahres-Überlebensrate mit Krebs ist deutlich abhängig von der Tumorlokalisation. Bestimmte Krebsarten, wie beispielsweise der Speiseröhrenkrebs oder der Krebs der Bauchspeicheldrüse haben sehr ungünstige Prognosen, wohingegen andere Krebsarten wie Hodenkrebs oder Brustkrebs deutlich günstigere Prognosen aufweisen. Generell liegt die durchschnittliche Überlebensrate für den Zeitraum von 5 Jahren bei Frauen in etwa bei 58%, bei Männern bei ca. 45%. Die Differenz der Überlebensrate ist auf das häufigere Vorkommen prognostisch ungünstiger Krebsarten bei Männern zurückzuführen (Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2006).

Das persönliche Erleben einer Krebserkrankung ist häufig mit dem Thema Tod und Sterben assoziiert. Nach Angaben des Statistischen Bundesamtes Deutschland (2005) sind in den Jahren 2002 bis 2004 jährlich in etwa 25% (ca. 215.000 Todesfälle) der in Deutschland registrierten Todesfälle auf Krebserkrankungen zurückzuführen. Lediglich Krankheiten des Kreislaufsystems nehmen mit etwa 45% (ca. 390.000) einen Platz vor den Krebserkrankungen in der Statistik der Todesursachen in Deutschland ein.

Bei denen als Krebs bezeichneten Krankheiten handelt es sich jedoch zum Teil um sehr verschiedene Erkrankungsformen. Gemeinsam ist diesem Krankheitsspektrum, dass es sich um maligne neoplastische Neubildungen im Organismus handelt (Tschuschke 2006). Die Klassifikation von Krebserkrankungen erfolgt üblicherweise nach der Lokalisation und nach dem betroffenen Organsystem.

Für die Einteilung von Tumorerkrankungen hat sich die so genannte TNM-Klassifikation etabliert (Deutsche Krebshilfe 2006):



- *T* steht für die Größe des Primärtumors: dieses Kriterium wird in verschiedene Abstufungen unterteilt und variiert von T0 (kein Anhalt für einen Tumor) bis hin zu T4 (Tumor infiltriert benachbartes Gewebe)
- *N* steht für regionäre Lymphknoten und ist im Wesentlichen in zwei Kategorien unterteilt. N0 (kein Befall benachbarter Lymphknoten) und N1 (Befall benachbarter Lymphknoten).
- *M* steht für Fernmetastasen und umfasst eine Unterteilung von M0 (keine Fernmetastasen) oder M1 (Fernmetastasen).

Die Entscheidung über die Art der medizinischen Behandlung wird maßgeblich durch das Stadium der Tumorerkrankung bedingt (Niebrügge 1999). Entsprechend lassen sich die Therapien unterteilen in kurative Behandlungen, deren Ziel die Heilung darstellt; die adjuvante Therapie, die ebenfalls auf eine Heilung beziehungsweise auf eine Verlängerung der Lebenszeit abzielt; und eine palliative Behandlung, deren primäres Ziel die Symptomlinderung in fortgeschrittenen Stadien der Krebserkrankung darstellt. Je nach individueller Entwicklung der Krankheit werden diese Strategien auch kombiniert durchgeführt. Kurative Behandlungen stehen dann im Vordergrund, wenn sich ein Tumor im Anfangsstadium befindet und dessen Entfernung möglich ist, wobei weitere Krebszellen im Körper nicht nachzuweisen sind. Adjuvante Therapie ist angesagt, wenn eine operative Entfernung des Tumors möglich ist und durch eine zusätzliche Therapieform (Chemotherapie, Hormontherapie bzw. Strahlentherapie) die Bildung von Metastasen verhindert werden soll. Palliative Behandlungen sind bei weit fortgeschrittener Krebserkrankung zur Symptomlinderung insbesondere am Lebensende wichtig. Die häufigsten medizinischen Behandlungsmöglichkeiten bei Krebserkrankungen sind chirurgische Eingriffe, Strahlen-, Chemo- und Hormontherapie.

Nach Schwarz (2005) treffen bei der Krankheit ‚Krebs‘ im Allgemeinverständnis zwei Haltungen aufeinander: Einerseits stellt Krebs die Krankheit mit dem höchsten Bedrohungspotenzial dar und andererseits wiederum wird die Krankheit als kaum beeinflussbar gesehen. Eine Krebsdiagnose und -behandlung ist mit enormen Anforderungen und Belastung für den Betroffenen verbunden. Faller (1998) beschreibt verschiedene Bereiche körperlicher, psychischer und sozialer Belastungen, mit denen Krebspatienten konfrontiert sein können:

- Konfrontation mit Tod und Sterben: Die Diagnose sowie die Behandlung führen bei den Betroffenen in der Regel zu einer Auseinandersetzung mit der eigenen Endlichkeit;
- Verletzung der körperlichen Integrität: Einerseits wird der eigene Körper als makelbehaftet erlebt und andererseits kann der Körper durch Amputation auch äußerlich stark verändert werden;
- Autonomieverlust: Die für die meisten Menschen grundlegende Haltung der Selbstbestimmung wird in Frage gestellt; häufig entwickeln Betroffene eher fatalistische Überzeugungen;
- Aktivitätenverlust: Gewohnte Alltagsaktivitäten werden möglicherweise nur noch eingeschränkt durchführbar, mit Fortschreiten der Krankheit reduziert sich die Quantität der Aktivitäten;
- Soziale Isolierung / Stigmatisierung: Die Interaktion zwischen dem Betroffenen und seinen sozialen Bezugspersonen kann sich ändern; Distanzierung von anderen oder durch andere ist eine mögliche Folge
- Bedrohung der sozialen Identität und des Selbstwertgefühls: Ergibt sich u. U. aus den genannten Belastungen, die die Person als Ganzes vor neue Aufgaben und Anforderungen stellt, denen gegenüber sie sich erst neu positionieren muss.

### 2.5.2 Prostatakarzinom

Nach Angaben der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (2006) erkranken national pro Jahr etwa 48.650 Männer an Prostatakrebs. Das PCa ist die häufigste Krebserkrankung bei Männern mit einem Anteil von 22,3% aller bösartigen Neubildungen bei männlichen Erwachsenen. Diese Krebsart tritt vornehmlich im höheren Alter auf. Erkrankungen vor dem 50. Lebensjahr sind sehr selten. Das mittlere Erkrankungsalter wird bei ca. 70 Jahren angegeben. Seit den 1980er Jahren ist ein Anstieg des PCa zu verzeichnen, der in erster Linie mit einer verbesserten Diagnostik und der demographischen Veränderung in Deutschland zusammenhängt. Dennoch ist die Sterberate durch PCa seit ca. 35 Jahren stabil geblieben. Die relative 5-Jahres-Überlebensrate liegt bei 82%. Von den Verlaufsformen lassen sich grob zwei Varianten unterscheiden. Eine aggressive, eher schnell metastasierende Verlaufsform mit einem Erkrankungsbeginn häufig vor dem 60. Lebensjahr und ein langsam voranschreitender Erkrankungsverlauf im höheren Alter. Folgende Möglichkeiten der Behandlung von PCa stehen zur Verfügung:

*Chirurgie:* In einem frühem Erkrankungsstadium ist die Entfernung der Prostata (radikale bzw. totale Prostatektomie (RP)) mit dem Ziel der Heilung möglich. Voraussetzung für diesen Eingriff ist, dass der Tumor auf die Vorsteherdrüse begrenzt ist. Befinden sich an den so genannten Schmittändern, den Organgrenzen der Prostata, noch keine Krebszellen, ist davon auszugehen, dass der Tumor vollständig entfernt werden konnte.

*Strahlentherapie:* Je nach Tumorstadium wird diese Behandlungsmöglichkeit ebenfalls mit dem Ziel der Heilung oder begleitend in Kombination mit anderen Therapiemethoden durchgeführt. Angestrebt wird eine Zerstörung der Tumorzellen in der Prostata. Zwei Bestrahlungsmethoden stehen zur Verfügung: die Bestrahlung von ‚außen‘ oder von ‚innen‘. Bei der Bestrahlung von außen wird eine möglichst hohe Dosis an Röntgenstrahlen auf das Tumorgewebe gerichtet, um den Tumor abzutöten und gleichzeitig gesundes Nachbargewebe weitestgehend zu schonen. Jedoch werden bei der Bestrahlung von außen alle Gewebe belastet, die mit den Röntgenstrahlen Kontakt haben. Bei der Bestrahlung von innen, der Prostata-Späckung, werden kleine radioaktive Stifte (sog. Seeds) in das Organ eingestochen, die dann über die Zeit ihre Radioaktivität an das umliegende Gewebe abgeben.

*Chemotherapie:* Diese basiert darauf, dass durch Zytostatika Krebszellen abgetötet werden. Bei Prostatakrebs wird die Chemotherapie erst dann angewendet, wenn durch operative Verfahren, Strahlentherapie und/oder die Hormontherapie die Erkrankung nicht kurativ behandelt werden konnte.

*Hormontherapie:* Diese ist als eine der systemischen Therapien bei Patienten mit PCa anwendbar, da über die Regulation des Testosteronhaushalts das Tumorstadium mitgesteuert werden kann. Die Testosteronbildung wird durch operative Kastration, Gabe von Luteinisierungshormon-Releasing Hormon-Analoga (LHRH) oder Gabe von Antiandrogenen beeinflusst. Operative Kastration umfasst die Entfernung der Hoden (Orchiektomie). Die Gabe von LHRH-Analoga führt ebenfalls zu einer Reduktion von Testosteron (chemische Kastration). Durch Antiandrogene wird entweder lediglich das Tumorgewebe gegenüber dem Testosteron abgeschirmt oder zusätzlich dazu noch die Testosteronbildung in den Hoden unterdrückt.

*Watchful Waiting:* Unter bestimmten Bedingungen ist es ausreichend, den Tumor lediglich zu kontrollieren, ohne eine der bisher aufgeführten Behandlungsarten durchzuführen.

Dies ist dann der Fall, wenn es sich bei der bösartigen Neubildung um einen Tumor mit günstiger Gewebsreife handelt, die Befunde zur Analyse des Tumors eher unauffällig sind (zum Beispiel ein PSA-Wert  $< 10$  ng/ml), ein hohes Lebensalter bei Diagnosestellung sowie gegebenenfalls das Vorliegen weiterer (lebens-)bedrohlicher Erkrankungen.

Als besondere Belastungen ergeben sich für PCa-Patienten tumor- beziehungsweise behandlungsbedingte Darmbeschwerden, Inkontinenz und erektile Dysfunktion. Diese Folgen der Behandlung und der Erkrankungen sind Faktoren, die möglicherweise die Lebensqualität von Betroffenen beeinflussen. Weiterhin können im Rahmen der genannten Behandlungsmaßnahmen die für diese Therapien generell bekannten Nebenwirkungen auftreten. Dazu zählen im Verlauf einer externalen Strahlentherapie Hautentzündungen oder auch Beeinträchtigungen von ebenfalls bestrahltem gesunden Gewebe (v.a. Darm und Blase). Im Rahmen der Chemotherapie ist das Auftreten von Haarausfall, eine Entzündung der Mundschleimhäute oder Übelkeit und Erbrechen möglich. Bei der Hormontherapie kann es zu Stimmungsveränderungen bis hin zu depressiven Auslenkungen und Antriebsarmut kommen. Eine langfristige Hormontherapie kann Stoffwechselstörungen in den Knochen verursachen. Insgesamt läßt sich jedoch sagen, dass die allgemeinen Therapienebenwirkungen bei Patienten mit Prostatakrebs eher selten auftreten und stark vom Ausmaß der Erkrankung und der spezifischen Therapie abhängen (Deutsche Krebshilfe e.V. 2006).

### 2.5.3 Empirische Befunde zu Prostatakrebs und Lebensqualität

Die Befundlage zur GLQ bei PCa-Patienten ist vielfältig. Einerseits zeigen Studien, dass die Diagnosestellung oder Behandlungsmethoden keinen bedeutsamen Einfluss auf die GLQ haben (Lim et al. 1995, Sanchez-Ortiz et al. 2000). Zudem existieren Studien, die Hinweise dafür liefern, dass sich die GLQ von PCa-Patienten in verschiedenen Phasen nach der Behandlung nicht von der Lebensqualität einer nichterkrankten Vergleichsgruppe unterscheidet (Albertsen et al. 1997, Johnstone et al. 2000). Andererseits gibt es eine Vielzahl an Befunden, die – je nach Krankheitsstadium, Behandlungsmethode oder Patientenkollektiv – zu gegenteiligen Ergebnissen kommen. Im Folgenden wird ein Überblick über diese Studien gegeben, ohne dass die hier vorgenommene Darstellung einen Anspruch auf Vollständigkeit erhebt.

## *Querschnittstudien zu Prostatakarzinom und Lebensqualität*

*Ergebnisse zur gesundheitsbezogenen Lebensqualität.* Davis und Kollegen (2001) fanden Unterschiede in der allgemeinen Lebensqualität von PCa-Patienten je nach der Behandlungsmethode (Radikale Prostatektomie (RP), Seeds und externe Strahlentherapie). Patienten, die mit externer Strahlentherapie behandelt wurden, wiesen in der Körperlichen Funktionsfähigkeit, Körperlichen Rollenfunktion sowie der Allgemeinen Gesundheitswahrnehmung im SF-36 statistisch signifikant schlechtere Werte auf als eine der beiden anderen Behandlungsmethoden. Fowler und Kollegen (1995) wiesen bei PCa-Patienten nach RP dann eine negativere Einschätzung der allgemeinen Lebensqualität nach, wenn die Befragten von erheblichen Kontinenzproblemen betroffen waren. Brandeis und Kollegen (2000) berichteten mehr Einschränkungen der GLQ bei PCa-Patienten, die eine Kombinationstherapie aus Seeds und externer Strahlentherapie erhielten, im Vergleich zu einer Gruppe, die lediglich mit Seeds behandelt wurde. Mols und Kollegen (2006) stellen aufgrund einer Repräsentativerhebung aller registrierten PCa-Patienten einer niederländischen Großstadt fest, dass im Langzeitverlauf (ca. 5 bis 10 Jahre postoperativ) einerseits die Allgemeine Gesundheitswahrnehmung im SF-36 schlechter war als die einer altersgemachten Vergleichsstichprobe, dafür die psychische Befindlichkeit (gemessen mit dem Quality of Life-Cancer Survivor Questionnaire) jedoch vergleichsweise positiver von den PCa-Patienten beurteilt wurde. Hinsichtlich der Physischen Funktionsfähigkeit zeigt sich in dieser Studie die RP der Watchful Waiting Strategie überlegen, wobei Patienten mit Watchful Waiting wiederum bessere physische Funktionswerte berichteten als Strahlentherapiepatienten. Hormontherapeutisch Behandelte wiesen die vergleichsweise geringsten Werte auf der Skala Physische Funktionsfähigkeit auf. Galbraith und Kollegen (2001) geben jedoch für Watchful-Waiting-Patienten Belastungen und Einschränkungen in der Allgemeinen Gesundheitswahrnehmung im SF-36 an.

Auf der Ebene der allgemeinen GLQ als Outcome-Parameter läßt sich kein klares Bild bezüglich der Auswirkungen von Behandlungen bei Patienten mit PCa ableiten. Kurzfristig scheint die (externale) Strahlentherapie möglicherweise eher zu Beeinträchtigungen der Allgemeinen Gesundheitswahrnehmung (gemessen durch den SF-36) sowie der Physischen Komponenten der GLQ führen. Bei der RP scheint das Ausmaß der Kontinenzprobleme ein moderierender Faktor zu sein. Es ergaben sich jedoch Hinweise, dass langfristig die Hormontherapie die Physische Komponente der GLQ stärker beeinträchtigt als die anderen Therapieverfahren.

*Ergebnisse zu behandlungsspezifischen Auswirkungen.* Eton und Kollegen (2001) kommen zu dem Ergebnis, dass im Vergleich zwischen Radikaler Prostatektomie (RP) und den Bestrahlungsformen (external vs. Seeds) kurzfristig die Behandlung mit Seeds die geringeren Beschwerden im Bereich der sexuellen und physischen Funktionsfähigkeit zur Folge hat. Davis und Kollegen (2001) kommen zu einem vergleichbaren Ergebnis. Kirschner-Hermanns und Jaske (2002) weisen nach, dass bei RP häufiger Probleme mit der Kontinenz sowie der Sexualität auftreten. Beim Vergleich der RP mit Strahlentherapie zeigen sich bei Bestrahlungen vergleichsweise mehr Darm- und Verdauungsschwierigkeiten im Zeitraum von ein bis fünf Jahren nach der primären Behandlung (Shrader-Bogen et al. 1997). Hu und Kollegen (2004) weisen darauf hin, dass lediglich 20 Prozent der mit RP behandelten PCa-Patienten nach einem Jahr das präoperative Ausgangsniveau ihrer sexuellen Potenz wiedererreicht hatten. Auch langfristig bleiben bei Patienten mit Prostatakrebs nach der Behandlung mit RP im Vergleich mit externaler Strahlentherapie sexuelle Schwierigkeiten bestehen (Schover et al. 2002). Clark und Kollegen (2003) kommen zu dem Schluss, dass PCa-Patienten nach RP und/oder Strahlentherapie im Vergleich zu einer gesunden Stichprobe Einschränkungen im Bereich des Urinal- und Darmtraktes sowie der Sexualität aufweisen. Fransson und Widmark (1999) fanden in einer Vergleichsstudie nach externer Strahlentherapie mit einer Stichprobe aus der gesunden Allgemeinbevölkerung auch nach mehreren Jahren eine erhöhte Rate an Problemen mit der sexuellen Funktionsfähigkeit, der Kontinenz und der Verdauungstätigkeit bei PCa-Patienten. Aufgrund des Querschnittsdesigns der Studie wurde jedoch nicht das Ausgangsniveau beider Gruppen kontrolliert. McCammon und Kollegen (1999) fanden beim Vergleich von RP mit externer Strahlentherapie mehr Inkontinenzprobleme bei Patienten nach chirurgischem Eingriff, aber mehr Verdauungsbeschwerden bei Strahlentherapiepatienten.

Zusammenfassend ergeben sich Hinweise darauf, dass nach RP die sexuelle Funktionsfähigkeit stärker beeinträchtigt zu sein scheint, als nach Strahlentherapie. Weiterhin scheint das Ausmaß der Beeinträchtigungen der Sexualität, Kontinenz und Verdauung nach der Behandlung bei Patienten mit PCa massiver zu sein, als dies in der gesunden Allgemeinbevölkerung der Fall ist. Ob dieser Befund auf die Behandlungen zurückzuführen ist, bleibt jedoch fraglich, da in Querschnittstudien das Ausgangsniveau nicht kontrolliert wird.

## *Längsschnittstudien zum Prostatakarzinom und Lebensqualität*

Die Ergebnisse der Längsschnittuntersuchungen zeigen ein ähnlich heterogenes Bild wie die Querschnittstudien hinsichtlich der langfristigen Entwicklung der GLQ sowie der krebs- und behandlungsspezifischen Auswirkungen.

*Ergebnisse zur gesundheitsbezogenen Lebensqualität.* Korfage und Kollegen (2005) weisen darauf hin, dass die GLQ von PCa-Patienten nach RP oder Strahlentherapie langfristig nicht von den Werten einer altersentsprechenden Stichprobe gesunder Personen unterscheidbar ist und bestätigen damit einen Befund von Hanlon und Kollegen (2001). Lubeck und Kollegen (1999) berichten in einer Vergleichsstudie, dass Patienten direkt nach Behandlung durch RP deutlich schlechtere Werte in der gesundheitsbezogenen und krebsspezifischen Lebensqualität angaben als Patienten nach Hormon-, Strahlentherapie oder Watchful-Waiting-Patienten. Allerdings zeigten die Patienten in der RP-Gruppe im Jahr nach der Behandlung die deutlichsten Verbesserungen in beiden Lebensqualitätsbereichen, die die Einschätzungen der anderen Gruppen zum Teil deutlich überstiegen.

Die berichteten Ergebnisse liefern Hinweise darauf, dass die GLQ von PCa-Patienten unabhängig von der Art der Behandlung langfristig das Niveau der gesunden Allgemeinbevölkerung erreicht, dass jedoch insbesondere bei Durchführung einer RP kurzfristig eine Beeinträchtigung der Lebensqualität resultieren kann.

*Ergebnisse zu behandlungsspezifischen Auswirkungen.* Litwin und Kollegen (1999a, 1999b) untersuchten die Auswirkungen der RP bei PCa-Patienten im Verlauf eines Jahres nach der Operation. Nach etwa sechs Monaten erreichte die Einschätzung der Darmbeschwerden wieder das präoperative Niveau. Jedoch ergaben sich nach einem Jahr postoperativ bei etwa 40% weiterhin Probleme in den Bereichen Kontinenz und sexuelle Funktionsfähigkeit. Auswertungen derselben Stichprobe zum Vergleich zwischen RP und externer Strahlentherapie erbrachte für die Patienten mit Strahlentherapie eine deutliche Abnahme der sexuellen Funktionsfähigkeit im zweiten Jahr nach der Behandlung, was bei Patienten nach RP nicht zu verzeichnen war. Madalinska und Kollegen (2001) fanden beim Vergleich der krebsspezifischen Lebensqualität bei Patienten nach RP oder externe Strahlentherapie mehr Inkontinenzprobleme und Beeinträchtigungen der sexuellen Funktionsfähigkeit bei RP-Patienten. Diese Patienten gaben bezüglich ihrer sexuellen Funktionsfähigkeit auch eine stärkere Belas-

tung an. Unterschieden sich die beiden Gruppen nicht im Ausmaß der Verdauungsbeschwerden auf objektivem Niveau, berichteten Strahlentherapiepatienten jedoch von intensiveren Belastungen durch Verdauungsschwierigkeiten. Schapira und Kollegen (2001) zeigten in einer Therapievergleichsstudie bei PCa-Patienten mit lokal begrenztem Tumor, dass Patienten im Jahr nach RP über deutlich mehr Schwierigkeiten in den Bereichen Inkontinenz und sexuelle Funktionsfähigkeit klagten, wobei Patienten nach Strahlentherapie lediglich Schwierigkeiten im Bereich der Sexualität angaben. Die Watchful Waiting-Patienten zeigten in allen untersuchten krankheitsspezifischen Bereiche die unauffälligsten Werte. In Bezug auf einen Vergleich einer nerverhaltenden Prostatektomie und einer nicht-nerverhaltenden RP wiesen Talcott und Kollegen (1997) darauf hin, dass sich zwar bei den nerverhaltend Behandelten eine signifikante Verschlechterung der sexuellen Funktionsfähigkeit im Zeitverlauf zeigte. Jedoch ließ sich kein statistisch signifikanter Unterschied zwischen den beiden Gruppen nachweisen. Litwin und Kollegen (1999b) kommen dagegen zu dem Schluss, dass bei einer nerverhaltend behandelten Gruppe von PCa-Patienten im Verlauf von bis zu 18 Monaten signifikant bessere Werte im Bereich der sexuellen Funktionsfähigkeit nachweisbar waren, dass jedoch die Werte beider Gruppen (nerverhaltend und nicht-nerverhaltend) in einem sehr niedrigen Bereich lagen. Wei und Kollegen (2002) verweisen auf ein Ergebnis, demzufolge Seeds im Vergleich mit externaler Bestrahlung und RP die deutlichsten Beschwerden und Belastungen nach sich ziehen. Miller und Kollegen (2005) weisen jedoch auf widersprechende Befunde hin. In dieser Studie zum Vergleich von Wirkungen und Nebenwirkungen von Behandlungen zeigt sich bei der Gruppe der RP-Behandelten langfristig eine stabile Lebensqualität, wohingegen die externe Bestrahlung sowie die Seed-Behandlung zu langanhaltenden Schwierigkeiten im Bereich der Kontinenz führen.

Insgesamt läßt sich aus den dargestellten Ergebnissen ableiten, dass RP häufiger mit Beeinträchtigungen im Bereich der Sexualität und der Kontinenz verbunden ist. Verdauungsschwierigkeiten scheinen bei dieser Art der Behandlung keine nennenswerte Bedeutung zu haben. Dies ist jedoch vor allem bei der externen Strahlentherapie der Fall, wobei diese hinsichtlich der Beeinträchtigung der Sexualität der internen Bestrahlung mittels Seeds langfristig überlegen zu sein scheint. Watchful Waiting-Patienten zeigen langfristig erwartungsgemäß die geringsten Veränderungen.

*Studien zur Lebensqualität bei PCa im fortgeschrittenen Stadium.* Curran und Kollegen (1997) wiesen nach, dass Männer mit hormonresistenten PCa im fortgeschrittenen Stadium



der Krebserkrankung größere Beeinträchtigungen in den Bereichen Rollen- und Physische Funktionsfähigkeit sowie der allgemeinen Lebensqualität aufweisen als eine Vergleichsgruppe im selben Erkrankungsstadium mit erfolgreicher Hormontherapie. Herr und Sullivan (2000) weisen bei einer Patientengruppe im fortgeschrittenen Stadium der Erkrankung, die erfolgreich mit Hormontherapie behandelt wurde, ein höheres Ausmaß an Fatigue und emotionaler Belastung nach, als bei einer nichtbehandelten Vergleichsstichprobe, die mit der Therapie ausgesetzt hatte. Nach Litwin und Kollegen (1998b) hat bei fortgeschrittenem Krankheitsverlauf die Form der Kastration (chirurgisch vs. medikamentös) keine Auswirkung auf die GLQ der von ihnen untersuchten Studienteilnehmer. Ergebnisse von Clark und Kollegen (2001) stützen diesen Befund. Ein Vergleich zwischen der chirurgischen Kastration und einer kombinierten Hormontherapie erbrachte keine bedeutsamen Unterschiede in der allgemeinen und spezifischen GLQ zwischen den Gruppen (Litwin et al. 1998a), wobei Moinpour und Kollegen (1998) nachweisen konnten, dass in einer randomisierten Kontrollgruppenstudie eine Gruppe mit chirurgischer Kastration und der Gabe von Placebo bessere Verdauungstätigkeiten aufwies als eine Gruppe mit kombinierter Hormontherapie. Die zuletzt genannte Gruppe zeichnete sich jedoch durch eine bessere Emotionale Funktionsfähigkeit drei und sechs Monate nach Behandlungsbeginn aus. Osoba und Kollegen (1999) konnten bei hormonresistentem PCa in fortgeschrittenem Stadium eine Verbesserung der allgemeinen Lebensqualität bei Männern mit einer kombinierten Hormon- und Zytostatikabehandlung im Vergleich zur Monotherapie mit Hormonpräparat nachweisen. Im Langzeitverlauf zeigte sich jedoch in der Gruppe mit Kombinationstherapie v.a. dann eine statistisch relevante Verbesserung der GLQ, wenn die Zytostatikabehandlung als Kombinationstherapie (Mitoxandron und Prednison) durchgeführt wurde. Ein Vergleich zwischen einer Monotherapie (Bicalutamid) mit einer kombinierten Hormontherapie bei fortgeschrittenem Prostatakrebs erbrachte eine statistisch bedeutsame Überlegenheit der Monotherapie im Bereich der sexuellen und sozialen Funktionsfähigkeit. Unterschiede im Krankheitsverlauf und in der Überlebenszeit gab es jedoch keine (Boccardo und Kollegen 1999).

Aus den bisherigen Befunden zur Lebensqualität bei fortgeschrittenem PCa ergeben sich Hinweise darauf, dass eine erfolgreiche Hormontherapie einerseits positive Auswirkungen auf die allgemeine Lebensqualität, Rollen- und Physische Funktionsfähigkeit zeigt. Andererseits zeigt sich eine Verschlechterung der Werte für Fatigue und emotionale Belastungen. Bezüglich der Art der Hormontherapie (chirurgisch vs. medikamentös) scheint es insgesamt keine Unterschiede in den Werten zur Lebensqualität der betroffenen PCa-Patienten zu geben.

*Allgemeine Bewertung der Untersuchungen zu PCa und gesundheitsbezogener Lebensqualität.* Nach Penson und Kollegen (2003) besitzt die GLQ bei Prostatakrebs in allen Phasen der Krankheitsbehandlung besondere Bedeutung. Wird der Tumor frühzeitig diagnostiziert und ist eine gute Prognose gegeben, kann die Lebensqualität aufgrund der verschiedenen Wirkungen und Nebenwirkungen der Behandlungen dennoch zumindest zeitweise negativ beeinflusst werden. Ist die Erkrankung weiter fortgeschritten, sind bereits vermehrte Metastasierungen bekannt oder handelt es sich um ein Stadium mit überwiegend palliativer Behandlung, ist die Lebensqualität des Betroffenen eines der wichtigsten Behandlungsziele.

Efficace und Kollegen (2003) geben einen Überblick über 25 randomisierte kontrollierte Studien (RCT), in denen PCa-Patienten hinsichtlich ihrer Lebensqualität untersucht wurden. Vornehmlich handelte es sich dabei um Studien, die den kurzfristigen Einfluss von Hormontherapie beim metastasierten Prostatakrebs untersuchten. Eine Aussage über den Vergleich der Wirksamkeit verschiedener Behandlungsmethoden ist den Autoren zufolge nicht möglich, da zu unterschiedliche Krankheitsstadien (lokal begrenzter Tumor vs. Ausstrahlungen) sowie eher partielle Therapievergleiche (verschiedene Präparate einer Behandlungsform, z. B. verschiedene Hormonpräparate bzw. verschiedene Zytostatika) vorgenommen wurden. Zusätzlich waren die Instrumente zur Erfassung der GLQ verschieden beziehungsweise manche Subskalen gebräuchlicher Fragebögen wurden nicht verwendet. Aufgrund der methodischen Kritikpunkte an den durchgeführten RCTs weisen die Autoren auf die begrenzte Aussagekraft der Einzelergebnisse hin. Allgemeine Aussagen über den Zusammenhang zwischen GLQ und PCa sowie dessen Behandlungsmethoden scheinen aufgrund der inhaltlichen und methodischen Vielfalt der publizierten Studien nur schwer ableitbar.

Entsprechend den dargestellten Studienergebnissen kommen Eton und Lepore (2002) in ihrem Übersichtsartikel zu dem Schluss, dass je nach Krankheitsstadium und Behandlungsmethode unterschiedliche Beschwerden und Lebensqualitätsveränderungen auftreten (können). Insbesondere krankheits- und behandlungsspezifische Beschwerden stehen jedoch bei den meisten Arbeiten im Vordergrund. Veränderungen auf der Ebene der allgemeinen und GLQ werden eher selten berichtet. Bedeutsam für den Einfluss der Erkrankung auf die GLQ scheint das Krankheitsstadium zu sein. Ist der Tumor auf die Vorsteherdrüse begrenzt, also nicht metastasiert, stehen Inkontinenz und sexuelle Dysfunktion im Vordergrund, wobei vor allem die Beeinträchtigung der Sexualität häufig dauerhaft ist. Hinsichtlich der allgemeinen GLQ lässt sich in Querschnittsstudien oftmals keine nennenswerte Beeinträchtigung nachwei-

sen, wobei diese Studien häufig einige Jahre nach der Behandlung durchgeführt wurden und mögliche Veränderungen über die Zeit nicht berücksichtigt werden konnten. In Längsschnittstudien wurden häufiger temporäre Veränderungen in verschiedenen Bereichen der GLQ festgestellt. Follow-up-Messungen nach bis zu einem Jahr zeigten jedoch in den meisten Studien auch für die genannten Komponenten Werte im Bereich ‚nicht-prostatakrebs-erkrankter‘ Männer oder des Ausgangsniveaus zu Studienbeginn. Ist die Krebserkrankung weiter fortgeschritten, gelten andere Behandlungsmethoden als Therapie der Wahl, was in der Konsequenz dazu führt, dass von den Betroffenen in geringerem Maße Probleme wegen Inkontinenz oder Darmbeschwerden berichtet werden. Beeinträchtigungen der sexuellen Funktionsfähigkeit sind teilweise krankheitsspezifisch, bestehen im fortgeschrittenen Stadium zumeist bereits vor der Behandlung, werden von den Betroffenen jedoch als weniger belastend wahrgenommen. Im Gegensatz zu den Betroffenen im früheren Erkrankungsstadium zeigen sich bei schwerer Erkrankten eher Beeinträchtigungen der allgemeinen und GLQ. Diese beziehen sich auf die Bereiche Schmerzen, Vitalität sowie auf andere Komponenten der GLQ. Häufig gezeigte Befunde einer sich normalisierenden GLQ zu Follow-up-Messungen können durchaus auch mit einem positiven Selektionseffekt in Zusammenhang gebracht werden, da Personen, deren Gesundheitszustand zu Beginn der Studie oder im Laufe der Zeit schlechter war/wurde, möglicherweise nicht bei den Analysen entsprechend berücksichtigt wurden.

Sommers und Ramsey (1999) kommen zu dem Schluss, dass generische Lebensqualitätsinstrumente nicht geeignet seien, Veränderungen der Lebensqualität bei Prostatakrebs-Betroffenen abzubilden. Diese Annahme scheint jedoch insbesondere für schwerer Erkrankte nicht mehr uneingeschränkt zutreffend zu sein. Kritisch an den Übersichtsartikeln von Eton und Lepore (2002) sowie Sommers und Ramsey (1999) ist anzumerken, dass zum Teil keine Längsschnittstudien und Studien berücksichtigt wurden, die Betroffene in fortgeschrittenem Krankheitsstadium untersuchten.

Obwohl die Studienergebnisse zu PCa und GLQ widersprüchlich sind, lassen sich nach Litwin und Talcott (2005) aus den bisherigen Befunden folgende Aussagen ableiten:

1. Beeinträchtigungen der sexuellen Funktionsfähigkeit, des Urogenitaltrakts und Darmbereichs sind nach der Behandlung von Prostatakrebs im Frühstadium häufiger als bisher angenommen;

2. Globalmessungen der Lebensqualität weisen zum Teil geringe Veränderungen über die Behandlungen beziehungsweise die Zeit auf, wohingegen krankheits- und behandlungsspezifische Maße eher Variabilität zeigen;
3. Die Beziehung zwischen organspezifischen Dysfunktionen und erlittenen Belastungen und Beeinträchtigungen sind interindividuell und intraindividuell sehr unterschiedlich;
4. Beeinträchtigungen variieren über die Zeit und je nach Behandlungsmaßnahme
5. GLQ ist ein subjektives Konstrukt, das sich im Verlauf der Erkrankung verändern kann und somit immer wieder im Verlauf der Erkrankungen bei Entscheidungen thematisiert und berücksichtigt werden sollte.

In den vergangenen Jahren sind vermehrt Studien zur Erfassung und Veränderung der GLQ bei PCa-Patienten durchgeführt worden. Darin zeigt sich die Bedeutung des Konstrukts für PCa-Patienten sowie deren Behandlung und den weiteren Krankheits- und Lebensverlauf der Betroffenen. Neben der konzeptuellen Frage, wie GLQ für PCa-Patienten beschrieben werden kann, und der behandlungsorientierten Frage nach den Auswirkungen von Therapie-maßnahmen auf die GLQ von Betroffenen, ist die methodische Erfassung des Konstrukts und dessen Veränderungen von großer Wichtigkeit. Das betrifft zum Einen die Evaluation von Maßnahmen mit GLQ als Outcome-Parameter. Zum Anderen ist das Konstrukt auch für die Bewertung individueller Therapieverläufe zunehmend wichtiger. Die wesentlichen Voraussetzungen für die Verwendung des Konstrukts für die gerade beschriebenen Zwecke sind dessen genaue Erfassung sowie die klare Interpretierbarkeit der gemessenen Werte. Dies ist im Rahmen von Veränderungsmessungen, die anhand der GLQ vorgenommen werden, von großer Bedeutung. Auf dieses Thema zielt die vorliegende Arbeit: Es soll untersucht werden, inwieweit sich ändernde Bewertungshintergründe einen Einfluss haben auf die Veränderungsmessung der GLQ bei PCa-Patienten.

## **2.6. Onkologische Rehabilitation bei Patienten mit Prostatakrebs**

Nach Angaben der Deutschen Rentenversicherung (2005) lag der Anteil stationärer onkologischer Rehabilitationsmaßnahmen im Jahre 2004 bei ca. 20% aller durchgeführten Maßnahmen der medizinischen Rehabilitation. Bei Frauen ist die onkologische Rehabilitation im stationären Bereich mittlerweile die zweithäufigste Indikationsstellung nach der Rehabilitation orthopädischer Erkrankungen. Bei Männern nimmt die onkologische Rehabilitation ebenfalls den zweiten Platz hinter der Rehabilitation von Krankheiten des Muskel-Skelett-Systems

ein, teilt sich diesen Rangplatz jedoch mit den Maßnahmen bei psychosomatischen Erkrankungen.

Ein Vergleich der aktuellen Zahlen der Deutschen Rentenversicherung mit Angaben aus dem Jahr 1995 zeigt eine Zunahme von Maßnahmen der stationären onkologischen Rehabilitation bei Männern um ca. 140 % sowie bei Frauen um 60 %. Im Jahre 2004 wurden 22.307 stationäre Leistungen zur medizinischen Rehabilitation bei benigner Neubildung der Prostata abgeschlossen. Das Durchschnittsalter der behandelten Männer lag bei 65,5 Jahren. Anschlussheilbehandlungen (AHB) bei Prostatakrebs wurden im Jahr 2004 bei 15.425 Männern durchgeführt. Deren Durchschnittsalter lag bei 65,6 Jahren. Voraussetzungen zur Erlangung von Leistungen der Rehabilitation sind gesetzlich festgelegt (§§ 9 – 12 SGB V und § 31 Abs. 1, Satz 3 SGBVI (Rentenversicherung)). Wesentlich für die Gewährleistung von Maßnahmen der stationären onkologischen Rehabilitation ist jedoch nicht die Wiederherstellung der Erwerbs- und Arbeitsfähigkeit. Für Krebskranke werden rehabilitative Leistungen als Nach- und Festigungskuren wegen Geschwulsterkrankungen begründet, unabhängig von deren Beeinträchtigung der Erwerbsfähigkeit.

Die Rehabilitation von Menschen mit Krebserkrankungen hat als wesentliches Ziel, die körperlichen, psychischen und sozialen Beeinträchtigungen, die durch das Krebsleiden selbst oder dessen Behandlung auftreten, zu reduzieren, wobei hervorgehoben wird, dass eine Verminderung funktioneller Ausfälle ohne Verbesserung der Lebensqualität oder ohne soziale Integration nicht als voller Rehabilitationserfolg anzusehen ist (AWMF 2002). Nach Niebrügge (1999) ergeben sich unterschiedliche Zielkriterien der stationären onkologischen Rehabilitation, je nach Art der Behinderung, Funktionseinschränkungen und Belastungen, wobei als übergeordnetes Ziel die Verbesserung der individuellen Lebensqualität des Patienten angegeben wird (dazu auch Koch 2000). Bei der rehabilitativen Behandlung onkologischer Patienten sind nach Herschbach und Keller (1997) sowie Lübbe (1998) spezifische Belastungen zu berücksichtigen. Dazu zählen u.a. die Ungewissheit über den weiteren Krankheitsverlauf, die Belastungen durch die Therapie, die Notwendigkeit der kontinuierlichen diagnostischen Überwachung und die möglicherweise lang anhaltende psychische Belastung für die erkrankte Person und häufig auch deren Angehörige.

Therapeutische Maßnahmen im Rahmen einer onkologischen Rehabilitation umfassen allgemeine (Physiotherapie, Ergometertraining, Muskelaufbau etc.) und spezifische therapeu-

tische Maßnahmen (manuelle Lymphdrainage, Schmerztherapie etc.), Ernährungs-, Schulungs- und Informationsangebote, psychologische Interventionen und Psychotherapie sowie soziale Hilfeleistungen und gegebenenfalls berufliche Maßnahmen. Die Rehabilitation bei PCa-Patienten zielt auf eine umfangreiche Reduktion von gesundheitlichen Beeinträchtigungen, die sich nach Heim und Schwerte (2006) in somatische, psychosoziale, aktivitätsbezogene und teilhabeorientierte Beeinträchtigungen gruppieren lassen (siehe Tabelle 4).

Tabelle 4: Beeinträchtigungen der Gesundheit bei Prostatakrebspatienten

somatisch	psychosozial	aktivitätsbezogen	teilhabeorientiert
direkte Operationsfolgen Strahlentherapiefolgen Therapiebedingte Hormonausfallerscheinungen Harninkontinenz Folgeerscheinungen zytostatischer Chemotherapie Sexuelle Funktionsstörungen	Probleme der Krankheitsverarbeitung Depression Rezidivängste Schlafstörungen Partnerschaftliche Probleme Fatiguesyndrom Posttraumatische Belastungsstörung	Verminderung der körperlichen Belastbarkeit Einschränkungen im Bereich der Fortbewegung Sozialer Rückzug	Probleme der Integration in soziales Umfeld Probleme bei der beruflichen Wiedereingliederung Einschränkung der Mobilität und Teilnahme am kulturellen Leben

Aus diesen Beeinträchtigungen leiten sich die allgemeinen Rehabilitationsziele ab. Diese sind in Tabelle 5 dargestellt, wobei aus Gründen der Übersichtlichkeit zu jedem Therapiebereich lediglich zwei Ziele vorgestellt werden. Für die Bewertung der Zielerreichung werden unterschiedliche Assessmentverfahren in der Praxis angewendet. Neben medizinischen Verfahren zur Erfassung körperlicher Prozesse und Veränderungen kommt der Erfassung durch Selbstbeurteilungsinstrumente zunehmend Bedeutung zu. Zur Evaluation der GLQ verweisen Heim und Schwerte (2006) auf den EORTC-QLQ-C30 als Fragebogen der Wahl.

Nach Haaf (2005) erfüllt die onkologische Rehabilitation in Deutschland ihren Versorgungsauftrag. In Evaluationsstudien zur Wirksamkeit zeigen sich Hinweise auf eine Verbesserung der körperlichen und psychischen Befindlichkeit der untersuchten Patienten. Für die vorliegende Arbeit gilt diese Aussage jedoch nur eingeschränkt, da sich Haaf in seinem Überblicksartikel in erster Linie auf die onkologische Rehabilitation bei Mamma-Ca-Patientinnen bezieht. Hergert und Kollegen (2009) erstellten eine systematische Literaturliteraturanalyse zur interventionsspezifischen Wirksamkeit bei der Rehabilitation von Patienten mit PCa und kommen zu dem Schluss, dass die spezifische Wirksamkeit einzelner Maßnahmen bisher kaum empirisch belegt werden kann.

Tabelle 5: Rehabilitationsziele bei Prostatakrebspatienten (Heim & Schwerte 2006)

---

#### Somatische Therapieziele

- Förderung der allgemeinen körperlichen Leistungsfähigkeit
- Verminderung postchirurgischer Probleme (Narbenschmerzen, Serome)

---

#### Funktionsbezogene Therapieziele

- Besserung der Harninkontinenz
- Umgehen mit sexuellen Funktionsstörungen, Besserung der erektilen Dysfunktion

---

#### Psychosoziale Therapieziele

- Förderung der Krankheitsbewältigung, Verbesserung von Selbstwahrnehmung und Selbstakzeptanz, emotionale Stabilisierung
- Hilfe bei Neuorientierung und beim Umgang mit krankheitsbedingten Einschränkungen (Inkontinenz, erektile Dysfunktion)

---

#### Soziale Therapieziele

- Vorbereitung der beruflichen Wiedereingliederung ggf. Einleitung berufsfördernder Maßnahmen
- Erhalt von Selbstversorgung, Haushaltsführung und Teilhabe am gesellschaftlichen Leben

---

#### Edukative Therapieziele

- Informationsvermittlung und medizinische Beratung zur Erkrankung, Tumortherapie, Sekundär- und Tertiärprävention, Nachsorge
- Erlernen des Kontinenztrainings und Übertragung auf Aktivitäten des Alltags

Abschließend ist festzuhalten, dass ein wesentliches Ziel der onkologischen Rehabilitation darin besteht, die Lebensqualität des Patienten zu verbessern. Daher sind Maßnahmen der onkologischen Rehabilitation bei Patienten mit PCa auf somatische, funktionsbezogene, psychosoziale, soziale und edukative Ziele ausgerichtet. Erfasst wird die Zielerreichung unter anderem durch den EORTC-QLQ-C30, einem krebspezifischen Lebensqualitätsfragebogen. Dieses Instrument liegt auch der vorliegenden Arbeit zugrunde. Fraglich ist bisher, welchen Einfluss eine Veränderung der Bewertungshintergründe der Patienten auf die Evaluation von Maßnahmen in der onkologischen Rehabilitation hat.

## 2.7 Zusammenfassung

GLQ ist ein Teilbereich der allgemeinen Lebensqualität und wird definiert als die subjektive Beurteilung der körperlichen Funktionsfähigkeit, des psychischen Befindens, der sozialen Beziehungen und Funktionsfähigkeit im Alltag. Im Zusammenhang mit Krebserkrankungen besitzt das Konstrukt vor allem im Rahmen der Evaluation von Behandlungsmaßnahmen große Bedeutung. Das ist zum Einen darauf zurückzuführen, dass Krebserkrankungen an sich mit großen Belastungen für die Betroffenen verbunden sind. Dazu zählt die Auseinandersetzung mit dem Thema Tod und Sterben genauso wie die Belastung durch Therapiemaßnahmen

und Behandlungsnebenwirkungen. Zum Anderen wird die GLQ immer mehr als Maß betrachtet, in dem die betroffenen Personen ihre individuellen Gewichtungen aller für sie persönlich bedeutsamen Gesundheitsaspekte vornehmen.

Boehmer und Luszczynska (2006) überprüfen ein Modell der GLQ bei Krebspatienten und unterscheiden zwischen Indikatoren der GLQ und so genannten Kausalvariablen. Als Indikatoren bezeichnen die Autorinnen die physische, soziale, emotionale, kognitive und Rollenfunktionsfähigkeit. Kausalvariablen umfassen Nebenwirkungen der Therapie. Diese können zu deutlichen Beeinträchtigungen führen, müssen jedoch nicht notwendigerweise vorhanden sein, wenn die GLQ schlecht ist. Zur Messung gesundheitsbezogener Lebensqualität bei Krebspatienten stehen verschiedene Instrumente zur Verfügung. Generische Lebensqualitätsinstrumente dienen dem Vergleich Krebskranker mit gesunden oder nichtkrebskranken Personen aus der Allgemeinbevölkerung. Krebspezifische Instrumenten ermöglichen Aussagen über Zusammenhänge oder Unterschiede in der Lebensqualität zwischen Betroffenen unterschiedlicher Krebserkrankungen. Mit krankheitsspezifischen Instrumenten können beispielsweise verschiedene Behandlungsstrategien bei Personen mit einheitlicher Grunderkrankung analysiert und bewertet werden. In der Literatur zeigt sich der Trend, dass spezifische Lebensqualitätsinstrumente Veränderungen in der GLQ eher abbilden. Fraglich ist hingegen, ob diese Instrumente tatsächlich immer GLQ erfassen, oder ob es sich dabei nicht lediglich um die bereits beschriebenen Kausalvariablen handelt.

Das PCa ist gegenwärtig die häufigste bösartige Neubildung beim Mann und somit von besonderer Wichtigkeit für die onkologische Rehabilitation, deren Ziel die Reduktion von somatischen, psychosozialen, aktivitätsbezogenen und teilhabeorientierten Beeinträchtigungen ist. Zur Überprüfung dieses Ziels bieten sich Maße zur Erfassung der GLQ an und finden in der Praxis bereits große Verwendung.

Eine einheitliche Aussage über die GLQ bei PCa-Patienten scheint derzeit nicht möglich. Die durchgeführten Studien unterscheiden sich hinsichtlich der Krankheitsstadien (lokal begrenzter Tumor, Metastasierung etc.) und der verglichenen Therapieformen. Trotz frühzeitiger Behandlung und guter Prognose kann die GLQ von PCa-Patienten aufgrund der verschiedenen Wirkungen und Nebenwirkungen der Behandlungen dennoch zumindest zeitweise beeinträchtigt sein. In fortgeschrittenem Erkrankungsstadium ist die Lebensqualität des Betroffenen eines der wichtigsten Behandlungsziele (Penson et al. 2003).



### 3. Response-Shift

GLQ bei Krebserkrankungen ist ein Maß, das in erster Linie über die Selbstbeurteilung der Betroffenen erfasst wird. Ausnahmen sind möglich, jedoch zeigen sich Fremdbeurteilungen häufig als nur wenig zusammenhängend mit den persönlichen Einschätzungen der Betroffenen (Norman 2003, Andresen et al. 2001). Ein Grund für die Diskrepanz mag darin liegen, dass die Betroffenen eine persönliche Bewertungsgrundlage für ihre GLQ heranziehen, die ‚von außen‘ nicht ersichtlich ist und die mehr oder weniger stark abweicht von dem, wie Gesunde ihre Lebensqualität bewerten würden (Breetvelt & van Dam 1991, Ubel et al. 2001).

Die Anwendung von Selbstbeurteilungsverfahren setzt einerseits voraus, dass die Befragten ein internalisiertes Bewertungssystem entwickelt haben, wie sie ihre eigenen Zustände und ihre Befindlichkeit auf verschiedenen Dimensionen beurteilen. Weiterhin wird in vielen wissenschaftlichen Studien davon ausgegangen, dass sich diese internalisierten Bewertungssysteme zwischen verschiedenen Gruppen (beispielsweise Interventions- und Vergleichsgruppen bei RCT) nicht unterscheiden und sich im Längsschnitt nicht verändern (beispielsweise bei Prä-Post-Messungen). Beides wird jedoch mittlerweile kritisch diskutiert (Postular et al. 2000, Sprangers 1996).

Ein Phänomen, das im Rahmen von Selbstbeurteilungen in den vergangenen Jahren immer häufiger thematisiert wurde, ist der so genannte *Response-Shift*. Darunter ist eine Veränderung des individuellen Bewertungshintergrunds für persönlich bedeutsame Lebensbereiche und Konzepte gemeint. Die Veränderung scheint besonders stark ausgeprägt zu sein, wenn sich die Person mit einer die eigene Identität bedrohenden kritischen Lebenssituation auseinandersetzen muss (Sprangers 2002). Die Diagnosemitteilung oder die Behandlung einer schwerwiegenden, lebensbedrohlichen (und zumeist chronischen) Erkrankung kann solch ein kritisches Ereignis sein.

Das Response-Shift-Phänomen ist aus methodischer Perspektive von großer Bedeutung. Wenn sich die Bewertungsgrundlagen von Personen über die Zeit verändern, muss diese Veränderung bei der Erfassung persönlich bedeutsamer Konzepte und Konstrukte – wie beispielsweise der GLQ – berücksichtigt werden (Wilson 1999). Insbesondere bei Veränderungsmessungen mit mehreren Messzeitpunkten ist zu beachten, dass sich die Bedeutung dessen, was eine Person zwei- oder mehrmals einschätzt, soweit verändert haben kann, dass die

inhaltliche Vergleichbarkeit der Zustandsmessungen in Frage zu stellen ist (Schwartz & Sprangers 2002). Rein hypothetisch können beide Messungen valide, die Differenz daraus jedoch nicht mehr quantitativ abbildbar sein (Schwartz et al. 2007).

Im Folgenden werden im Kapitel 3.1 verschiedene Modelle und Konzepte zum Response-Shift vorgestellt. Anschließend wird im Kapitel 3.2 auf Methoden zur Erfassung von Response-Shift eingegangen. Im Kapitel 3.3 werden bisher publizierte Studien zum Thema zusammengetragen und dargestellt. Im abschließenden Kapitel 3.4 werden die wichtigsten Informationen zusammengefasst.

### **3.1 Response-Shift – Definition und Modelle**

Bereits Campbell und Stanley (1963) wiesen darauf hin, dass eine mögliche Bedrohung der internen Validität bei experimentellen Studien darin zu sehen ist, dass bei Selbstbeurteilungen zwischen zwei Messzeitpunkten Veränderungen in den internen Standards der Probanden auftreten können. Cronbach und Furby (1970) sprechen in diesem Zusammenhang von einer gemeinsamen Metrik, die als wesentliche Voraussetzung einer validen Differenzbildung bei Prä-Post-Vergleichen auf der Grundlage von Selbstbeurteilungen gegeben sein muss.

Das Phänomen der veränderten Standards lässt sich einerseits mit dem Einflussfaktor ‚Instrumentation‘ in Zusammenhang bringen, der definiert wird „... as changes in the calibration of a measuring instrument or changes in raters’ standards“ (zitiert nach Howard et al. 1979, S.1). Instrumentation als Veränderung interner Standards tritt nach Ansicht von Shadish und Kollegen (2002) vor allem mit Blick auf Veränderungen über die Lebensspanne auf, so dass bei Längsschnittstudien gemessene Konstrukte für die Studienteilnehmer unterschiedliche Bedeutung haben können. Auch hier ist der zentrale Einfluss darin zu sehen, dass „... the meaning of a variable may change over time“ (Shadish et al. 2002, S. 60). Weiterhin ist denkbar, Response-Shift auch im Sinne eines Reifungseffektes zu verstehen, den Shadish und Kollegen definieren als „...natural changes that would occur even in the absence of treatment, such as growing older, [...] or [getting (eingefügt durch den Autor)] more experienced“ (ebd., S. 57).

In der Interventionsforschung wird dieses Problem umso bedeutsamer, je stärker eine durchgeführte Maßnahme direkt auf die Veränderung eines Konstrukts abzielt, das gleichzeitig als Outcome-Kriterium definiert und durch Selbstbeurteilungen erfasst wird. Ein For-

schungsfeld, in dem dieser Zusammenhang häufig besteht, ist das der Evaluation von Schulungs- und Fortbildungsmaßnahmen (Aiken & West 1990, Skeff et al. 1992).

### 3.1.1 Response-Shift nach Howard und Kollegen

Pioniere der Response-Shift-Forschung sind Howard und Dailey (1979) sowie Howard und Kollegen (1979), die den Einfluss von Fortbildungsmaßnahmen auf verschiedenen Outcome-Variablen untersuchten. Aufgrund der Analyse paradoxer Ergebnisse – beispielsweise beurteilten Fortbildungsteilnehmer von Kommunikationskursen ihren Gesprächsstil nach der Maßnahme als ‚dogmatischer‘ als vorher – unterscheiden die Autoren vier Formen eines Response-Shifts (siehe Tabelle 6).

Tabelle 6: Formen des Response-Shift nach Howard und Kollegen (1979)

<b>Response-Shift</b>	<b>Beschreibung</b>
positive expanding shift	Negativer Endpunkt der internen Skala bleibt bestehen, der positive Endpunkt schließt neue, vorher nicht berücksichtigte Aspekte mit ein
positive contracting shift	Negativer Endpunkt der internen Skala bleibt erhalten, der positive Endpunkt wird um einige Stufen reduziert
negative expanding shift	Positiver Endpunkt der internen Skala bleibt bestehen, der negative Endpunkt schließt neue, vorher nicht berücksichtigte Aspekte mit ein
negative contracting shift	Positiver Endpunkt der internen Skala bleibt erhalten, der negative Endpunkt wird um einige Stufen reduziert

Bei einem positiv erweiternden Shift ist zu beobachten, dass Personen realisieren, dass es positive Kriterien oder Ziele gibt, deren sie sich in ihren bisherigen Bewertungen nicht bewusst waren. Ein positiv reduzierender Shift würde bedeuten, dass bestimmte positiv besetzte Vorstellungen des persönlichen Bewertungsmaßstabs als unrealistisch angesehen werden und die Skala um einige positive Stufen ‚gekürzt‘ wird. Das Gleiche gilt dementsprechend für die negative Seite der Bewertungsdimension.

Unberücksichtigt nach der Definition von Howard und Kollegen (1979) bleiben Mischtypen. So werden keine Angaben gemacht über die simultane Veränderung der negativen und positiven Eckpunkte einer Dimension. Diese könnten folgende Formen annehmen (Tabelle 7):

Tabelle 7: Weitere potenzielle Formen des Response-Shift

<b>Response-Shift</b>	<b>Beschreibung</b>
positive and negative expanding shift	Negativer und positiver Endpunkt werden derart verlagert, dass an beiden Endpunkten Erweiterungen der Dimension stattfinden
positive and negative contracting shift	An beiden Seiten der Skala werden vorher enthaltene Abstufungen in der internen Skalierung nicht mehr berücksichtigt.
negative expanding and positive contracting shift	Negativer Endpunkt der internen Skala schließt neue Ereignisse mit ein, jedoch der positive Endpunkt verkürzt sich um Abstufungen
negative contracting and positive expanding shift	Positiver Endpunkt der internen Skala schließt neue Ereignisse mit ein, jedoch der negative Endpunkt verkürzt sich um Abstufungen

Obwohl die Einführung und Diskussion des Terminus Response-Shift auf Howard und Kollegen (1979) zurückgeführt wird, hat schon Upshaw (1962) ein Modell der Skalen-Rekalibrierung formuliert. Das aus der Sozialpsychologie stammende Modell bezieht sich ursprünglich auf die Bewertung sozialer Objekte und deren Veränderung über verschiedene Messzeitpunkte, beinhaltet jedoch bereits die wesentlichen Kernprozesse des Response-Shifts nach Howard. Dazu gehören drei Komponenten: (a) die Bewertung eines Objekts, (b) die psychologische beziehungsweise individuelle Bandbreite der Referenzobjekte und (c) die Unterteilung der Bandbreite in Antwortkategorien. Die Bewertung eines Objekts findet statt, indem dieses auf eine Position einer Skala verortet wird, die durch individuell definierte Endpunkte gekennzeichnet ist. Ausgehend von diesen Endpunkten wird die gesamte Dimension in (gleichgroße) Abschnitte unterteilt. Die individuelle Definition der Endpunkte ist jedoch nicht invariant, sondern durchaus veränderbar.

### 3.1.2 Alpha, Beta und Gamma Change nach Golembiewski

Golembiewski und Kollegen (1976, 1980) untersuchten verschiedene Formen der Veränderungen im Zusammenhang mit Maßnahmen zur Organisationsentwicklung. In vorangegangenen Studien der Autoren zeigten sich ‚paradoxe‘ Ergebnisse im Sinne eines Response-Shifts. Daraufhin wurden verschiedene Formen der Veränderungen als Alpha-, Beta- und Gamma-Veränderungen definiert (siehe Tabelle 8). Diese drei Arten von Veränderungen stellen eine Erweiterung des Ansatzes nach Howard und Kollegen (1979) dar, da diese über die verschiedenen Formen der Veränderung der internen Skalierung eines Konstrukts hinausgehen und auch die Möglichkeit der Neudefinition des Konstrukts beinhalten.

Tabelle 8: Formen der Veränderung nach Golembiewski und Kollegen (1976)

Form der Veränderung	Beschreibung
Alpha-Veränderung	Veränderung zwischen Zustandserfassungen, wobei der Bewertungshintergrund zu den verschiedenen Messzeitpunkten bei den Personen gleich bleibt und auch das Instrument in gleicher Weise verwendet wird = ‚wahre Veränderung‘.
Beta-Veränderung	Veränderung zwischen Zustandserfassungen, die durch eine unterschiedliche Kalibrierung des internen Standards der Person zu verschiedenen Messzeitpunkten erfolgen, wobei die Bedeutung der Kerndimension über die Zeit stabil bleibt.
Gamma-Veränderung	Veränderungen zwischen Zustandserfassungen, die durch eine Veränderung der Bedeutung des zentralen Bewertungskonstrukts erfolgen.

Die Bedeutung der Einteilung von Veränderungen in Alpha-, Beta- und Gamma-Veränderungen ist darin zu sehen, dass lediglich im Falle einer Alpha-Veränderung davon ausgegangen werden kann, dass der gefundene Differenzwert eindeutig interpretierbar ist. Im Falle einer Beta-Veränderung ist der Differenzwert zweier Messungen nicht mehr eindeutig quantitativ interpretierbar. Bei einer Gamma-Veränderung werden theoretisch zwei qualitativ unterschiedliche Konstrukte miteinander in Beziehung gesetzt.

### 3.1.3 Response-Shift nach Sprangers und Schwartz

Sprangers und Schwartz (2000) bezeichnen *Response-Shift* als ein Resultat einer Veränderung des Bedeutungsgehalts eines zentralen internen Konzepts aufgrund von drei verschiedenen Prozessen (Tabelle 9). Die Response-Shift Definition nach Sprangers und Schwartz bildet die Grundlage der vorliegenden Arbeit.

„The working definition of Response-Shift [...] refers to a change in the meaning of one’s self-evaluation of a target construct as a result of: (a) a change in the respondent’s internal standards of measurement (scale recalibration, in psychometric terms); (b) a change in the respondent’s values (i.e. the importance of component domains constituting the target construct), or (c) a redefinition of the target construct (i.e. reconceptualization)” (Sprangers & Schwartz 2000, S.12).

Sprangers und Schwartz nehmen die Definitionen von Howard und Kollegen sowie Golembiewski und Kollegen auf, wobei sie die Gamma-Veränderung unterteilen in die Repriorisierung und die Rekonzeptualisierung. Quantitativ ist die Rekonzeptualisierung die Extremform der Repriorisierung, qualitativ ist sie eine Veränderung des Bewertungskonzepts auf

einer der Reprioritisierung übergeordneten Ebene. Die Autorinnen weisen darauf hin, dass die Art der Beziehungen dieser Response-Shift-Komponenten untereinander unklar ist.

Tabelle 9: Formen des Response-Shift nach Sprangers und Schwartz (2000)

<b>Response-Shift</b>	<b>Beschreibung</b>
Rekalibrierung	Veränderung der internen Skalierung, die für die Bewertung eines Konzepts grundlegend ist, durch eine individuelle Verschiebung der Ankerpunkte einer Skala, die zur Definition des Konstrukts notwendig sind.
Reprioritisierung	Veränderung der quantitativen bzw. relativen Bedeutsamkeit einzelner Werte bzw. Ziele eines Menschen, die dieser für wesentlich im Zusammenhang mit einem zentralen Lebensqualitätskonstrukt hält.
Rekonzeptualisierung	fundamentale bzw. qualitative Veränderung im Wertesystem einer Person i. d. S., dass eine vorher relevante Dimension (Werte bzw. Ziele) ihre Bedeutung gänzlich verliert oder eine vorher nicht bedeutsame Dimension neu hinzukommt.

Response-Shift kennzeichnet das Ausmaß der (intraindividuellen) Instabilität von Konzepten im Zusammenhang mit Bewertungsprozessen im Zeitverlauf. Sprangers und Schwartz (1999) formulieren ein *Prozessmodell von Response-Shift und Lebensqualität* (Abbildung 2). Dieses Modell setzt sich aus fünf verschiedenen Bestandteilen zusammen (deutsche Übersetzung Güthlin 2004): Katalysator (1), Vorbedingungen (2), Mechanismen (3), Response-Shift (4) und wahrgenommene Lebensqualität (5). Diese Bestandteile werden wie folgt in dem Modell zur Veränderung der GLQ integriert: Ein Katalysator ist ein Ereignis, das eine Auseinandersetzung mit der GLQ fördert, wie beispielsweise eine Veränderung des Gesundheitszustandes durch eine Krebsdiagnose oder -behandlung. Die Wahrnehmung der GLQ wird einerseits durch bestimmte Merkmale der Person mitbestimmt, wie beispielsweise durch das Alter, den Familienstand und andere soziodemographischen Variablen sowie persönliche Eigenschaften etc. Andererseits wendet die betroffene Person Strategien an, um sich mit der verändernden Lebenssituation auseinander zu setzen. Zu diesen Mechanismen oder Strategien zählen in erster Linie Bewältigungsprozesse wie handlungsbezogenes, emotionales und kognitives Coping. Weiterhin werden im Zusammenhang mit kritischen gesundheitlichen Lebensereignissen noch soziale Vergleichsprozesse oder verschiedene Möglichkeiten der Ressourcenaktivierung explizit hervorgehoben (Spiritualität, Soziale Unterstützung). Die persönlichen Vorbedingungen wirken dem Modell zufolge einerseits ‚direkt‘ auf den Response-Shift, andererseits aber auch ‚indirekt‘, vermittelt über die von der Person genutzten Strategien. Der Response-Shift führt in Form der oben genannten drei Veränderungsprozesse des Bewertungshintergrunds schließlich zu einer Veränderung der wahrgenommenen Lebensqualität.

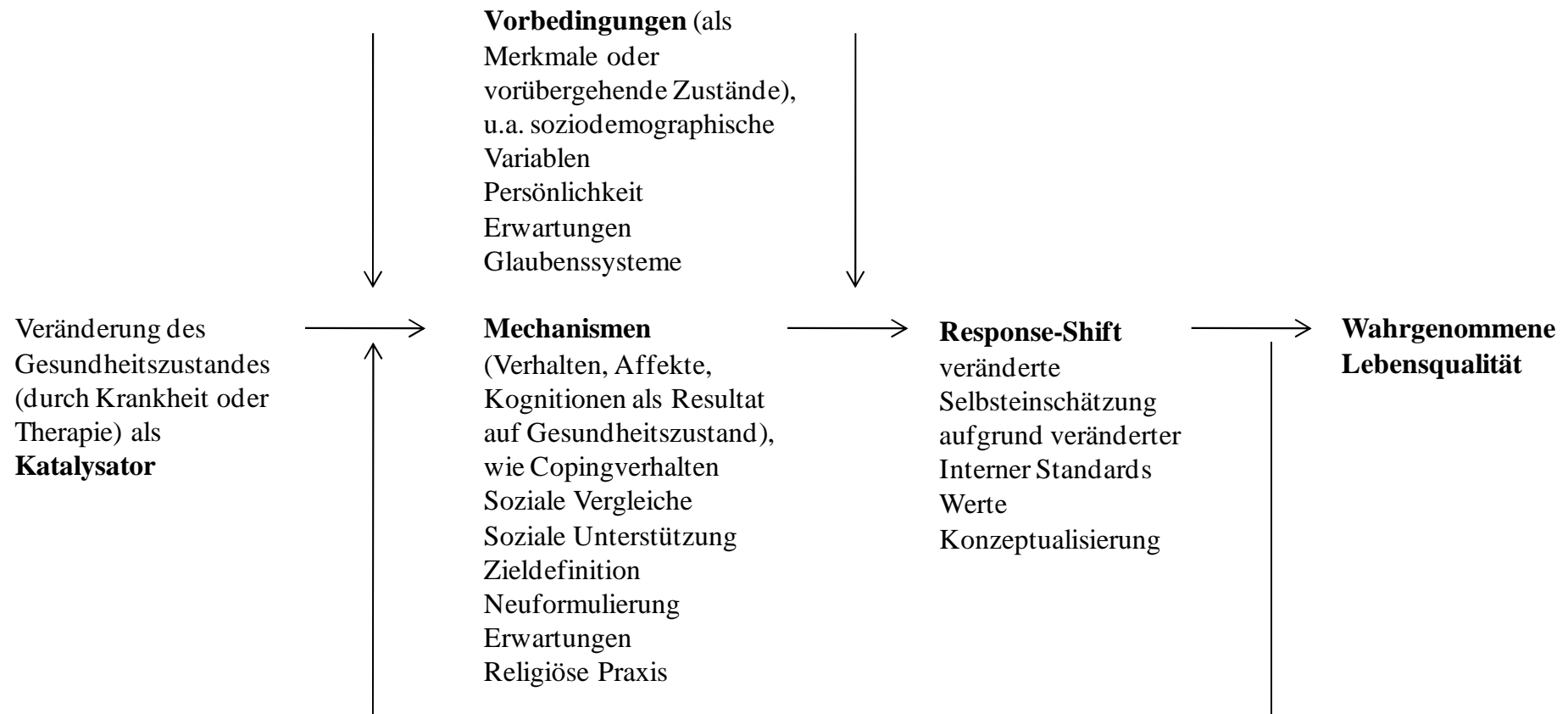


Abbildung 2: Prozessmodell von Response-Shift und Lebensqualität nach Sprangers und Schwartz (1999, Übersetzung von Gütlin 2004)

Aus den Modellparametern Katalysator, Vorbedingungen, Mechanismen und wahrgenommene Lebensqualität lassen sich auch die zentralen Fragen ableiten, die zum Response-Shift-Phänomen noch weitgehend ungeklärt sind.

*Katalysator:* Response-Shift tritt um so eher auf, je kritischer ein Lebensereignis ist, mit dem die Person konfrontiert wird (Sprangers und Schwartz 2000). Unklar ist, wann ein Ereignis tatsächlich response-shift-relevant ist und wann nicht. Faktoren, die einen Response-Shift beeinflussen können sind unter anderem die Dauer eines Ereignisses oder der Auseinandersetzung mit diesem, dessen Schweregrad und die Gewichtung (positiv vs. negativ) sowie die Spontaneität beziehungsweise ‚Plötzlichkeit‘ des Auftretens. Weiterhin wurde darauf hingewiesen, dass Response-Shift vor allem im Rahmen von Selbstbeurteilungen relevant ist. Diese Selbstbeurteilungen betreffen nicht nur die wahrgenommene Lebensqualität als Response-Shift-Resultat, sondern auch die auslösenden Bedingungen. Wann ein Ereignis im konkreten Einzelfall als Katalysator fungiert, ist wiederum abhängig von individuellen Faktoren.

*Vorbedingungen:* Unklar ist bisher, welche personalen Faktoren für einen Response-Shift bedeutsam sind. Angenommen wird, dass soziodemographische Faktoren wie Alter, Geschlecht, Bildung etc. einen Response-Shift beeinflussen. Persönlichkeitseigenschaften beeinflussen die Auseinandersetzung mit einem Ereignis und deshalb wird angenommen, dass dies auch auf Response-Shift zutrifft. Es ist jedoch weitgehend ungeklärt, wie welche Persönlichkeitseigenschaften in diesem Zusammenhang wirken. Sprangers und Schwartz (2000) nehmen an, dass positive Affektivität und damit verbundene Eigenschaften wie beispielsweise Optimismus und ein hohes Selbstwertgefühl eher response-shift-förderlich sind, wohingegen eine negative Affektivität, Depressivität oder eher externale Kontrollüberzeugungen das Auftreten eines Response-Shift vermindern könnten.

*Mechanismen:* Eine Palette von Faktoren wird als potenzielle Mechanismen genannt, die im Zusammenhang mit Response-Shift zu beachten sind. Unklar ist, welche Prozesse und Strategien entscheidend sind und wie diese miteinander interagieren. Hervorgehoben wird die Bedeutung von Copingprozessen (Richards & Folkman 2000, Folkman & Greer 2000). Welche verschiedenen Ebenen von Bewältigungsmechanismen (kognitiv, emotional, handlungsbezogen) betroffen sind und welche Ebene von größerer Bedeutung für einen Response-Shift ist, wird aktuell diskutiert.



*Wahrgenommene gesundheitsbezogene Lebensqualität:* Offen ist die Frage, ob es bestimmte Bereiche oder Komponenten der Lebensqualität gibt, die anfälliger für einen Response-Shift sind als andere. Nach Studienergebnissen von Haes und Kollegen (1992) ist anzunehmen, dass kognitive Prozesse (beispielsweise Zufriedenheitseinschätzungen) stärker response-shift-beinträchtigt sind als affektive (beispielsweise Stimmung). Weiterhin wird postuliert, dass Response-Shift vor allem in den Lebensqualitätsbereichen auftritt, die stärker subjektiv geprägt sind, wie zum Beispiel die Wahrnehmung und Einschätzung von Schmerzen, Fatigue oder sozialen Beziehungen, wohingegen eher ‚objektivierbare‘ Bereiche, wie zum Beispiel die funktionale Kapazität, weniger starken Verzerrungen unterliegt. Ebenfalls beachtenswert in diesem Zusammenhang ist die Formulierung der Items, anhand derer Lebensqualität erfasst werden soll. Eher verhaltensnahe Itemformulierungen (Häufigkeit von Tätigkeiten) sollen weniger response-shift-anfällig sein als eher affektiv- oder kognitiv-bewertungsorientierte Items (Belastung durch Erkrankung bzw. Zufriedenheit mit gesundheitlichen Kapazitäten). In diesem Zusammenhang unterscheiden Schwartz und Rapkin (2004) drei verschiedene Beurteilungsebenen: eine *performanz-basierte* Beurteilung (‚performance-based‘), bei der Leistungen von Personen geprüft werden (beispielsweise Treppensteigen oder Zeitmessung bei bestimmter Gehstrecke), wobei das Urteil und die Messung der Leistung zwei unterschiedliche Dinge sind und die tatsächliche Leistung nur durch einen Fehler in der Messung verzerrt werden kann. Die *wahrnehmungs-basierte* Beurteilung (‚perception-based‘) bezieht sich auf Leistungen oder Handlungen, die im Alltag durchgeführt werden und über die Angaben gemacht werden sollen (beispielsweise „Wie oft steigen Sie Treppen?“). Das Urteil kann Einflüssen unterliegen, die eher als ‚Response-Bias‘ zu bezeichnen und darauf zurückzuführen sind, dass zum Beispiel verschiedene Beurteiler mit unterschiedlichen Standards messen, wobei aber allgemeingültige, objektivierbare Standards existieren, auf die die Urteiler ‚geeicht‘ werden können. Davon zu unterscheiden sind so genannte *evaluations-basierte* Beurteilungen (‚evaluation-based‘), denen subjektive Standards zugrunde liegen (beispielsweise Beurteilung der *Schwierigkeit*, Treppen zu steigen). Bei diesem Urteilsprozess gelten individuelle Bezugspunkte, die nur schwer generalisierbar sind. Beurteilungen des zuletzt genannten Typs unterliegen nach Schwartz und Rapkin (2004) einem Response-Shift.

Kritisch zum Modell von Sprangers und Schwartz ist die Zirkularität des Response-Shift-Modells und die damit verbundenen unscharfen Trennungen zwischen Modellbestandteilen anzumerken. Einerseits wird die Repriorisierung als Response-Shift-Komponente definiert, die durch Ziel- oder Werteveränderungen gekennzeichnet ist, andererseits sind Veränderun-

gen persönlich bedeutsamer Ziele auch Prozesse kognitiver Umstrukturierungen und somit auch gleichzeitig Mechanismen. Wiederum unklar ist die genaue Definition von Response-Shift als Prozess oder Ergebnis einer Auseinandersetzung mit Anforderungen. In diesem Zusammenhang ist ebenfalls darauf hinzuweisen, dass im Modell keine Informationen darüber enthalten sind, ob es eine Rangfolge im Ablauf eines Response-Shifts gibt.

Da in der vorliegenden Arbeit Einflussfaktoren im Sinne von Moderatorvariablen auf Response-Shift geprüft werden, soll auf diese überblicksartig eingegangen werden. Im Folgenden werden die potenziellen Moderatoren beschrieben und in den Kontext der Lebensqualitätsforschung bei PCa gestellt. Dabei sollen hier die Variablen positive und negative Affektivität sowie Selbstwirksamkeit vorgestellt werden. Auf Benefit Finding wird in Kapitel 3.1.6 ausführlich eingegangen. Es handelt sich dabei um eine mögliche inhaltliche Erklärung für einen Response-Shift beziehungsweise ein Konstrukt, das Response-Shift sehr nahe steht.

#### *Exkurs: Positive und negative Affektivität*

In der Bewältigungsforschung ist eine zentrale Aufgabe die Analyse der persönlichen Auseinandersetzung mit verschiedenen emotionalen Zuständen (Krohne et al 1996). Eine Voraussetzung dafür ist die genaue Definition und Operationalisierung emotionalen Befindens und emotionaler Prozesse als stabile Persönlichkeitsmerkmale. Neben Ansätzen der Definition verschiedener spezifischer Basisemotionen (bspw. Ekman 1982, Izard 1991) postulierten Watson und Tellegen (1985) ein Modell, in dem emotionale Zustände auf zwei voneinander unabhängigen Dimensionen verortet sind. Bei diesen zwei Dimensionen handelt es sich um positive und negative Affektivität. Watson und Kollegen definieren *positive Affektivität* (PA) wie folgt: „... (PA) reflects the extent to which a person feels enthusiastic, active, and alert. High PA is a state of feel high energy, full concentration, and pleasurable engagement, ...” (Watson et al. 1988a, S.1063). Neben der Beschreibung dieser positiven emotionalen Zustände eignet sich das Konstrukt ebenso zur Operationalisierung stabiler individueller Unterschiede in der Affektivität über die Zeit. Positive Affektivität lässt sich somit auch als Disposition bezeichnen, die positiv korreliert mit Eigenschaften wie beispielsweise Extraversion (Tellegen 1985). *Negative Affektivität* (NA) wird von Watson und Kollegen entsprechend folgendermaßen beschrieben: „... (NA) is a general dimension of subjective distress and unpleasurable engagement that subsumes a variety of aversive mood states, including anger, contempt, disgust, guilt, fear, ...” (Watson et al. 1988a, S. 1063). Auch negative Affektivität

ist neben der Beschreibung von Zuständen zur Erfassung stabiler Persönlichkeitsmerkmale geeignet. Dabei gehen die Autoren davon aus, dass diese positiv zusammenhängt mit Neurotizismus beziehungsweise Ängstlichkeit. Negative Affektivität erfasst eine allgemeine Dimension von Distress (Watson et al. 1988b). Andererseits haben Watson und Pennebaker (1989) festgestellt, dass negative Affektivität positiv korreliert mit Symptombelastungen: Je mehr negative Affektivität eine Person schildert, desto mehr körperliche Beschwerden werden wahrgenommen. Für positive Affektivität konnte kein Zusammenhang mit Symptombelastungen nachgewiesen werden. Die Autoren diskutieren drei verschiedene Annahmen bezüglich der Beziehung zwischen negativer Affektivität und körperlichen Beschwerden: die *psychosomatische Hypothese*, wobei negative Affektivität als Ursache für gesundheitliche Probleme bezeichnet wird, die *Disability-Hypothese*, wobei die gesundheitlichen Beeinträchtigungen zu einem erhöhten Ausmaß an negativer Affektivität führt, sowie die *Symptom-Perception-Hypothese*, die annimmt, dass Personen mit erhöhter negativer Affektivität sensitiver sind bezüglich der Wahrnehmung und Äußerung von körperlichen Prozessen. Während die ersten beiden Hypothesen von den Autoren aufgrund der Ergebnisse einer Literaturübersicht keine Bestätigung finden, wird die dritte Hypothese als zumindest teilweise bestätigt bezeichnet. Positive Affektivität korreliert Watson und Pennebaker (ebd.) zufolge stärker mit sozialen Ereignissen und Beziehungen.

Kressin und Kollegen (2000) untersuchten an drei unterschiedlichen (nicht krebserkrankten) Stichproben die Beziehung zwischen negativer Affektivität und GLQ (SF-36). Es ließ sich ein statistisch signifikanter Zusammenhang zwischen negativer Affektivität und der Psychischen Summenskala feststellen. Negative Affektivität und die Körperliche Summenskala korrelierten nicht miteinander.

Blank und Belizzi (2006) kommen aufgrund regressionsanalytischer Auswertungen zu dem Schluss, dass ein hohes Maß an subjektivem Wohlbefinden von Männern mit PCa beeinflusst wird durch dispositionellen Optimismus. Belizzi und Blank (2007) wiesen anhand einer Querschnittstudie mit 490 PCa-Patienten ein bis acht Jahre nach Diagnosestellung nach, dass die Attribution „Survivor“ und „Someone who has conquered cancer“ (Bellizi & Blank 2007, S. 44) mit einem höheren Ausmaß an positiver Affektivität korrelieren, als andere Attributionen (bspw. „Patient“). Roesch und Kollegen (2005) gehen aufgrund der Ergebnisse einer Metaanalyse zu Bewältigungsstrategien bei PCa-Patienten davon aus, dass problemorientiertes Coping einhergeht mit einem erhöhten Ausmaß an positiver Affektivität und GLQ. Prob-

lemorientierte Bewältigung war negativ korreliert mit Ängstlichkeit und Depressivität. Wrosch und Scheier (2003) sehen in dispositionellem Optimismus einen wesentlichen Einflussfaktor auf die allgemeine Lebensqualität, wobei diese Beziehung einerseits dazu führt, dass problemorientiertes Coping in angemessener Weise eingesetzt wird, wenn dies der gegenwärtigen Erkrankungssituation entspricht. Optimismus und allgemeine Lebensqualität können jedoch auch in einer positiven Beziehung stehen, wenn emotionsorientiertes Coping eingesetzt wird. Das ist dann der Fall, wenn Zieländerungen (i.S. einer Repriorisierung) vorzunehmen sind, da aktives Problemlösen keine adäquate Bewältigungsstrategie mehr ist.

Voogt und Kollegen (2005) untersuchten positive und negative Affektivität in einer Stichprobe mit heterogenen Krebserkrankungen. Insgesamt zeigten die Personen in einer fortgeschrittenen Erkrankungsphase weniger positive Affektivität als die gesunde Normalbevölkerung. Es ließen sich jedoch keine Unterschiede zwischen diesen beiden Gruppen hinsichtlich des Ausmaßes der negativen Affektivität feststellen. Prädiktoren für ein geringeres Maß an positiver Affektivität waren der Studie zufolge: männliches Geschlecht, höheres Alter, weniger problemorientierte Bewältigungsstrategien, weniger Sinnfindung sowie die Planung einer chirurgischen Behandlung. Prädiktoren für höhere Werte der negativen Affektivität waren: Personen mit einer Gesundheitsversicherung, die Planung eines chirurgischen Eingriffs und ein geringes Ausmaß an Sinnfindung. Insgesamt waren ein geringes Ausmaß an positiver Affektivität sowie ein hohes Ausmaß an negativer Affektivität korreliert mit höheren Ängstlichkeits- und Depressivitätswerten.

#### *Exkurs: Selbstwirksamkeit*

Neben den aus dem Modell nach Sprangers und Schwartz abgeleiteten Vorbedingungen positive Affektivität und negative Affektivität wird mit der vorliegenden Arbeit die Bedeutung der Mechanismus-Variable *Allgemeine Selbstwirksamkeit* untersucht. Diese wird definiert als positive Beurteilung eigener Kompetenzen, um schwierige, neue oder anforderungsreiche Aufgaben kontrollieren und bewältigen zu können (Bandura 1997, Schwarzer & Jerusalem 1995, Luszczynska et al. 2005). Personen mit einer hohen Selbstwirksamkeitserwartung suchen vermehrt herausfordernde Situationen, mit denen sie sich länger und intensiver beschäftigen als Personen mit einer geringen Selbstwirksamkeit.

Nach Bandura (1997) besteht ein positiver Zusammenhang zwischen einem ausgeprägten Maß an Selbstwirksamkeit, einem hohen Engagement in gesundheitsförderlichem Verhalten sowie in Bestrebungen im Krankheitsfall möglichst schnell zu genesen. Die mit einer hohen Selbstwirksamkeit einhergehenden positiven Überzeugungen bezüglich der Einschätzung, genügend Kompetenzen zum Umgang mit schwierigen Situationen zu besitzen, korreliert positiv mit positiven Affekten sowie negativ mit Gefühlen wie Ängstlichkeit. In Belastungssituationen erleben selbstwirksame Personen ein geringeres Maß an negativen Emotionen, die mit Stress und Depressivität in Zusammenhang stehen. Außerdem zeigen sie mehr Problemorientierung und angemessene kognitive Bewältigungsprozesse und damit einhergehend weniger intrusive Gedanken sowie Grübeln. Cunningham und Kollegen (1991) bestätigen eine positive Korrelation zwischen Selbstwirksamkeit und allgemeiner Lebensqualität. Studien zur Untersuchung der Beziehungen zwischen Selbstwirksamkeit und positiver Befindlichkeit (Optimismus) sowie allgemeiner Lebensqualität verweisen auf einen positiven Zusammenhang zwischen den Konstrukten (Bandura 1997, Schwarzer 1992, Luszczynska et al. 2005). Ein negativer Zusammenhang zwischen Selbstwirksamkeit und negativer Affektivität sowie Hilflosigkeit wird ebenfalls berichtet.

In der Auseinandersetzung mit Krebserkrankungen gilt eine hohe Selbstwirksamkeit als protektiver Faktor in der Bewältigung der Erkrankung (Lev 1997). Weber und Kollegen (2004) fanden einen negativen Zusammenhang zwischen Selbstwirksamkeit und Depressivität bei PCa-Patienten. Northouse und Kollegen (2002) sehen in der wahrgenommenen Selbstwirksamkeit den wichtigsten Einflussfaktor auf die psychische Befindlichkeit bei Brustkrebspatientinnen. Frauen mit einer deutlich ausgeprägten Selbstwirksamkeit sind hoffnungsvoller, zeigen weniger maladaptive Bewältigungsstrategien und eine bessere Lebensqualität. Böhmer und Kollegen (2007) wiesen eine positive direkte Beziehung zwischen wahrgenommener Selbstwirksamkeit und physischer, sozialer und emotionaler Befindlichkeit als Komponenten der Lebensqualität bei einer heterogenen Stichprobe von Krebspatienten nach chirurgischen Eingriffen nach. Graves (2003) publizierte eine Metaanalyse zur Effektivität von Interventionsprogrammen bei Patientinnen und Patienten mit Krebserkrankungen. Die Autorin kam zu dem Schluss, dass Maßnahmen, die sozial-kognitive Ausrichtungen mit dem Ziel einer gesteigerten Selbstwirksamkeit und der Förderung einer positiven Behandlungserwartung anstreben, bei den Teilnehmern zu einer Verbesserung der positiven Affektivität, der psychischen Befindlichkeit, sozialen Lebensqualität und objektiver physischer Parameter führten. Maßnahmen mit geringerer sozial-kognitiver Ausrichtung wiesen deutlich kleinere Effekte auf.

Die genannten Befunde zum Zusammenhang zwischen Selbstwirksamkeitserwartung und Auseinandersetzung mit Krebserkrankungen sprechen dafür, diese Ressource als Mechanismus-Variable in das Modell von Sprangers und Schwartz (1999) aufzunehmen.

#### 3.1.4 Erweiterung des Modells nach Sprangers durch Rapkin und Schwartz

Rapkin und Schwartz (2004) erweitern das Modell von Sprangers und Schwartz (2000), indem die Bedeutung von Bewertungsprozessen hervorgehoben wird. Die Aktivierung eines individuellen Referenzrahmens für die Beantwortung eines Items zur GLQ wird beeinflusst von der Bedeutung, die das Item für die beantwortende Person besitzt, sowie den Anforderungen durch die Befragungssituation selbst. Es geht dabei um das allgemeine Verständnis der Frage. Die individuelle Gestaltung des Referenzrahmens ist dabei umso größer, je globaler eine Frage formuliert wird. Daraus folgt jedoch, dass die aktivierten inhaltlichen Bereiche oder Kategorien, die eine Person zur Beantwortung der Frage heranzieht, umso variabler und in ihrer Anzahl interindividuell sehr verschieden sein können. Beispielhaft sei hier auf eine Studie von Rapkin (2000) verwiesen. Bei der Analyse von persönlich bedeutsamen Zielen bei Menschen mit HIV zeigte sich, dass die Zielkombinationen der einzelnen Befragten vollkommen unabhängig waren. „In other words, individuals who were likely to mention goals in one category were no more or less likely to mention goals in other content categories. This variety in content and patterning of personal goals speaks to the many frames of references that people use to appraise their QOL” (Rapkin 2000, S.61).

Rapkin und Schwartz (2004) verweisen auf vier verschiedene Prozesse, die bei der Beantwortung von Fragebogenitems zur GLQ nach dem Response-Shift-Modell bedeutsam sind: die Einführung eines Bewertungshintergrunds im Sinne eines Referenzrahmens (1), die Auswahl relevanter Erfahrungen (2), das Setzen und Aktivieren persönlich wichtiger Vergleichsmaßstäbe (3) sowie ein subjektiver Algorithmus, der die Gewichtung dieser Prozesse miteinander in Beziehung setzt (4). Diese vier Prozesse sind angelehnt an Modelle der kognitiven Informationsverarbeitung, die entwickelt wurden, um zu beschreiben, wie Personen Informationen aus Fragebogenitems verarbeiten und Antworten auf diese Fragen generieren (Jobe 2003). Auf die genannten Strategien wird im Folgenden nochmals genauer eingegangen:

Der *Referenzrahmen* (1): Innerhalb des idiosynkratischen Bezugsrahmens werden durch Auswahlstrategien bestimmte Erinnerungen, Vorstellungen und Annahmen aktiviert. Diese

ergeben sich aus einem Zusammenspiel biographischer Aspekte und wiederum ebenfalls der Itemformulierung, der allgemeinen Fragebogengestaltung und der Untersuchungssituation.

*Die Auswahl relevanter Erfahrungen (2):* Bedeutsam in diesem Zusammenhang ist, dass diese Strategien, anhand derer die Person Erfahrungen auswählt, über die Zeit variabel sind, so dass selbst bei stabilem Bezugsrahmen im Verlauf der Zeit verschiedene Auswahlstrategien zu unterschiedlichen Bewertungen von ein und demselben Item führen können.

*Aktivieren eines persönlichen Vergleichsmaßstabs (3):* Die ausgewählten Erfahrungen werden mit subjektiven Standards verglichen. Persönliche Referenzpunkte werden aktiviert und mit dem verstandenen Iteminhalt in Beziehung gesetzt. Dies kann beispielsweise geschehen, indem frühere Funktionsfähigkeiten in verschiedenen Bereichen berücksichtigt und das Ausmaß an aktuellen Einbußen oder die Veränderung in bestimmten Lebensbereichen bewusst herangezogen oder Extremerfahrungen als Vergleichsmaßstäbe aktiviert werden. Die Aktivierung eines solchen Vergleichsmaßstabs beruht auf Selektionsprozessen, die lediglich eigene beziehungsweise individuelle Erfahrungen heranziehen. Denkbar sind jedoch auch soziale Vergleiche wie die Beobachtung anderer Patienten mit der Konsequenz eines sozialen Ab- oder Aufwärtsvergleichs oder auch die Ableitung eines Maßstabs aufgrund der Interaktion mit Behandlungspersonal und deren kommunizierter Ziele und Erwartungen. Inhalt und Höhe der Standards können – je nach persönlicher Situation – über die Zeit variieren.

*Anwendung eines individuellen Algorithmus (4):* Letztlich ist eine Verknüpfung der verschiedenen Punkte nötig, um die entsprechenden Prozesse zu integrieren und eine Antwort auf die Selbstbeurteilung zu formulieren. Dabei werden bei jeder Einschätzung eines Items Kombinationen aus Bewertungsgrundlagen und Erfahrungen vorgenommen, wobei diese je nach aktueller Bedeutung für die Person von dieser unterschiedlich gewichtet werden.

### 3.1.5 Buffer-Modell nach Lepore und Eton

Lepore und Eton (2000) stellen zwei unterschiedliche Modelle zur Wirkung von Response-Shift im Zusammenhang mit GLQ vor. Ausgangspunkt ist das Modell von Sprangers und Schwartz (2000), nach dem Response-Shift einerseits als Moderatorvariable auf die Beziehung zwischen Katalysator und gesundheitsbezogener Lebensqualität fungieren kann. Andererseits kann Response-Shift auch als Mediatorvariable zwischen dem Katalysator und der

GLQ wirken. Das Moderatormodell wird von Lepore und Eton (2000) als Buffer-Modell, das Mediatormodell als Suppressor-Modell bezeichnet.

Nach dem *Buffer-Modell* (Lepore & Eton 2000) wird Response-Shift als Moderatorvariable eingestuft, die Einfluss nehmen kann auf die Beziehung zwischen den gesundheitlichen Problemen und der Lebensqualität einer Person. In diesem Modell wird davon ausgegangen, dass Response-Shift wiederum unter dem Einfluss weiterer Faktoren wie dem Ausmaß sozialer Unterstützung oder verschiedener Bewältigungsstile steht. Gemäß der Funktion eines Moderators ist zu erwarten, dass je nach Ausprägung verschiedener Einflussfaktoren bei entsprechenden Personen ein Response-Shift auftritt und bei anderen wiederum nicht. Im Moderatormodell ist hinsichtlich korrelativer Zusammenhänge anzunehmen, dass Personen mit einer Verschlechterung im Gesundheitszustand keine negativen Veränderungen der GLQ angeben, wenn bei ihnen ein Response-Shift stattfindet. Bei Personen jedoch, die keinen Response-Shift zeigen, sollte sich ein direkter Zusammenhang zwischen Gesundheitsveränderung und GLQ zeigen. Als Beispiel für den Einfluss von Response-Shift als Moderatorvariable in Form der Rekalibrierung (Veränderung der internen Standards) beschreiben die Autoren die Bedeutung von sozialen Abwärtsvergleichen („downward social comparison“). PCa-Patienten, die Personen erleben, denen es gesundheitlich viel schlechter geht als ihnen selbst, erweitern ihre eigene Skala für (beispielsweise) körperliche Funktionsfähigkeit um einige negative Einheiten (negative expanding Response-Shift nach Howard et al. 1979). Dies hat zur Folge, dass die Personen ihre GLQ trotz eigener Einbußen im Bereich der körperlichen Funktionsfähigkeit noch immer in einem mittleren Bereich angeben. Personen ohne sozialen Abwärtsvergleich hingegen würden nach gesundheitlichen Einbußen ihre körperliche Funktionsfähigkeit deutlich schlechter einschätzen. Im Falle eines Response-Shift in Form einer Repriorisierung (quantitative Veränderung des Wertesystems) ist zu erwarten, dass bei Personen, deren Wertesystem sich verändert, keine signifikante Beziehung zwischen gesundheitlicher Veränderung und GLQ aufgedeckt werden kann, da beispielsweise die Wertigkeit der erlebten Einbuße als persönlich weniger bedeutsam für die GLQ eingeschätzt wird.

Im *Suppressor-Modell* gilt Response-Shift als Mediatorvariable zwischen gesundheitlichen Beschwerden und Lebensqualität. Die Beziehung zwischen den beteiligten Variablen ist demnach folgende: Gesundheitliche Probleme verursachen einen Response-Shift, d.h. eine Veränderung der internen Standards (Rekalibrierung) und/oder des Wertesystems (Repriorisierung). Der Response-Shift wiederum führt zu einer veränderten Bewertung der individuel-



len Lebensqualität. Eine Annahme, die mit diesem Modell verbunden ist, lautet, dass Response-Shift immer dann auftritt, wenn gesundheitliche Probleme und Beschwerden auftreten. Diese Annahme lässt sich noch in verschiedene weitere Annahmen zerlegen: (a) Je extremer das Ausmaß gesundheitlicher Veränderung, desto intensiver ist der Response-Shift, (b) Response-Shift beeinflusst die wahrgenommene GLQ positiv und (c) wenn die Beziehung zwischen gesundheitlicher Veränderung und wahrgenommener GLQ gemessen wird und gleichzeitig der Response-Shift statistisch kontrolliert wird (i.S. einer Partialkorrelation), sollte die Beziehung zwischen Gesundheit und Lebensqualität positiv sein. Anhand einer Rekalibrierung und einer Repriorisierung wird dieser Zusammenhang wie folgt beschrieben: PCa-Patienten bewerten ihre Ängstlichkeit vor der Diagnosestellung als moderat, erleben während der Diagnose und Behandlung des Prostatakrebs jedoch eine starke Angst und bewerten im Anschluss an die Behandlung ihre Ängstlichkeit bedingt durch Therapie- und Nebenwirkungen wiederum als moderat, wobei sich ihre interne Skala jedoch durch das Erleben der starken Angst verschoben hat und sie ihre Ängstlichkeit vor Diagnosestellung zum späteren Zeitpunkt nicht mehr als moderat, sondern als gering einschätzen würden. Dieser Prozess findet nach dem Suppressormodell bei jedem PCa-Patienten statt. Dies gilt ebenso für den Repriorisierungsprozess. Wenn dieser normativ im Zusammenhang mit gesundheitlichen Veränderungen auftritt, dann wäre dies ein Beleg für das Suppressor- beziehungsweise Mediatormodell.

Der Modellvergleich erbrachte jedoch mehr Hinweise für die Annahme des Buffer- beziehungsweise Moderatormodells. Die Überprüfung des Suppressormodells zeigte keine statistisch signifikanten Zusammenhänge zwischen den gesundheitlichen Veränderungen und dem Ausmaß an Response-Shift. Weiterhin waren Veränderungen in der GLQ unkorreliert mit Veränderungen der gesundheitlichen Beschwerden, unabhängig davon ob ein Response-Shift statistisch kontrolliert wurde oder nicht. Zur Bestätigung des Buffer- oder Moderatormodells konnten statistisch bedeutsame Interaktionseffekte für gesundheitliche Beeinträchtigungen und Response-Shift-Variablen aufgezeigt werden: Im Falle sich verschlechternder urinaler Funktionsfähigkeit resultierte eine verbesserte Lebensqualität, wenn die Betroffenen ihr Wertesystem veränderten.

Eine Bestätigung erhält das Buffermodell durch eine Untersuchung von Hagedoorn und Kollegen (2002). Bei einer Gruppe mit heterogenen Krebserkrankungen wurde zu zwei Messzeitpunkten Daten zur allgemeinen GLQ und zu spezifischen Lebensqualitätsdomänen (Physische und Emotionale Lebensqualität) erfasst. Außerdem wurden die Befragten gebeten, Aus-

sagen im Sinne eines sozialen Vergleichsprozesses zu treffen. Im Ergebnis zeigte sich, dass diejenigen Krebspatienten, die eine statistisch signifikante Abnahme in der Physischen Lebensqualität zeigten, nur dann eine signifikante Reduzierung der allgemeinen GLQ aufwiesen, wenn sie zusätzlich angaben, dass es ihnen im Vergleich zu anderen Personen in einer vergleichbaren Situation schlechter gehe. Personen mit einer Verschlechterung in der Physischen Lebensqualität zeigten demnach stabile allgemeine Lebensqualität, wenn es ihnen ihrer eigenen Einschätzung nach nicht schlechter erging als anderen Personen in einer ähnlichen Lebenssituation.

Zwei wesentliche Ergebnisse bestätigen somit das Buffermodell von Lepore und Eton: Zum Einen findet nicht bei jedem untersuchten Krebspatienten, bei dem eine Verschlechterung der physischen Funktionsfähigkeit festgestellt wurde, auch ein Response-Shift statt. Zum Anderen werden soziale Abwärtsvergleiche als response-shift-fördernde Faktoren im Buffermodell beschrieben. In der von Hagedoorn und Kollegen (2002) durchgeführten Studie konnte diesbezüglich ein mit den Angaben von Lepore und Eton übereinstimmendes Ergebnis für die Beziehung zwischen gesundheitlichen Beeinträchtigungen, Response-Shift und GLQ nachgewiesen werden.

### 3.1.6 Dem Response-Shift nahestehende Konzepte und Konstrukte

Obwohl Response-Shift bereits seit gut 40 Jahren als messmethodisches Phänomen bekannt ist, gibt es kaum Modelle und Theorien, die das Konstrukt berücksichtigen. Ein Grund ist nach Güthlin (2004) darin zu sehen, dass Response-Shift inhaltlich eng mit verschiedenen anderen Konstrukten in Zusammenhang gebracht wird. Zu diesen Konstrukten zählen Selbstregulation, Coping und Benefit Finding, die für die Response-Shift Thematik von besonderer Bedeutung sind. Diese werden in den folgenden Abschnitten skizziert. Bei allen Konstrukten sind die Art der Auseinandersetzung und die Anpassung an sich ändernde Lebensbedingungen von zentraler Bedeutung.

#### *Selbstregulation*

Modelle der Selbstregulation betrachten die Integration von handlungsbezogenen, kognitiven und emotionalen Prozessen im Rahmen der Anpassung an sich verändernde Lebensbedingungen (Carver & Scheier 1982, Heckhausen & Schulz 1995, Lachman & Burack 1993). Es werden verschiedene dieser Modelle mit Bezug zum Response-Shift vorgestellt. Dabei

handelt es sich einerseits um ein Modell nach Brandstädter und Renner (1990) mit einer Unterscheidung zwischen assimilativen und akkomodativen Anpassungsstrategien, ein Modell von Heckhausen und Schulz (1995) mit primären und sekundären Kontrollprozessen sowie ein Modell der adaptiven Selbstregulation nach Carver und Scheier (2000).

Brandstädter und Renner (1990) unterscheiden in ihrem Modell der personalen Selbstregulation zwei verschiedene Strategien zur Optimierung der Balance zwischen Gewinn und Verlust im individuellen Entwicklungsverlauf. Das gilt somit auch für die Auseinandersetzung mit kritischen Lebensereignissen wie einer Krebserkrankung. Besteht eine Diskrepanz zwischen der aktuellen und der von der Person gewünschten Entwicklung, existieren zwei Strategien, diese Diskrepanz zu minimieren. *Assimilation* führt dazu, dass die Person ihre Lebensbedingungen und -umstände so verändert, dass die gewünschte Entwicklung eintritt oder die Wahrscheinlichkeit, diese zu erreichen, erhöht wird. Persönliche Präferenzen, Bewertungskonzepte und Ziele bleiben bei diesem Prozess stabil und bilden einen konstanten Referenzpunkt für die Erreichung des gewünschten Zustandes. Im Gegensatz dazu wird die *Akkommodation* als (eher passive) Neutralisation des gegenwärtigen Zustandes gewertet. Hier werden die persönlich bedeutsamen Bewertungshintergründe zugunsten neuer, zur gegenwärtigen Lebenssituation besser passender Bewertungsgrundlagen aufgegeben. Akkomodative Strategien werden als Mechanismen zur Erklärung paradoxer Befunde in der Zufriedenheitsforschung herangezogen. Dem Modell zufolge setzen sie ein, wenn assimilative Strategien nicht mehr erfolgreich angewendet werden können. Über Bewertungen der bisherigen Bemühungen zur Zielerreichung, der Beurteilung der gegenwärtigen Lebensbedingungen, Widersprüchen zwischen Kontrollüberzeugungen und Erfahrungen in der Auseinandersetzung mit Anforderungen und Belastungen sowie einem negativen emotionalen Befinden geprägt durch Hoffnungslosigkeit, Resignation und/oder Trauer, setzen zur Stabilisierung der individuellen Befindlichkeit akkomodative Prozesse ein. Diese werden inhaltlich beschrieben als Anpassung von Zielvorstellungen, Revision von Werten und Bewertungsstandards inklusive selbstbestärkender Vergleichsprozesse (v.a. ein sozialer Abwärtsvergleich) sowie eine Neutralisation aversiver Erlebniszustände durch selektive Fokussierung auf bestimmte Themen oder Ziele und sinnstiftende kognitive Bewertungsprozesse (Brandstädter & Grewe 1994). In diesem Sinne ist Response-Shift das Resultat einer akkomodativen Strategie im Rahmen der Auseinandersetzung mit kritischen Lebensereignissen. Eine Aufrechterhaltung unrealistischer Ziele und Werte, die vor dem Ereignis durchaus angemessen waren, würden zu dauerhafter Frustration führen.

Diese beiden unterschiedlichen Anpassungsprozesse werden von Heckhausen und Schulz (1995) auch als primäre und sekundäre Kontrollstrategien beschrieben. *Primäre Kontrollstrategien* beinhalten alle Möglichkeiten einer Person, auf ihre Lebenssituation in der Art Einfluss zu nehmen, dass sie ihre persönlichen Ziele, Bedürfnisse und Werte beibehalten und erreichen kann. *Sekundäre Kontrollstrategien* umfassen die internalen Veränderungen, die eine Person vornimmt, um eine positive Befindlichkeit zu erreichen. Heckhausen und Schulz (1995) unterscheiden sekundäre Strategien nochmals in *selektive* und *kompensatorische* sekundäre Kontrollprozesse. Erstere umfassen die Konzentration auf bereits vorhandene Ziele und Bedürfnisse, wobei andere, vorher relevante Ziele zu deren Gunsten vernachlässigt werden. Letztere Strategien beziehen sich auf eine Identifikation neuer Ziele, wenn die bisherigen nicht mehr erreichbar sind. Den Autoren zufolge ist für die Wahl der Kontrollstrategien in hohem Maße bedeutsam, welche Möglichkeiten der Person in ihrer Lebensumwelt zur Erreichung ihrer Ziele gegeben sind. Je mehr Möglichkeiten die Umwelt der Person bietet, desto wahrscheinlicher werden primäre Strategien aktiviert. Mit zunehmendem Alter – und damit einhergehenden Verlust- und Misserfolgserlebnissen – werden sekundäre Strategien bedeutsamer. Insbesondere die sekundären Kontrollstrategien, die darauf ausgerichtet sind, Erwartungen und Ziele zu reduzieren, ist für den Response-Shift-Kontext von Bedeutung, da damit ein Loslassen von vorher bedeutsamen Zielen im Sinne einer Repriorisierung verbunden ist (Wrosch et al. 2000).

Carver und Scheier (2000) beschreiben in ihrer *Theorie der adaptiven Selbstregulation* zwei verschiedene Feedbackschleifen, eine eher kognitiv-verhaltensorientierte und eine emotionale. Beide funktionieren als Kontrollprozesse bei der Verarbeitung von kognitiven und emotionalen Erlebnissen und besitzen für die individuelle Entwicklung eine zentrale Funktion. Entsprechend eines Regelkreismodells bestehen beide Feedbackschleifen aus vier Bestandteilen: einem Input, einem Referenzwert, einer Messinstanz und einem Output. Weiterhin nimmt der Output wiederum Einfluss auf die persönliche Lebenswelt der Person, die im weiteren Verlauf einen neuen Input bewirken kann. Ein Regelkreis für die kognitive Selbstregulation wäre, dass eine Person mit der Diagnose Krebs konfrontiert wird (Input) und diese Erfahrung persönliche Bedeutung bezüglich der individuellen Werte, Ziele und Erwartungen (Referenzpunkte) erhält. Möglicherweise sieht die Person Ziele gefährdet, die sie bisher definiert hatte. Jegliche Form der (Un-)Stimmigkeit zwischen Referenzpunkt und aktuellem Zustand löst entsprechende (Anpassungs-)Prozesse (Output) aus. Diese können dazu führen, sich mehr für das Erreichen des Referenzpunktes zu engagieren (Assimilation) oder den Referenz-

punkt zu verändern (Akkomodation). Kognitive Prozesse dienen als Messinstanz zur Erfassung der Diskrepanz zwischen Referenzpunkten und gegenwärtigem Zustand (Carver & Scheier 1981). Da der beschriebene Regelkreis für die Entwicklung und Bewältigung von Emotionen ebenso Gültigkeit beansprucht, resultiert aus der ständigen Misserfolgswahrnehmung und den damit verbundenen ungenügenden Anpassungsleistungen ein Insuffizienzgefühl von leichter negativer Affektivität bis hin zu Depression (Carver & Scheier 1990). Der emotionale Regelkreis wird durch den kognitiven ausgelöst und wirkt bei dauerhaftem negativem Output als Initiator für die Neudefinition von Referenzpunkten auf kognitiver Ebene. Die beschriebenen Prozesse erfolgen in jeglicher Anpassungssituation. In besonders schwierigen Lebenssituationen ist kurzfristig eher mit einer Steigerung der Anpassungsbemühungen zu rechnen. Besteht eine Belastung jedoch langfristig, ohne dass sich durch vermehrte Aktivität und Schaffung bestmöglicher Umweltbedingungen etwas verändert, bewirkt dies eine Adjustierung der Referenzwerte. Diesen zuletzt beschriebenen Prozess bezeichnen die Autoren als Response-Shift. Von besonderer Bedeutung für die diskrepanzreduzierenden Feedbackschleifen (Carver & Scheier 1999) sind im Falle einer Akkomodation die Veränderungen der Ziele und Bedürfnisse (Wrosch et al. 2003). Diese können entweder erfolgen, indem ein Ziel völlig an Bedeutung verliert (*goal disengagement*) oder durch ein anderes ersetzt wird (*goal reengagement*). Beide Prozesse sind voneinander unabhängig, d.h. es ist nicht zwangsläufig eine Aufgabe persönlicher Ziele notwendig, um Alternativen zu entwickeln, und umgekehrt. Sind in bestimmten Lebensphasen Ziele und Bedürfnisse für eine Person nicht mehr erreichbar beziehungsweise erfüllbar, verursacht dies eine negative affektive Befindlichkeit (Carver & Scheier 1990). In einer solchen Situation ist zur Veränderung der Referenzwerte eine Aufgabe bedeutsamer Ziele eine notwendige Voraussetzung, um den Distress dauerhaft zu vermeiden. Die Lösung von bedeutsamen Zielen führt in einer solchen Situation zum Ausgleich der emotionalen Befindlichkeit. Außerdem bieten sich durch die freiwerdenden Ressourcen Möglichkeiten zur Wahl von alternativen, angemesseneren Zielen. Kommt es zu einer tatsächlichen Aktivierung alternativer Ziele und Bedürfnisse, sind drei Aspekte zur Stabilisierung der Befindlichkeit wichtig: die Identifikation von Alternativen, die Beimessung von Bedeutung für diese Alternativen sowie die Ausrichtung des Verhaltens nach diesen. Die Verfolgung neuer Ziele, die für die betreffende Person sinnstiftenden Charakter besitzen, führt einerseits zur Verbesserung der psychischen Befindlichkeit (Wrosch et al. 2002, 2007a). Andererseits konnten Wrosch und Kollegen (2007b) nachweisen, dass auch die körperliche Gesundheit durch die Aufgabe unerreichbarer Ziele und das Streben nach erreichbaren Alternativzielen verbessert werden kann.

## Coping

Aus dem Bereich der Coping-Forschung wird ein Ansatz von Folkman und Greer (2000) vorgestellt, in dem das Konzept des *Meaning-based Coping* beschrieben wird. Ausgangspunkt der Modellentwicklung stellt das Coping-Modell nach Lazarus und Folkman (1990) dar. Diesem Modell zu Folge ist psychologischer Stress „...a relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being“ (Lazarus & Folkman 1990, S. 19). Im Sinne des Coping-Modells ist die individuelle Bewertung der Situation sowie der subjektiv wahrgenommenen Möglichkeiten, mit situationalen Gegebenheiten umzugehen, ausschlaggebend für die Bewältigung von Anforderungen und Belastungen. Ausgehend von der zentralen Funktion kognitiver Bewertungsprozesse für die Auseinandersetzung mit Anforderungen ist die subjektive Bewertung durch die Person entscheidend. Nach Richards und Folkman ist Response-Shift „... a result of specific meaning-based coping processes that come into play when previous beliefs, expectations, and goals are no longer tenable“ (Richards & Folkman 2000, S. 35). Folkman (1997) zufolge findet Response-Shift statt, wenn ein Ereignis von einer Person nicht bewältigt werden kann. In diesem Fall findet *Meaning-based Coping* statt, das als gleichbedeutend mit Response-Shift gesehen wird. Folkman und Greer (2000) stellen ein Modell vor, in dem der allgemeine Coping-Prozess um das so genannte Meaning-based Coping erweitert wird (Abbildung 3). Dieser Prozess tritt häufig auf im Rahmen der Auseinandersetzung von Menschen mit schweren chronischen bis lebensbedrohlichen Erkrankungen.

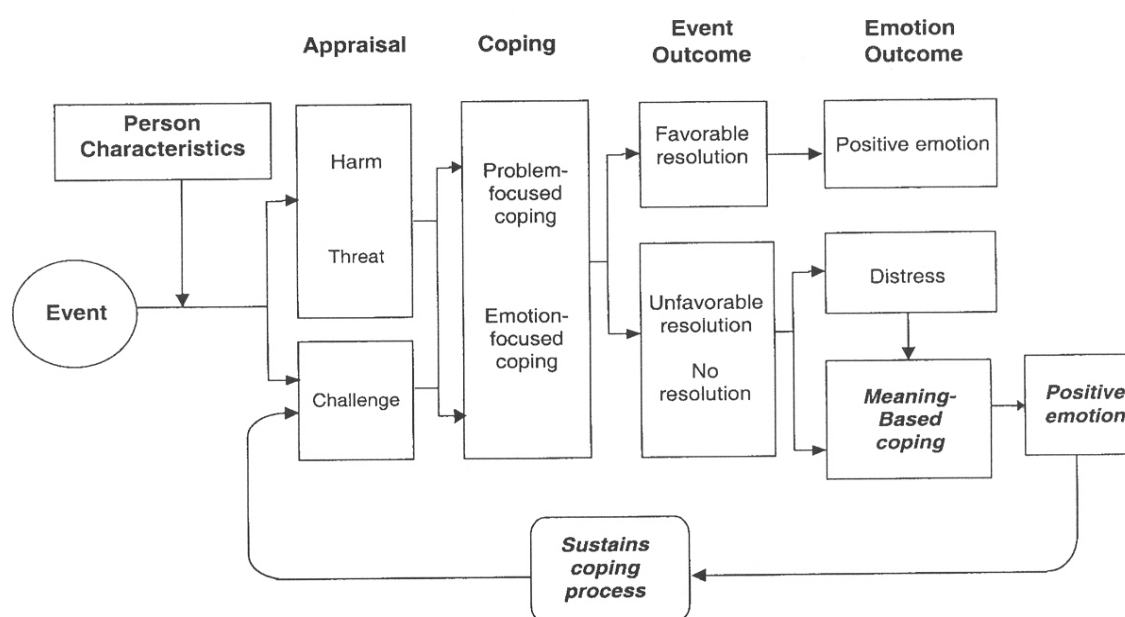


Abbildung 3: Theoretisches Modell der Bewertungs- und Copingprozesse nach Folkman und Greer (2000)

Grundlegend für das Modell sind zwei Prozesse: Appraisal (Bewertung) und Coping (Bewältigung). Mit Appraisal sind Bewertungen bezüglich der persönlichen Bedeutsamkeit eines Ereignisses (primary appraisal) und der vorhandenen Ressourcen, mit diesem Ereignis umgehen zu können (secondary appraisal), gemeint. Coping bezieht sich auf die Art der Bewältigung des Ereignisses, wobei drei verschiedene Varianten unterschieden werden: *emotionsorientiertes Coping* in Form von Gedanken und Handlungen zur Bewältigung des Distress-Erlebens, das durch ein Ereignis ausgelöst wird, *problemorientiertes Coping* als Behebung beziehungsweise Bewältigung der Ursache des Distress-Erlebens sowie *bedeutungsorientiertes Coping* (meaning-based coping) zur Aufrechterhaltung des psychischen Wohlbefindens (Folkman und Greer 2000).

Tritt ein Ereignis ein, das die zentralen Annahmen und Ziele einer Person in Frage stellt, kann dieses Ereignis als Schaden (harm), Bedrohung (threat) oder Herausforderung (challenge) bewertet werden. Ein Schaden oder Verlust steht für ein Ereignis, das bereits eingetreten ist, Bedrohung und Herausforderung für die Möglichkeit, auf ein Ereignis zu reagieren mit dem Ziel eines persönlichen Gewinns oder Wachstums. Wie ein Ereignis bewertet wird, hängt in hohem Maße ab von den persönlichen Wertvorstellungen, Überzeugungen und Grundannahmen einerseits und den vorhandenen persönlichen Ressourcen der Person für den Umgang mit diesem Ereignis andererseits. Insbesondere der persönlichen Kontrollierbarkeit des Ereignisses kommt große Bedeutung zu, denn nach Folkman und Greer (2000) beeinflusst das Ausmaß der Kontrolle die Art des Bewältigungsprozesses: Kann die Person das Ereignis kontrollieren, ist ein eher problemorientiertes Coping wahrscheinlich. Verfügt die Person nur über geringe Kontrolle über das Ereignis, ist ein emotionsorientiertes Coping erwartungsgemäß. Resultieren diese Bewältigungsbemühungen in angemessene Konsequenzen für die betroffene Person, ist psychisches Wohlbefinden erreicht. Ergibt sich aus den emotions- oder problemorientierten Ansätzen jedoch kein angemessenes psychisches Wohlbefinden, besteht weiterhin Distress. Dies führt zum bedeutungsbezogenen Coping. Die Person gibt nicht-haltbare Ziele oder Grundüberzeugungen auf und entwickelt neue, angemessenere Perspektiven, um sich so den bestehenden Anforderungen entsprechend anzupassen. Schafft es eine Person, ihre Grundüberzeugungen und Ziele zu relativieren, führt dies zu einem positiven psychischen Befinden, wobei dies wiederum aktuelle und zukünftige Copingbemühungen beeinflusst.

Meaning-based Coping lässt sich wiederum in zwei Bereiche unterteilen (Park & Folkman 1997, Richards & Folkman 2000, Folkman & Greer 2000): allgemeine Überzeugungen

(,global meaning') und situative Überzeugungen (,situational meaning'). Beide Bereiche sind keine disjunkten Kategorien, da sie sich wechselseitig beeinflussen. Park und Folkman (1997, S. 116) definieren allgemeine Überzeugungen als „... people's basic goals and fundamental assumptions, beliefs, and expectations about the world. Global meaning influences people's understanding of the past and the present, and it influences their expectations regarding the future“. Globale beziehungsweise allgemeine Überzeugungen bestehen zum größten Teil aus grundlegenden Gesetzmäßigkeiten, mit denen sich Personen Zusammenhänge in ihrem Leben erklären, und aus Zielen, die sich diese Personen in ihrem Leben setzen. Situative Überzeugungen werden von den Autorinnen definiert als „... interaction of a person's global belief and goals and the circumstances of a particular person-environment transaction“ (ebd. 1997, S. 121). Situative Überzeugungen bestehen aus drei verschiedenen Komponenten; zum einen der Bewertung von Bedeutungen, die konkrete Situationen für eine Person haben, zum zweiten aus Copingprozessen, in denen die Betroffenen versuchen, Bedeutungen in stressvollen Lebenssituationen zu relativieren, und zum dritten aus einer veränderten Bedeutung der erlebten Situation beziehungsweise des erfahrenen Ereignisses als Produkt des Bewältigungsprozesses. Vor diesem Hintergrund ist Response-Shift nach Richards und Folkman (2000, S. 35) das Resultat eines Prozesses, den sie beschreiben als „... realistic appraisal of changing circumstances, the consequent need to revise beliefs and expectations at both global and situational levels, and the adoption of new expectations and goals that are realistic and meaningful at both global and situational level“.

Der Ansatz nach Richards und Folkman (2000) ist eine Differenzierung der Coping-Prozesse, die an einem Response-Shift beteiligt sein können. Er stellt einen geeigneten Ausgangspunkt für weitere Analysen relevanter Bewertungsprozesse und Copingstrategien im Zusammenhang mit Response-Shift dar. Gemeinsamkeiten zum Response-Shift-Ansatz bestehen in der Veränderung zentraler Überzeugungen, Grundannahmen, Erwartungen und Ziele. Diese kognitiven Konstrukte stehen beim Copingansatz im Zentrum der inhaltlichen Analyse des Meaning-based Coping. Die Auseinandersetzung mit Anforderungen und Herausforderungen, die das individuelle System kognitiver Strukturen einer Person in Frage stellen, kann zu einer Veränderung des Wertesystems und somit zu psychischem Wohlbefinden führen. Auch dies entspricht dem Response-Shift-Modell von Sprangers und Schwartz (2000). Unterschiede zwischen dem Coping- und dem Response-Shift-Ansatz ergeben sich nicht aus der Bedeutung der Mechanismen, die zu einem Response-Shift beitragen. Vielmehr wird Response-Shift nach Sprangers und Schwartz (2000) als Phänomen auf einer abstrakteren Ebene



definiert. Der Unterschied ist vergleichbar mit der Definition eines latenten Konstrukts und dessen Operationalisierung: Veränderung von Zielvorstellungen, Erwartungen und Grundüberzeugungen sind nach Sprangers und Schwartz Konsequenzen eines Response-Shifts und daher geeignet, diesen zu operationalisieren. Das Phänomen an sich ist die Modifikation des individuellen Werte- und Bedeutungssystems, vergleichbar mit einem sich wandelnden Koordinatensystem, bei dem sich einerseits der Nullpunkt und andererseits die Längen der Achsen, die dieses System aufspannen, verschieben. Meaning-based Coping ist somit ein möglicher inhaltlicher Einflussfaktor auf das individuelle Koordinatensystem der Person, genau so wie andere Coping- und Bewertungsprozesse.

### *Benefit Finding*

In der Forschung zum Umgang mit kritischen Lebensereignissen und deren Folgen gewinnt seit einigen Jahren eine auf positive Veränderungen gerichtete Perspektive immer mehr an Bedeutung. Dabei werden unterschiedliche Konzepte und Konstrukte definiert, die auf die Möglichkeit einer positiven Weiterentwicklung einer Person nach belastenden Ereignissen, wie Verlustereignisse durch Tod oder Trennung, Bewältigung einer schweren Krankheit oder Arbeitslosigkeit, eingehen. Dazu gehören *Stress-related Growth* (Park et al. 1996), *Thriving* (O'Leary et al. 1998), *Positive Psychological Changes* (Yalom & Lieberman 1991) oder *Adversarial Growth* (Linley & Joseph 2004). Am häufigsten sind die Bezeichnungen *Posttraumatic Growth* (PTG, Calhoun & Tedeschi 1999, 2001, Maercker & Zoellner 2004, Zoellner & Maercker 2006) und *Benefit Finding* (Affleck & Tennen 1996).

Affleck und Tennen (1996) verstehen unter Benefit Finding einen kognitiven Prozess, der dazu führt, aufgrund widriger Umstände Möglichkeiten für positive persönliche Entwicklungen zu sehen und gegebenenfalls umzusetzen. Die Autoren unterscheiden zwischen *Benefit Finding* als adaptive Überzeugung, aus negativen Ereignissen Positives ableiten zu können, und *Benefit Reminding* als kognitive Copingstrategie. Diese wird zur Bewältigung eines Stressors verwendet, indem die Person sich positive Entwicklungen aus früheren belastenden Lebensereignissen in der Biographie selbst bewusst macht. Potenzielle Konsequenzen eines Benefit Finding sind beispielsweise die bewusstere Wahrnehmung des eigenen Lebens und der Umwelt, das Setzen neuer relevanter Ziele und Werte, eine Zunahme an persönlicher Stärke und Kraft, Wahrnehmung neuer Entwicklungsmöglichkeiten, mehr Nähe und Offenheit zu engen Bezugspersonen oder spirituelles Wachstum (Tedeschi et al. 1998).

Eine aktuelle Übersichtsarbeit in Form einer Metaanalyse zur Untersuchung von Benefit Finding im Zusammenhang mit verschiedenen psychosozialen Outcome-Parametern und darauf wirkende Moderatoren wurde von Helgeson und Kollegen veröffentlicht (2006). Die Autoren nahmen 87 Querschnittstudien in die Analyse auf. Es wurden die Beziehungen zwischen Benefit Finding und Depressivität, Ängstlichkeit, positive Befindlichkeit (positive Affektivität, Selbstwert, Lebenszufriedenheit), allgemeinem Stress (negative Affektivität, Stimmung), belastenden Intrusionen, allgemeiner Lebensqualität und subjektiv wahrgenommener physischer Funktionsfähigkeit untersucht. Als Moderatoren wurden demographische Merkmale (Geschlecht, Nationalität, Familienstand, Alter, Einkommen), Merkmale der Stressoren (Dauer, objektiver Schweregrad, wahrgenommener Stress), Persönlichkeitsmerkmale (Optimismus, Neurotizismus, Religiosität) und Copingstrategien (positive Neubewertung, Akzeptanz, Verdrängung) untersucht. Im Ergebnis zeigte sich, dass Benefit Finding negativ mit Depressivität ( $r = -.09$ ), positiv mit positiver Befindlichkeit ( $r = .22$ ) und dem Ausmaß an intrusiven Gedanken ( $r = .18$ ) korreliert. Mit Ängstlichkeit, allgemeinem Stress, allgemeiner Lebensqualität und subjektiver physischer Funktionsfähigkeit wurden querschnittlich keine statistisch signifikanten Zusammenhänge gefunden. Weiterhin wichtig für die vorliegende Studie ist, dass der Befund zur Beziehung zwischen Benefit Finding und allgemeiner Lebensqualität über alle Moderatorvariablen homogen ist. Bezüglich der überprüften Moderatorvariablen stellten die Autoren fest, dass die vergangene ‚Zeit seit Ereignis‘ statistisch bedeutsam ist. Benefit Finding korrelierte negativ mit Depressivität und positiv mit positiver Affektivität, wenn das Ereignis länger als zwei Jahren zurücklag. Lag das Ereignis weniger als zwei Jahre zurück, korrelierte Benefit Finding positiv mit allgemeinem Distress. Auch die Art des Ereignisses spielt als Moderator eine Rolle. Studien, die Benefit Finding im Zusammenhang mit (physischen) Erkrankungen untersuchten, zeigen eine negative Korrelation mit allgemeinem Distress. Studien, die andere traumatisierende Ereignisse untersuchten, zeigten keinen statistisch bedeutsamen Zusammenhang zwischen diesen Variablen. Weiterhin zeigten sich folgende Variablen als Korrelate von Benefit Finding: Frauen und Minderheiten (Nationalität) sowie jüngere Betroffene neigen eher zu Benefit Finding und dieses ist umso intensiver ausgeprägt, je gravierender das auslösende Ereignis objektiv beurteilt und subjektiv wahrgenommen wurde. Dieses Ergebnis steht zum Teil im Widerspruch zu Ergebnissen nach Stanton und Kollegen (2006), die für Krebspatienten nachweisen konnten, dass zwar die subjektiv erlebte Belastung, nicht aber der objektive Schweregrad der Krebserkrankung mit Benefit Finding zusammenhängt. Besonders hervorzuheben ist nach Helgeson und Kollegen (2006) der Befund, dass Benefit Finding einerseits negativ korreliert mit Depressivität und positiv mit

intrusiven Gedanken, was häufig zu inkonsistenten Bewertungen der publizierten Einzelstudien geführt hat (Carboon et al. 2005). Eine Erklärung für dieses Ergebnis wird darin gesehen, dass intrusive Gedanken als Beleg für die kognitive Verarbeitung des belastenden Ereignisses gesehen werden und nicht als Anzeichen mangelnder psychischer Befindlichkeit (Lepore 2001). Intrusive Gedanken sind nach dieser Auffassung Kennzeichen der internen Verarbeitung des Traumas mit gleichzeitiger positiver Ausrichtung dieser Verarbeitung im Sinne einer Umstrukturierung relevanter persönlicher Lebenskonzepte und -ziele.

Einen Überblick über die Auswirkungen von Benefit Finding bei Krebspatienten bietet Thornton (2002). Demnach zeigen sich in drei Bereichen positive intrapersonale Veränderungen während der Auseinandersetzung mit einer Krebserkrankung: Auswirkungen auf die Lebensperspektive und -philosophie, veränderte Bedeutung persönlicher und sozialer Beziehungen und selbstbezogene Veränderungen (beispielsweise im Bereich Spiritualität, selbstbezogene Aufmerksamkeit). Widows und Kollegen (2005) konnten im Längsschnitt für Patienten nach Knochenmarktransplantation nach einer Krebserkrankung negative Zusammenhänge feststellen zwischen Benefit Finding und dem Alter sowie Bildungsstand. Positive Korrelationen ergaben sich für Benefit Finding mit positiver Umdeutung sowie Problemlösestrategien. Schulz und Mohamed (2004) wiesen einen positiven Zusammenhang nach zwischen sozialer Unterstützung und Benefit Finding. Selbstwirksamkeit als personale Ressource hatte einen indirekten Einfluss auf Benefit Finding, vermittelt über soziale Abwärtsvergleiche.

Zum Zusammenhang von Benefit Finding und Lebensqualität stellten Roberts und Kollegen (2006) aufgrund regressionsanalytischer Untersuchungen zum Einfluss von sozialer Unterstützung, Benefit Finding und intrusiven Gedanken auf die psychische Befindlichkeit als Komponente der GLQ fest, dass Benefit Finding und Intrusionen das Ausmaß der psychischen Befindlichkeit statistisch bedeutsam beeinflussen. Nach Schwarzer und Kollegen (2006) steigt das Ausmaß an Benefit Finding nur dann mit der Zeit an, wenn das Ausgangsniveau dieser Strategie eher gering ist. Weiterhin zeigte sich, dass aktuell erfasstes Benefit Finding zu keinem von drei Messzeitpunkten mit Lebensqualität korreliert, Veränderungen im Benefit Finding hingegen schon.

Für PCa-Patienten konnten Kinsinger und Kollegen (2006) bedeutsame positive Zusammenhänge zwischen Benefit Finding und dem Ausmaß aktiver Copingstrategien sowie sozialer Unterstützung nachweisen. Subjektive GLQ war unkorreliert mit Benefit Finding. Thorn-

ton und Perez (2006) fanden im Längsschnitt ebenfalls keine signifikante Beziehung zwischen Benefit Finding und allgemeiner Lebensqualität. Als Prädiktoren für Benefit Finding ergaben sich vermehrte präoperative negative Affektivität, und postoperative positive Umdeutung sowie emotionale Unterstützung. Ein Jahr nach der Operation unterschieden sich die Lebenspartner der untersuchten Patienten nicht von den Krebspatienten im Ausmaß des Benefit Finding.

### *Integratives Modell der Anpassung an chronische Krankheiten*

Ein Modell zur Anpassung an chronische Krankheiten, das die oben genannten Konzepte in Beziehung zum Response-Shift setzt, haben Sharpe und Curran (2006) formuliert. Die Autoren integrieren Selbstregulations-, Coping-, Benefit Finding-Prozesse und Response-Shift in einem allgemeinen Anpassungsmodell. Dem Modell zufolge finden Anpassungsprozesse an eine sich ändernde Lebensbedingung (interne, externe) aufgrund einer chronischen Erkrankung erst dann statt, wenn durch die neue Lebenssituation Erfahrungen gemacht werden, die mit den bisherigen kognitiven Repräsentationen in Konflikt stehen. Das Ausmaß der Beteiligung der oben beschriebenen Prozesse der Selbstregulation, Bewältigung, Sinnfindung und des Response-Shifts ist jedoch abhängig vom Ausmaß der Veränderungen kognitiver Annahmen, Grundannahmen und Schemata, die wiederum mit der emotionalen Verarbeitung des Geschehens vernetzt sind.

Ausgangspunkt des Modells bilden allgemeine basale Grundannahmen und Schemata als kognitive Repräsentationen der eigenen Person, der Umwelt und der Erkrankung. Diese Repräsentationen stehen in engen Beziehungen zu verschiedenen weiteren Kernkonzepten wie persönlichen Zielen und Eigenschaften. Aus diesen persönlichkeitsdefinierenden Kernannahmen lassen sich bedingte Annahmen ableiten (intermediatory beliefs) im Sinne von ‚wenn, dann‘-Sätzen (beispielsweise „wenn ich von niemanden Hilfe benötige, dann bin ich ein unabhängiger Mensch“). Entwickelt sich eine chronische Erkrankung, ohne dass die bestehenden Grundannahmen und bedingten Annahmen der Person dadurch in Konflikt geraten, resultiert daraus eine positive Befindlichkeit. Kann eine Person die Erfahrung der chronischen Krankheit jedoch nicht ohne weiteres in ihre bestehenden Strukturen integrieren, starten Prozesse der Veränderung der auf die Krankheitssituation bezogenen Kognitionen. Dazu gehören Neubewertungsprozesse der Situation (Reappraisal, „ich habe gute Möglichkeiten, um mit der schwierigen Situation umzugehen“), des Abwärtsvergleichs („anderen geht es noch schlechter

als mir“) und des Benefit Finding („die Situation hat auch etwas Gutes“). Dem Modell zufolge ändern sich dadurch noch nicht die bedingten Annahmen sowie die Grundannahmen der betroffenen Person. Findet die Person jedoch kein neues Gleichgewicht zwischen Anforderungen und Haltungen, führt dies zu einer Änderung der bedingten Annahmen. Diese werden verändert und der Situation entsprechend angepasst (beispielsweise „wenn ich von niemandem Hilfe *im Alltag* brauche, bin ich ein unabhängiger Mensch“). Resultiert daraus ein neues Gleichgewicht, entsteht für die Person eine positive Befindlichkeit. Ist das nicht der Fall, setzen Response-Shift-Prozesse ein, die zur Konsequenz haben, dass die Grundannahmen und zentralen Schemata der Person verändert werden (beispielsweise „ich bin von anderen Menschen abhängig“), beziehungsweise zentrale Ziele (i.S. einer Repriorisierung nach der Sprangers’schen Response-Shift-Definition) werden adaptiert („Unabhängigkeit von Anderen ist weniger wichtig, eine gute Beziehung zu Bezugspersonen dafür wichtiger“). Wird auch durch diese Anpassungsleistungen kein für die Person stabilisierendes Gleichgewicht bezüglich ihrer zentralen kognitiven Strukturen und den Anforderungen aus der veränderten Lebenssituation erreicht, folgt daraus ein negatives Befinden. Andernfalls entsteht durch eine angemessene Anpassung der internen Strukturen eine positive Befindlichkeit.

Das Modell bietet einen allgemeinen Überblick über Beziehungen zwischen zentralen Anpassungsprozessen wie Selbstregulation, Coping, Benefit Finding und Response-Shift. Positiv anzumerken ist die ausgeprägte Orientierung an therapeutischen Ansatzmöglichkeiten, hier insbesondere der kognitiven Verhaltenstherapie, um im Falle einer maladaptiven Entwicklung Interventionsstrategien abzuleiten. Vernachlässigt werden jedoch verschiedene Prozessvariablen mit besonderer Bedeutung aus der tiefenpsychologischen beziehungsweise psychoanalytischen Therapie, wie beispielsweise Abwehrprozesse. Als Ausgangspunkt für eine Integration verschiedener Anpassungsprozesse hat das Modell jedoch einen heuristischen Wert, der darin zu sehen ist, dass auf die verschiedenen Komponenten, die Berücksichtigung finden sollten, hingewiesen und das Konzept des Response-Shift zu integrieren versucht wird. Unklar bleibt jedoch die Abfolge und Abgrenzung der einzelnen Konzepte. Kritisch ist zum Einen die unidirektionale Verknüpfung der einzelnen Komponenten. Ein Beleg für die Abfolge der beschriebenen Prozesse ist empirisch sicherlich schwer zu finden. Eine multidirektionale Beziehung zwischen den beteiligten Konzepten und der Befindlichkeit der Person als Ergebnis der Anpassung scheint naheliegender. Andeutungen dieser wechselseitigen Beziehungen sind möglicherweise in der Definition einzelner Anpassungsprozesse zu sehen. So wird beispielsweise der soziale Abwärtsvergleich als ein wichtiger Prozess im Rahmen eines

Response-Shifts gesehen, der dann in der Folge zu einer Rekalibrierung, Reprioritisierung oder Neukonzeptualisierung zentraler Konzepte und Strukturen führt. Dies gilt ebenso für den Prozess des Benefit Finding. Hier deuten sich konzeptuelle Unschärfen in der Verwendung des Response-Shift-Konzepts an. Somit ist das Modell von Sharpe und Curran lediglich als Ausgangspunkt für weitere integrative Ansätze zur Verknüpfung zentraler Selbstregulationsmechanismen zu sehen. In Abbildung 4 wird das Modell dargestellt.

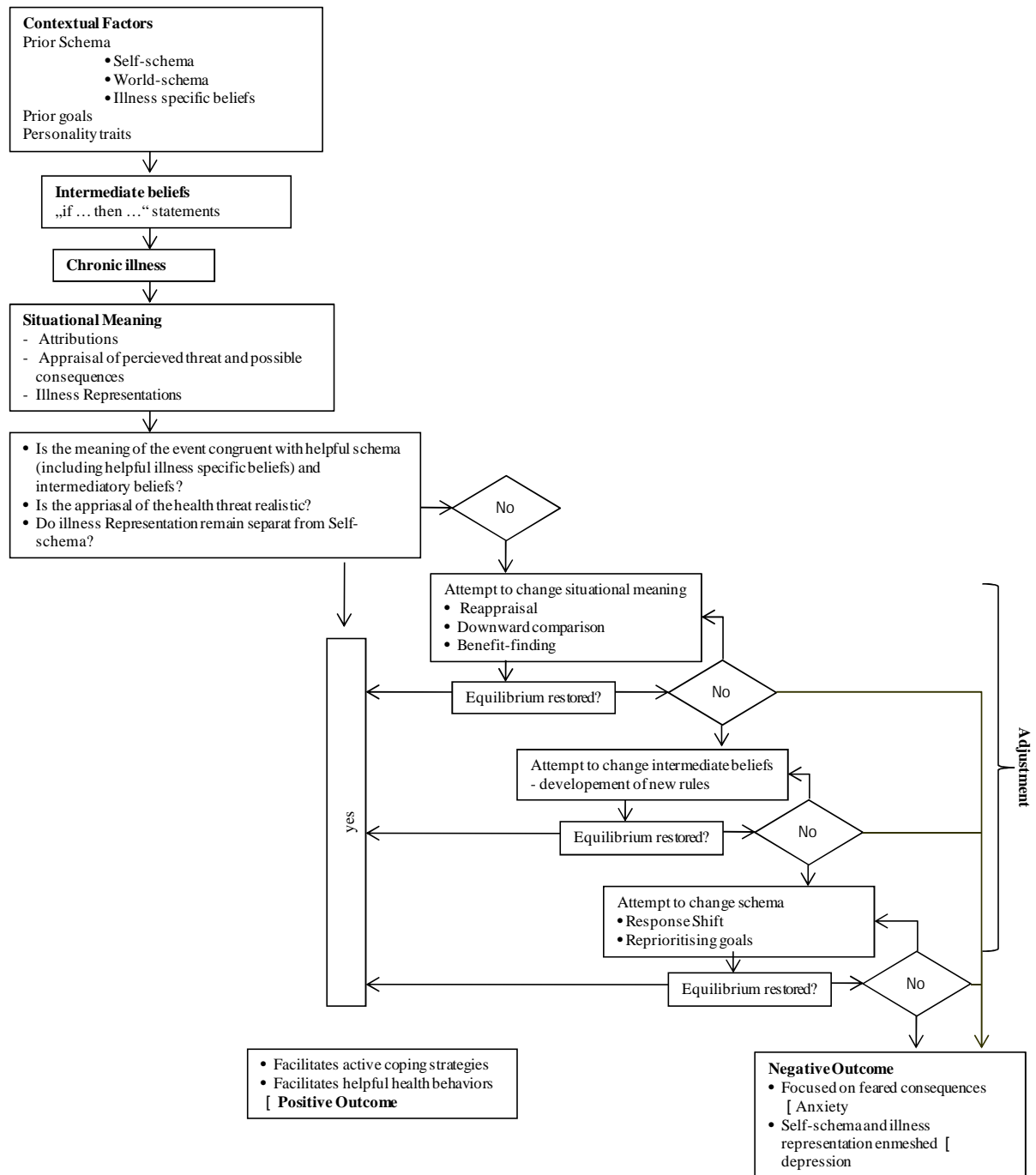


Abbildung 4: Modell der Anpassung an chronische Krankheiten nach Sharpe und Curran (2006, adaptiert nach Park und Folkman 1997)

### 3.2 Methoden zur Response-Shift-Erfassung

Die genaue methodische Erfassung von Response-Shift ist für die Bewertung und Evaluation von Behandlungsmaßnahmen von großer Bedeutung. Es ist möglich, anhand der Berücksichtigung von Response-Shift Behandlungseffekte genauer zu bestimmen, wodurch möglicherweise auch Aussagen über den Verlauf beziehungsweise die Veränderungen der Erkrankungen selbst spezifiziert werden können (Sprangers & Schwartz 2000). Es bestehen verschiedene Ansätze zur Erfassung eines Response-Shift, die sich grob in zwei Kategorien einordnen lassen, wobei insgesamt eine relativ große Schnittmenge der beiden Bereiche existiert, da diese häufig in Kombination miteinander verwendet werden. Die erste Gruppe der Verfahren stellt die direkte oder indirekte Erfassung von Response-Shift-Effekten auf inhaltlicher Ebene in den Vordergrund. Hierzu zählen verschiedene Ansätze, die persönlich bedeutsame Ziele, Werte oder Ankerpunkte für interne Skalierungen erfassen und miteinander in Beziehung setzen. Die zweite Gruppe der Verfahren und Strategien legen einen größeren Schwerpunkt auf statistische Auswertungen. Zu diesen Ansätzen zählen beispielsweise der Vergleich verschiedener Mittelwerte wie beim so genannten Thentest oder statistische Auswertungsstrategien zum Beispiel durch Faktorenanalysen oder Strukturgleichungsmodelle. Es wird darauf hingewiesen, dass die vorgestellten Ansätze nicht explizit für die Response-Shift-Erfassung im Bereich der GLQ, sondern der Lebensqualität allgemein entwickelt wurden.

Einen Überblick über verschiedene Strategien zur Response-Shift-Erfassung bieten Sprangers und Schwartz (2000) sowie Daig und Lehmann (2007). Sprangers und Schwartz (2000) nehmen eine Einteilung der Verfahren in folgende Bereiche vor: individuell-orientierte Ansätze, Design-Ansätze und Kovarianz-Ansätze. Diese drei methodischen Herangehensweisen werden anhand einiger Beispiele skizziert. Auf den Ansatz zur Erfassung des Response-Shift mittels Konfirmatorischer Faktorenanalyse (KFA) wird an entsprechender Stelle vertiefend eingegangen.

#### 3.2.1 Individuell-orientierte Ansätze

Sprangers und Schwartz (2000) unterteilen diesen Bereich der Strategien zur Erfassung von Response-Shift in vier Gruppen: einzelfallbezogene Methoden, präferenz-basierte Methoden, sukzessive Vergleichsmethoden und qualitative Methoden. Das Spektrum möglicher Ansätze ist kaum überschaubar, da eine Flut an Strategien existiert, um die Wertigkeit und

persönliche Bedeutung von Lebensqualitätsaspekten und deren Veränderung über die Zeit zu erfassen.

Der *einzelfallbezogene Ansatz* ist darauf ausgerichtet, die für die betroffene Person relevanten Dimensionen oder Komponenten der Lebensqualität zu erfassen. Dieses kann in Form einer Exploration oder eines Fragebogens geschehen. Die individuelle Zusammensetzung der Lebensqualität stellt den Ausgangspunkt der Response-Shift-Erfassung dar. Zu mindestens zwei Messzeitpunkten werden subjektiv relevante Komponenten der Lebensqualität erhoben. Letztlich werden über inhaltliche Veränderungen persönlich relevanter Lebensqualitätsdimensionen und Werte sowie Änderungen der Gewichtung einzelner Lebensqualitätsdimensionen und Werte Response-Shift-Einflüsse erfasst. Beispiele dieser Gruppe sind u.a. die Repetory-Grid-Technique (Thunedborg et al. 1993), die Subjective Domains of Quality of Life Measure (Bar-On & Amir 1993), die Schedule for Evaluation of Individual Quality of Life (SEIQoL, O'Boyle et al. 2000) und der Patient Generated Index (Ruta et al. 1994). Die beiden zuletzt genannten Verfahren sollen beispielhaft ausführlicher dargestellt werden. Das *Schedule for Evaluation of Individual Quality of Life* (SEIQoL, O'Boyle et al. 2000, Sharpe et al. 2005, Waldron et al. 1999) dient der Erfassung der subjektiven Lebensqualität und der Analyse persönlich bedeutsamer Lebensqualitätskomponenten und deren Wertigkeit für die Person. Mehrere Faktoren sollen mit dem SEIQoL erfasst werden. Es werden die fünf für die Person bedeutsamsten Bereiche der subjektiven Lebensqualität erfragt. Die Wahl dieser Bereiche steht dem Befragten frei. Es werden keine Vorgaben gemacht. Anschließend wird für jede der genannten Komponenten auf einer visuellen Analogskala eingeschätzt, inwieweit diese aktuell ausgeprägt ist (Ankerpunkte: bestmöglich vs. nicht vorhanden). Im nächsten Schritt wird die persönliche Bedeutung jeder einzelnen Komponente gewichtet. Aus den Ausprägungen der aktuellen Zustandsbeschreibungen der fünf bedeutsamsten Lebensbereiche in Kombination mit deren subjektiver Gewichtung wird ein SEIQoL-Score für die allgemeine Lebensqualität abgeleitet. Response-Shift kann mit dem SEIQoL erfasst werden, indem durch Mehrfachmessungen überprüft wird, ob sich die genannten Komponenten zu zwei Messzeitpunkten inhaltlich unterscheiden (Rekonzeptualisierung) und / oder ob bei Nennung gleicher Komponenten Veränderungen in deren persönlicher Gewichtung auftreten (Repriorisierung). Eine Rekalibrierung wird mit der aktuell konzipierten Form des SEIQoL nicht erfasst. Der *Patient Generated Index* nach Ruta und Kollegen (1994, 1999) folgt dem gleichen Prinzip. Der Fokus des Ansatzes liegt jedoch auf Einschränkungen der Lebensqualität. Das Verfahren beginnt mit der freien Nennung von maximal fünf Lebensbereichen oder Aktivitäten, die mit GLQ in Zu-



sammenhang stehen, und gegenwärtig durch die vorliegende Erkrankung am deutlichsten eingeschränkt sind. In einem zweiten Befragungsschritt soll die Person für die genannten Aspekte einschätzen, wie sie ihren aktuellen Stand im Vergleich zu dem, was sie gegenwärtig unter idealen Bedingungen erreichen könnte, in den einzelnen Bereichen bewertet. In einem dritten Schritt werden Einschätzungen zur persönlichen Wichtigkeit der genannten Aspekte verlangt. Dieselbe Befragung wird zu einem späteren Zeitpunkt wiederholt. Änderungen in der Anzahl oder der Themeninhalte wird als Neukonzeptualisierung definiert. Bei Konstanz der Inhalte kann eine Repriorisierung durch die veränderte Gewichtung der einzelnen Bereiche erfasst werden.

Mit der *präferenzbasierten Methode* soll die Bedeutung von Gesundheit beziehungsweise gesundheitlichen Beschwerden oder Behandlungsnebenwirkungen mit Lebensqualität in Beziehung gesetzt werden. Die so genannten *Time-Trade-off-Ansätze* (Torrance 1986) bilden eine Möglichkeit zur Erfassung der Bedeutung von gesundheitsbezogener Lebensqualität im Kontext einer Erkrankung und deren Verlauf. Hauptanliegen dieser Methode ist es, eine quantitative Aussage über das Verhältnis zwischen möglichst langer Lebenszeit und einem möglichst guten Gesundheitszustand ableiten zu können. Durch Mehrfachmessungen können Veränderungen in der Bedeutung persönlicher Zielkriterien erfasst werden. Beispielhaft soll hier die Q-TWIST-Methode vorgestellt werden (Schwartz et al. 1995). Die *quality-adjusted time without symptoms of disease and toxicities* (Q-TWIST) wurde entwickelt, um bei Personen, die an schweren Krankheiten leiden (beispielsweise Krebs, AIDS, Epilepsie), die Kosten der Behandlung in Form von Nebenwirkungen und körperlichen und psychischen Belastungen dem Nutzen in Form von verbesserter Gesundheit gegenüberzustellen. Ziel der Q-TWIST-Methode ist es, Maßnahmen zur Behandlung von - beispielsweise - Krebserkrankungen anhand ihrer Bedeutung für die individuell erlebte Lebensqualität zu bewerten. Die Überlebenszeit des Patienten wird in Beziehung gesetzt zu der von ihm erwarteten Lebensqualität in einzelnen Bereichen (Cole et al. 2004). Aus der Q-TWIST-Methode resultiert ein Wert, der Aussagen ermöglicht über die Akzeptanz von belastenden Krankheitssymptomen und Therapienebenwirkungen durch den Betroffenen in Relation zu einer erwarteten (belastungsfreien) Überlebenszeit. Diese belastungsfreie Überlebenszeit wird als Operationalisierung der Lebensqualität verstanden („the best quality of life that a patient can expect given that s/he has a disease“ (Schwartz et al. 1995, S. 136)). Grundgedanke ist, je besser ein Patient die Möglichkeit einer belastungsfreien Überlebenszeit einschätzt, desto eher ist er bereit, aktuelle Krankheitssymptome und Therapienebenwirkungen auf sich zu nehmen. Einen Über-

blick über die Q-TWIST-Methode zur Beurteilung der GLQ bei Krebspatienten bieten Revi-cki und Kollegen (2006). Zur Response-Shift-Erfassung haben die Autoren die Q-TWIST-Methode erweitert, indem sie einzelne Komponenten der Lebensqualität in Beziehung zu Beschwerden setzen und diese Lebensqualitätskomponenten durch die Betroffenen gewichten lassen. Unterschiedliche Gewichtungen dieser Komponenten zu verschiedenen Messzeitpunkten weisen auf eine Repriorisierung hin. Gleichzeitig berücksichtigt diese Methode auch die Veränderung von Toleranz gegenüber den genannten Beschwerden und Belastungen.

Die *Sukzessive Vergleichsmethode* basiert auf Entscheidungsprozessen, mit deren Hilfe Rangreihen von Lebensqualitätsdimensionen festgelegt und deren Veränderung im längsschnittlichen Verlauf überprüft werden. Schwartz und Sprangers (2000) stellen zwei Methoden vor, die zur Response-Shift-Erfassung eingesetzt werden können: den paarweisen Vergleich (Edwards 1957) sowie den *Card Sort Approach* (Schwartz 1996). Bei beiden Ansätzen ist es möglich, entweder die zu vergleichenden Lebensqualitätsdimensionen (theoretisch begründet) vorher festzulegen oder aber die Person selbst entscheiden zu lassen, welche Lebensqualitätsdimension in den Vergleich aufgenommen werden. Die Wahl für die eine oder andere Variante ist für die Erfassung des Response-Shift von Bedeutung, da durch die dem Probanden überlassene freie Auswahl der von ihm als relevant angesehenen Lebensqualitätsdimensionen eine Rekonzeptualisierung mit größerer Wahrscheinlichkeit erfasst wird. Der Proband bildet eine Rangreihe aus vorgegebenen oder frei gewählten Lebensqualitätskomponenten oder Werten entsprechend einer mehrstufigen Skala. Beispielsweise sollen Personen auf einer 7-stufigen Skala von ‚sehr wichtig‘ bis ‚gar nicht wichtig‘ für jede angegebene Skalenabstufung eine Komponente der Lebensqualität angeben, die dieser Abstufung entspricht. Ändern sich bei Mehrfachmessungen die Rangfolgen in der Weise, dass manche Lebensqualitätskomponenten nicht mehr in die Rangfolge aufgenommen werden oder andere neu hinzukommen, so ist davon auszugehen, dass eine Rekonzeptualisierung des Konstrukts Lebensqualität stattgefunden hat. Ändern sich die Lebensqualitätskomponenten lediglich in der Abfolge der Rangordnung, so erfolgte eine Repriorisierung. Durch diesen Ansatz wird die Rekalibrierung nicht erfasst. Bei dem *paarweisen Vergleich* werden jeweils zwei Komponenten der Lebensqualität miteinander verglichen und der Proband muss eine Entscheidung darüber treffen, welche der beiden Komponenten für seine Lebensqualität bedeutsamer ist. Auf diese Weise können alle Komponenten miteinander verglichen und eine Rangreihe über die Anzahl der Komponenten gebildet werden.

Ein *qualitativer Ansatz* zur Erfassung des Response-Shifts wurde von Rapkin und Fisher (1992) und Rapkin und Schwartz (2004) entwickelt, wobei diese im Laufe der Zeit eher zu einer Kombination aus qualitativem und quantitativem Vorgehen erweitert wurde. Im ersten Teil der Untersuchung wird der Befragte gebeten, Anhaltspunkte zu seinem Lebensqualitätskonzept zu formulieren. Dies erfolgt beispielsweise durch eine Formulierung dessen, was Lebensqualität für die befragte Person aktuell bedeutet. Anschließend werden Fragen nach den wichtigsten Zielen, Problemen, Anliegen etc. der Person gestellt. Im weiteren Verlauf soll die Person angeben, welcher der genannten Inhalte für sie während des Interviews für die Beantwortung der Fragen am meisten bestimmend war. Anschließend werden durch standardisierte Itemformulierungen Einschätzungen zu Response-Shift-Prozessen und relevanten Mechanismen nach dem Modell von Sprangers und Schwartz (2000) erhoben. Insbesondere werden soziale Vergleichsprozesse, positive und negative Affektivität erfasst. Abschließend wird die Person um eine Einschätzung ihres gegenwärtigen Gesundheitszustandes auf einer 11-stufigen Skala gebeten (0 = sehr schlecht bis 10 = sehr gut). Response-Shift wird nach diesem Ansatz erfasst, indem zu einem späteren Messzeitpunkt der Person die Nennungen ihrer offenen Fragen vorgelesen werden und diese zu entscheiden hat, ob dieses Ziel, dieses Problem etc. noch immer diese Bedeutung für sie hat oder ob es mittlerweile andere Inhalte gibt, die wichtiger geworden sind. Auf diese Weise sollen eine Neukonzeptualisierung oder eine Repriorisierung erfasst werden. Durch eine retrospektive Einschätzung der Gesundheit zum Zeitpunkt der letzten (bzw. ersten) Befragung auf einer 11-stufigen Skala soll das Ausmaß der Rekalibrierung im Sinne eines Thentest (siehe unten) erfasst werden.

### 3.2.2 Design-Ansätze

Der am häufigsten eingesetzte Ansatz zur Erfassung von Response-Shift ist der so genannte *Thentest* beziehungsweise *retrospektive Prätest*, der in verschiedenen Bereichen der Evaluation von Interventionen angewandt wurde (beispielsweise Howard et al. 1981, Terborg & Davis 1982). Bereits Howard und Dailey (1979) verweisen auf die Möglichkeit der Erfassung einer Veränderung der internen Skalierung durch dieses Verfahren. Dabei wird dem Probanden ein Instrument, das zur Prä-Post-Messung eingesetzt wird, zum Postmesszeitpunkt zweimal vorgelegt, einmal, um eine aktuelle Bewertung, und zum zweiten Mal, um eine rückblickende Einschätzung des Konstrukts zum Zeitpunkt der Prämessung zu erfassen. Es wird angenommen, dass die retrospektive Bewertung (für den ersten Messzeitpunkt) und die zeitgleich erfasste aktuelle Bewertung des Konstrukts vor dem gleichen individuellen Bewertungshintergrund vorgenommen werden. Die Differenz zwischen den so erhaltenen Messun-

gen gilt – im Gegensatz zu den herkömmlichen Prä-Post-Messungen – als response-shift-be-reinigt. Ein Response-Shift in Form einer Rekalibrierung drückt sich in der Differenz des zum ersten Messzeitpunkt tatsächlich erhobenen Prä-Wertes und des retrospektiv erhobenen Prä-Wertes (Thentest) zum Zeitpunkt des zweiten Messzeitpunkts aus. Je größer diese Differenz, desto größer ist das Ausmaß der Veränderung interner Standards. Eine Repriorisierung und eine Neukonzeptualisierung können mit diesem Ansatz nicht erfasst werden. Deshalb wird empfohlen, den Thentest mit anderen Verfahren zur Response-Shift-Erfassung zu kombinieren (siehe qualitativer Ansatz nach Rapkin & Schwartz 2004). Empfehlungen für Verwendung des Thentest zur Erfassung von Response-Shift bieten Schwartz und Sprangers (2010).

In manchen Studien mit retrospektivem Prätest wird die Veränderung interner Standards in zwei Formen unterteilt: einen positiven und einen negativen Response-Shift (Jansen et al. 2001, Osborne et al. 2006, Sprangers & Schwartz 2000). Unterschieden wird bei dieser Einteilung die Richtung der Veränderung. Je nach Polung der Lebensqualitätsskala richtet sich die Bezeichnung des Response-Shifts. Mit einer positiven Ausprägung eines Konstrukts sind meist höhere Werte auf der entsprechenden Skala verbunden. In diesem Fall würde ein positiver Response-Shift resultieren, wenn die Differenz zwischen dem Prä-Wert zum ersten Messzeitpunkt und dem Thentest-Wert positiv ist ( $MZP 1 - THEN-TEST > 0$ ). Im retrospektiven Prätest wird der Zustand zum ersten Messzeitpunkt schlechter eingeschätzt, als dies in der prospektiven Prämessung tatsächlich der Fall war. Ein negativer Response-Shift resultiert demgegenüber aus einer ‚Überschätzung‘ des tatsächlichen Wertes zum ersten Messzeitpunkt durch den retrospektiven Prä-Wert ( $MZP 1 - THEN-TEST < 0$ ). Diese Unterscheidung ist für die herkömmliche Auswertung von Prä-Post-Vergleichen von besonderer Bedeutung. Im Falle eines positiven Response-Shift und unter den Annahmen (a) der Thentest bilde den tatsächlichen Prä-Wert valider ab sowie (b) die GLQ hat sich verbessert, kommt es zu einer Unterschätzung der Differenz durch den klassischen Prä-Post-Vergleich. Liegt ein negativer Response-Shift vor, resultiert unter denselben Annahmen (a) und (b) wie im vorangegangenen Satz aus dem klassischen Prä-Post-Vergleich eine Überschätzung der Differenz. Dies ist für die Bewertung von Interventionen im Rahmen von Prä-Post-Vergleichen (mit oder ohne Kontrollgruppe) für die Einschätzung des Effektes einer Intervention wichtig.

Eine weitere Strategie zur Erfassung eines Response-Shift im Sinne einer Rekalibrierung stellt der Ideal-Skalen-Ansatz dar (Armenakis 1988). Eine Person beantwortet zu zwei Messzeitpunkten (Prä-Post) den Fragebogen zur Messung des latenten Konstrukts jeweils zweimal.

Einmal soll die Person eine Bewertung zum gegenwärtigen Zustand geben. Die zweite Vorlage des Fragebogens soll so beantwortet werden, wie es dem Betroffenen in seiner aktuellen Situation idealerweise gehen könnte. Änderungen in der internen Skalierung des Zielkonstrukts würden durch einen Vergleich der „Idealskalen“ zu den beiden Messzeitpunkten erfassbar. Dieser Ansatz dient der Erfassung der Veränderung dessen, was als positiver Ankerpunkt der internen Skala des jeweiligen Konstrukts anzusehen ist. Ändern sich diese Ankerpunkte, weist dies auf eine Rekalibrierung hin. Eine genauere Abschätzung der tatsächlichen Veränderung resultiert aus einer Relativierung der Differenzen aus den beiden Einschätzungen zur gegenwärtigen Situation an der Veränderung der beiden Idealskalen. Repriorisierung und Neukonzeptualisierung werden durch diesen Ansatz nicht erfasst.

### 3.2.3 Statistische Ansätze

Im Mittelpunkt stehen Ansätze, die über einen Vergleich der Kovarianz und Mittelwerte von Messvariablen zu unterschiedlichen Messzeitpunkten Veränderungen in der zugrunde liegenden Varianz/Kovarianzmatrix erfassen. Für die vorliegende Arbeit von Bedeutung sind faktorenanalytische Ansätze, deren Besonderheit unter anderem darin liegen, Response-Shift nicht auf individueller, sondern auf Gruppenebene zu erfassen. Gegenwärtig wird die Konfirmatorische Faktorenanalyse (KFA) als das statistische Verfahren angesehen, das am ehesten geeignet ist, Response-Shift ökonomisch und differenziert zu erfassen (Oort 2004). Die Voraussetzung für die Verwendung der KFA zur Response-Shift-Erfassung ist, dass Response-Shift bei einem substantiellen Anteil der Stichprobe auftritt (Oort et al. 2005). Weitere Verfahren sind sog. Residualanalysen (Mayo et al. 2008), Wachstumsanalysen (Brossart 2002, Schwartz & Sprangers 2000) sowie die Verwendung von Hierarchischen Linearen Modellen (HLM, Lowy & Bernhard 2004). Auf diese Verfahren wird ebenfalls eingegangen.

#### *Explorative und Konfirmatorische Faktorenanalyse*

Die Response-Shift-Erfassung mittels faktorenanalytischer Verfahren wird im Folgenden ausführlich dargestellt, wobei insbesondere Bezug auf die Arbeiten von Oort (2005a, 2005b) sowie Visser und Kollegen (2005) genommen wird. Vorab werden jedoch methodische Grundlagen vorgestellt, um den Ansatz in einem breiteren Kontext darzustellen. Dabei handelt es sich um zentrale Annahmen der Klassischen Testtheorie als Basis der Strukturgleichungsmodelle sowie um die Invarianz beziehungsweise Äquivalenz von Messungen.

Messtheoretisch beruht der Ansatz der Response-Shift-Erfassung nach Oort (2005a) einerseits auf den Grundlagen der *Klassischen Testtheorie* (KTT, Lord & Novick 1968). Bei der KTT handelt es sich um eine Kombination verschiedener Axiome, die Aussagen darüber zulassen, wie gut manifeste, tatsächlich messbare Variablen geeignet sind, um latente, nicht direkt messbare Variablen (wahre Werte) zu schätzen (DeVellis 2006). Ohne auf die verschiedenen Axiome dieser Testtheorie genauer einzugehen, soll hier das erste Axiom vorgestellt werden: Der tatsächliche Testwert  $X$  einer Person setzt sich zusammen aus einem Anteil eines wahren Wertes  $T$  und einem Anteil eines Messfehlers  $E$  (Gleichung 1).

$$X_{ijk} = T_{ij} + E_{ijk} \quad (1)$$

$X$  ist der Testwert der  $k$ -ten Beobachtung der  $i$ -ten Untersuchung des  $j$ -ten Phänomens, das untersucht werden soll. Dementsprechend ist  $T$  der *wahre Wert* des  $j$ -ten Phänomens beziehungsweise Konstrukts in der  $i$ -ten Untersuchung und  $E$  der *Messfehler* der  $k$ -ten Beobachtung der  $i$ -ten Untersuchung des  $j$ -ten Konstrukts. Der Messfehler beinhaltet der Theorie nach alle unsystematischen Einflüsse, die den tatsächlichen Testwert beeinflussen, ohne Bestandteil des wahren Wertes zu sein. Auf die Bedeutung des Messfehlers wird weiter unten noch ausführlicher eingegangen.

Gleichung (2) resultiert aus den beschriebenen Zusammenhängen und stellt diese auf der Ebene von Varianzen dar.

$$\sigma^2_X = \sigma^2_T + \sigma^2_E \quad (2)$$

Die Varianz der Messung eines Konstrukts setzt sich zusammen aus einem Varianzanteil bedingt durch den wahren Wert sowie einem Anteil an Messfehlervarianz. Strukturgleichungsmodelle (SEM) im Allgemeinen und Konfirmatorische Faktorenanalysen (KFA) im Speziellen bieten die Möglichkeit, den wahren Wert eines latenten Konstrukts sowie den Einfluss eines Messfehlers zu schätzen. Dabei werden die Kovarianzen oder Korrelationen zwischen verschiedenen manifesten Variablen, also Variablen, die konkret erfassbar sind, gemessen, die inhaltlich einem latenten Konstrukt, also Variablen, die nicht konkret erfassbar sind, zugeordnet sind. Aus der gemeinsamen Kovarianz der gemessenen manifesten

Variablen lässt sich das latente Konstrukt schätzen. Faktorenanalytisch stellt dies einen Faktor dar. Der Anteil der einzelnen manifesten Variablen, der nicht auf den Einfluss des latenten Konstrukts zurückzuführen ist, wird auf entsprechende Messfehler zurückgeführt.

Gleichung (3) stellt die statistische Formulierung eines Messmodells dar, wie es in Strukturgleichungsmodellen beziehungsweise Konfirmatorischen Faktorenanalysen abgebildet wird. Zusätzlich zu den Variablen ‚wahrer Wert‘ und Messfehler kommt auf der Ebene der Kovarianzberechnungen eine weitere Variable hinzu, der Achsenschnittpunkt  $\tau$ .

$$x = \tau + \xi + \delta \quad (3)$$

Der tatsächliche Messwert einer Person setzt sich zusammen aus dem wahren Wert  $\xi$ , einem Messfehler  $\delta$  sowie einem Wert  $\tau$ , der in einer Regressionsgleichung den Wert für  $x$  annimmt, wenn der wahre Wert  $\xi$  gleich Null ist. In der graphischen Darstellung entspricht  $\tau$  dem Achsenschnitt der Geraden mit der Abszisse. Für die Überprüfung von Messmodellen mittels Konfirmatorischer Faktorenanalyse werden jedoch mehrere manifeste Variablen benötigt, deren Gemeinsamkeit ein zu untersuchendes latentes Konstrukt ist. An dieser Stelle soll nicht weiter auf Voraussetzungen für die Berechnung von Strukturgleichungsmodellen und KFA eingegangen werden. Üblicherweise gilt der Grundsatz, dass wenigstens drei manifeste Variablen verwendet werden sollen, um eine latente Variable zu schätzen (siehe dazu Kline 2005, sowie auch zu den Voraussetzungen für die Verletzung dieser Berechnungsvoraussetzung bei SEM bzw. KFA). Da in der vorliegenden Arbeit ein Messmodell mit fünf manifesten Variablen und einer latenten Variable untersucht wird, dient diese Variablenkombination der Darstellung der statistischen Parameter im Gleichungssystem (4), das in matrixschreibweise die Zusammenhänge zwischen den beteiligten Variablen darstellt.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \end{pmatrix} + \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} \xi + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{pmatrix} \quad (4)$$

Die Werte der fünf manifesten Variablen  $x_{1-5}$  werden jeweils durch einen eigenen Achsenschnittpunkt  $\tau_{1-5}$  sowie durch einen eigenen Messfehleranteil  $\delta_{1-5}$  bestimmt. Zusätzlich dazu kommt noch ein gewichteter Anteil der latenten Variable  $\xi$ , wobei die Parameter  $\lambda_{1-5}$  die Ladungen der entsprechenden manifesten Variablen auf dem Faktor angeben.

Anknüpfend an diese kurze Darstellung der theoretischen und statistischen Grundlagen der KFA werden verschiedene Bereiche der Invarianz- beziehungsweise Äquivalenzmessungen vorgestellt. Hinsichtlich der Überprüfung der Vergleichbarkeit von Messungen im Längsschnitt wird angenommen, dass sich die Parameter  $\tau_{1-5}$ ,  $\lambda_{1-5}$  und  $\delta_{1-5}$  zu unterschiedlichen Messzeitpunkten nicht voneinander unterscheiden, so dass Änderungen in den gemessenen Werten  $x_1$  bis  $x_5$  lediglich durch Änderungen der latenten Variablen  $\xi$  zustande kommen. Auf diese Annahmen und entsprechende Einschränkungen wird im nächsten Abschnitt ausführlich eingegangen.

### *Invarianz von Messungen*

Mittelwerts- und Kovarianzanalysen (Mean and Covariance Structures (MACS) Analysis) bieten einen geeigneten Ansatz, um die Äquivalenz von Messungen zwischen Gruppen, aber auch innerhalb einer Gruppe zu verschiedenen Messzeitpunkten zu überprüfen (French & Finch 2006, Little 1997, Meredith & Teresi 2006, Ployhart & Oswald 2004, Steenkamp & Baumgartner 1998, Vandenberg & Lance 2000). Als statistisches Verfahren zur Überprüfung dienen Strukturgleichungsmodelle beziehungsweise Konfirmatorische Faktorenanalysen (Brown 2006, Gregorich 2006, MaCallum & Austin 2000, Raju et al. 2002, Reise et al. 1993, Reise & Widaman 1999, Thompson 2004). Aus methodisch-statistischer Sicht entwickelte sich der Ansatz der Response-Shift-Erfassung mittels Strukturgleichungsmodellen aus der Äquivalenz beziehungsweise Invarianz von Messungen (Oort 2005a). Vandenberg und Lance (2000) bieten einen Überblick über verschiedene Invarianztestungen im Zusammenhang mit der Vergleichbarkeit von Gruppen und deren statistischen Kennwerten im Quer- und Längsschnitt. Der Ansatz der Messung auf Invarianz zwischen Gruppen ist in der Organisationspsychologie ein in den vergangenen Jahren zunehmend eingesetztes Verfahren, das in erster Linie der Bestätigung der Vergleichbarkeit der Messungen in verschiedenen Gruppen dient. Hinsichtlich der testtheoretischen Gütekriterien zielt die Invarianztestung auf eine Untersuchung der Konstruktvalidität des Instruments ab, mit dem gemessen wird (Bagozzi & Edwards 1998).



Aus Gleichung (4) sind verschiedene Hypothesen ableitbar, die bei der Testung auf Äquivalenz von Messungen überprüfbar sind. Die für die vorliegende Arbeit zentralen Hypothesen der Äquivalenzmessung werden in der Tabelle 10 dargestellt (in Anlehnung an Vandenberg & Lance 2000). In der vorliegenden Arbeit soll die Invarianz von Messmodellparametern im Längsschnitt überprüft werden. Deshalb werden in der Tabelle Aussagen über längsschnittliche Messungen einer Gruppe mit zwei Messzeitpunkten formuliert. Auf eine Darstellung der Invarianzhypothesen hierarchischer Strukturgleichungsmodelle wird verzichtet, da diese über die statistischen Hypothesen und Auswertungen der vorliegenden Arbeit hinausgehen.

Zur Überprüfung der Invarianzhypothesen schlagen Vandenberg und Lance (2000) sowie Horn und McArdle (1992) verschiedene Prüfschritte vor, wobei die Überprüfung des nachfolgenden Schrittes nur dann statistisch sinnvoll ist, wenn auf der jeweils vorgenannten Stufe die Invarianz der Parameter bestätigt wurde (siehe dazu auch Anderson & Gerbing 1988).

Tabelle 10: Hypothesen der Testung auf Invarianz von Messungen (nach Vandenberg & Lance 2000)

Parameter	Vergleichbarkeit ...
$\Sigma_{t1} = \Sigma_{t2}$	... der Kovarianzmatrizen zu beiden Messzeitpunkten
$\lambda_{t1} = \lambda_{t2}$	... der Faktorladungen eines Messmodells zu beiden Messzeitpunkten
$\tau_{t1} = \tau_{t2}$	... der Achsenschnittpunkte eines Messmodells zu beiden Messzeitpunkten
$\delta_{t1} = \delta_{t2}$	... der Varianz der Messfehler zu beiden Messzeitpunkten
$\Phi_{t1} = \Phi_{t2}$	... der Varianz eines latenten Faktors zu beiden Messzeitpunkten
$\kappa_{t1} = \kappa_{t2}$	... der Mittelwerte eines latenten Faktors zu beiden Messzeitpunkten

Der *Omnibustest* umfasst die Überprüfung der Gleichheit der Kovarianzmatrizen zu beiden Messzeitpunkten ( $\Sigma_{t1} = \Sigma_{t2}$ ). Hierbei wird angenommen, dass zu zwei Messzeitpunkten eine Übereinstimmung hinsichtlich der Anzahl latenter Faktoren sowie der Stabilität von manifesten Variablen pro latenten Konstrukt gegeben ist. Invarianz bezüglich latenter Faktoren wird auch als *dimensionale Invarianz* und Gleichheit hinsichtlich Items pro Faktor wird auch als *konfigurale Invarianz* bezeichnet (Gregorich 2006). Gleichheit der gesamten Kovarianzmatrizen wird angenommen, wenn die  $\chi^2$ -Statistik sowie weitere Gütemaße – unter der Voraussetzung der Annahme der Stabilität der Messmodelle über die Zeit – eine vergleichbare Passung der Modelle mit den empirischen Daten zu beiden Messzeitpunkten bestätigen. Die Ablehnung der Hypothese des Omnibustest lässt keine Aussage über spezifische Parameterinvarianzen zu. Nach Byrne und Kollegen (1989) erfolgt der Omnibustest, indem das formu-

lierte Messmodell in beiden Gruppen unabhängig voneinander getestet wird. Wird durch diesen Vergleich das Messmodell in beiden Stichproben bestätigt, können beide Stichproben hinsichtlich des untersuchten Modells auch als einheitliche Gruppe bezeichnet werden. Bei Oort (2005a) stellt der Omnibustest eine Voraussetzung für die Response-Shift-Erfassung dar, da auf diese Weise das Basismodell zur weiteren Untersuchung eines Response-Shifts zu beiden Messzeitpunkten auf Passung getestet wird.

Die Überprüfung der *metrischen Invarianz* bezeichnet die Testung auf Invarianz der Faktorladungen, wobei die Faktorladungen einzelner manifester Variablen zu beiden Messzeitpunkten gleichgesetzt werden. Die Hypothese der metrischen Invarianz ist abzulehnen, wenn sich das Ausmaß einer Faktorladung zwischen den Messzeitpunkten statistisch signifikant verändert. Nimmt die Faktorladung einer manifesten Variable zum zweiten Messzeitpunkt statistisch signifikant zu, so erhält die manifeste Variable zum zweiten Messzeitpunkt ein stärkeres Gewicht. Nach Vandenberg und Lance (2000) ist eine Vergleichbarkeit der Messungen nicht gegeben, wenn metrische Invarianz widerlegt wird. Die Konsequenz ist ein Abbruch der weiteren Invarianztestungen. Außerdem steht auch die Vergleichbarkeit der latenten Konstrukte in Frage, da sich diese zwischen den Messzeitpunkten zu sehr verändert haben. Byrne und Kollegen (1989) vertreten den Ansatz, dass unter Voraussetzung einer *partiellen Invarianz* (der Äquivalenz eines Teils eines Messmodells) durchaus weitere Analysen der ausstehenden Invarianztestungen vorgenommen werden können und ein Vergleich von Gruppen- und Messzeitpunktparametern vertretbar ist.

Der Test auf *skalare Invarianz* (Meredith 1993) stellt die Überprüfung der Gleichheit der Achsenschnittpunkte dar. Die Voraussetzung der Invarianztestung dieser Modellparameter beruht auf entweder genauen theoretischen Überlegungen beziehungsweise zu erwartenden Unterschieden in den Achsenschnittpunkten, da in einem Messmodell die Achsenschnittpunkte die Mittelwerte der Variablen darstellen, wenn die latente Variable  $\xi$  den Wert ‚0‘ erhält. Beim Gruppenvergleich bedeutet das, beide Gruppen (beispielsweise Männer und Frauen) haben auf der manifesten Variable  $x_1$  denselben Mittelwert unter der Voraussetzung, dass in beiden Gruppen der Wert für  $\xi$  gleich Null ist. Ist dies theoretisch und empirisch zu erwarten, ist das ein Hinweis auf die Äquivalenz der Ausprägungen der manifesten Variablen bei beiden Gruppen (Geschlechtern).

Die Testung auf *unique Invarianz* bezieht sich auf die Äquivalenztestung der Messfehlerinflüsse. Nach den Annahmen der Klassischen Testtheorie ist davon auszugehen, dass Messfehler-Einflüsse weder mit einer manifesten Variablen, noch mit anderen Messfehlern korreliert sind. In Messmodellen von Strukturgleichungsmodellen geben sie den Anteil an Varianz der manifesten Variablen an, der nicht durch den latenten Faktor erklärt werden kann (Bollen 1989). Deshalb werden die Messfehleranteile einer manifesten Variablen auch als deren Mangel an Reliabilität gewertet (Kline 2005). Raju und Kollegen (2002) bezeichnen die Invarianz der Messfehler auch als „Reliability Invariance“. In der Messfehlervarianz einer manifesten Variablen bildet sich der Einfluss aller nicht kontrollierten Faktoren ab. In Längsschnittstudien mittels Messmodellen werden jedoch Annahmen formuliert, die das Axiom der Unkorreliertheit der Messfehler einer Variablen zu verschiedenen Messzeitpunkten außer Kraft setzen. Neben Einflussfaktoren wie Konzentration, Motivation, momentane Befindlichkeit etc., die im Messfehler enthalten sein können, ist bei einer längsschnittlichen Untersuchung auch davon auszugehen, dass die Messfehler zumindest teilweise korreliert sind, da beide Messungen (häufig, und in der vorliegenden Arbeit ist das der Fall) mit dem gleichen Instrument vorgenommen werden. Das Messinstrument wiederum besitzt eine entsprechende Reliabilität, wobei gewisse Anteile der Messungenauigkeit des Instruments auf dessen Skalierung und andere messmethodische Aspekte zurückzuführen sind. Kline (2005, S.176) führt dazu aus: “Measurement errors reflect two kinds of unique variance: (1) random error of the type estimated by the complements of reliability coefficients (i.e.  $1-r_{xx}$ ), and (2) a systematic variance due to things that the indicator measures besides its underlying factors, such as the effects of a particular method of measurement”. Der Einfluss des Methodeneffekts wird in Längsschnittanalysen mittels KFA beziehungsweise SEM als zwischen den MZP korreliert angenommen. Die Definition des Messfehlers als Kombination aus zufälligen und spezifischen Einflussfaktoren auf die manifeste Variable wird im Linear Latent Variable Model von Oort (2004) ausführlich beschrieben.

Der Test auf *Invarianz latenter Faktoren* überprüft die Gleichheit der Varianz der latenten Konstrukte zwischen zwei Messzeitpunkten ( $\Phi_{t1} = \Phi_{t2}$ ), wobei eine Veränderung der Varianz über die Zeit eine Veränderung der Kalibrierung des wahren Wertes in der Stichprobe beschreibt. Kann die Hypothese der Varianzgleichheit eines latenten Faktors nicht widerlegt werden, so ist davon auszugehen, dass innerhalb der untersuchten Stichprobe das Kontinuum der Ausprägung des latenten Konstrukts gleich geblieben ist. Eine Ablehnung der Annahme weist dementsprechend auf eine Ausweitung oder Einschränkung der Ausprägung des laten-

ten Konstrukts hin. Für die Response-Shift-Erfassung mittels KFA ist die Kontrolle der Varianz des latenten Faktors eine notwendige Voraussetzung zur Erfassung der non-uniformen Rekalibrierung.

Der Test auf *Invarianz der Mittelwerte des latenten Faktors* untersucht die Vergleichbarkeit der Mittelwerte des latenten Faktors über die Zeit ( $\kappa_{t1} = \kappa_{t2}$ ), wobei eine Veränderung des Mittelwertes die Zu- oder Abnahme des Wertes des latenten Konstrukts darstellt.

Auf die genaue Darstellung weiterer Invarianzannahmen (partielle Invarianz etc.) wird verzichtet, da diese für die Fragestellung der vorliegenden Arbeit vernachlässigbar sind. Byrne und Kollegen (1989) bezeichnen die ersten vier Formulierungen als Schritte zur Prüfung der *Invarianz der Messungen*, da es sich um Beziehungen zwischen manifesten und latenten Variablen handelt. Die Schritte fünf und sechs werden als Prüfung der *strukturellen Invarianz* bezeichnet, da sie sich auf die Ebene der latenten Variablen beziehen.

#### *Entwicklung der Invarianztestung im Zusammenhang mit Response-Shift*

Im Folgenden wird die Entwicklung der KFA-Methode zur Response-Shift-Erfassung skizziert. Ausgangspunkt stellt die bereits genannte Arbeit von Golembiewski und Kollegen (1976) dar, die auf der Grundlage der *Explorativen Faktorenanalyse* (EFA) durchgeführt wurde.

Die Studie von Golembiewski und Kollegen (1976) stellt die erste Response-Shift-Untersuchung auf Basis einer faktorenanalytischen Auswertung dar. Es wurden die Auswirkungen der Einführung einer Arbeitszeitgestaltung im Sinne eines Gleitzeit-Modells als Maßnahme zur Organisationsentwicklung untersucht. Als Studiendesign wurde ein Prä-Post-Kontrollgruppen-Design mit einer 12-Monats-Katamnese verwendet. In der Interventionsgruppe und in der Kontrollgruppe wurde zu den drei Messzeitpunkten ein 18-Item-Fragebogen zur Erfassung von Einstellungen erfasst. Untersucht wurde, ob ein *Gamma-Change* stattgefunden hat, der eine Neukonzeptualisierung des Outcome-Kriteriums darstellt. In einem ersten Auswertungsprozess wurden die Daten für jede Gruppe zu jedem Messzeitpunkt mittels explorativer Faktorenanalyse auf deren Faktorenstruktur überprüft. Neben der Variation der Anzahl der Faktoren zu den drei Messzeitpunkten wurde die Kongruenz zwischen den Kovarianzmatrizen mittels eines Ansatzes nach Ahmavaara (1954) überprüft. Dabei wurde für jeweils zwei Mat-

Matrizen geprüft, ob durch Rotationen Faktorenstrukturen der Matrizen herstellbar waren, die sich möglichst ähnlich waren. Durch weitere Matrizen-Transformationen lassen sich Intraclass-Korrelationskoeffizienten sowie Produkt-Moment-Korrelationen zwischen diesen Matrizen berechnen, die als Kongruenzmaß verwendet wurden (Harman 1967). Ergebnisse der Untersuchung waren eine Inkongruenz zwischen den Matrizen der Kontroll- und der Interventionsgruppe einerseits sowie zwischen den Matrizen der Interventionsgruppe im Zeitverlauf. Diese zuletzt genannte Inkongruenz wurde als Hinweis auf eine Neukonzeptualisierung gewertet.

Armenakis (1988, Armenakis & Zmud 1979, Zmud & Armenakis 1977) übernahm die faktorenanalytische Strategie von Golembiewski und Kollegen (1976) und überprüfte die Übereinstimmung der Faktorenstrukturen zu zwei Messzeitpunkten mittels eines Kongruenzkoeffizienten (*Coefficient of Congruence*, Bedian et al. 1988). Lindell und Drexler (1979) kritisierten die Schlussfolgerung, eine veränderte Faktorenstruktur sei immer ein Zeichen für eine Neukonzeptualisierung, und stellten anhand der artifiziellen Manipulation korrelativer Zusammenhänge dar, dass die Ursache für eine solche Veränderung auch auf eine Rekalibrierung beziehungsweise eine Veränderung des zu messenden Konstrukts zurückführbar ist. Gamma-Change sollte dementsprechend gar nicht mit der Methode nach Golembiewski und Kollegen untersucht werden. Als Methode zur Erfassung der Rekalibrierung (beta-change) schlugen die Autoren vor, diese entweder direkt mittels eines dafür konstruierten Fragebogens zu erfragen oder die Endpunkte der zu bewertenden Items durch den Probanden selbst definieren zu lassen (Lindell & Drexler 1980).

Terborg und Kollegen (1980) kombinierten das faktorenanalytische Vorgehen mit dem Thentest. Die Besonderheit des Ansatzes nach Terborg liegt darin, dass einerseits Itemwerte und andererseits Skalenprofile miteinander in Beziehung gesetzt werden. Dabei lassen sich auf individueller Ebene Korrelationen für die Itemwerte einer Skala und Skalenprofile im Prä-Post-, Thentest-Post- und Prä-Thentest-Vergleich berechnen. Korrelieren die Thentest-Post-Werte signifikant höher miteinander als die jeweils anderen beiden Wertepaare, so gehen Terborg und Kollegen von einem Gamma-Change nach Golembiewski aus. Aus Unterschieden zwischen den Skalenprofilen und der Streuung innerhalb der Skala lässt sich eine Rekalibrierung messen. Randolph und Elloy (1989) vergleichen die beiden Prozeduren nach Golembiewski und Terborg zur Response-Shift-Erfassung und kommen zu dem Schluss, dass sie teilweise deutlich unterschiedliche Ergebnisse zeigen, wobei beide Verfahren gleichermaßen von einer objektiv erfassten Veränderung abweichen. Weder aus differenzwertbezogener,

noch aus faktorenanalytischer Sicht stellen die Ansätze von Armenakis und Zmud (1978) sowie Terborg und Kollegen (1980) eine Weiterentwicklung der Response-Shift-Erfassung dar.

Schmitt (1982) verwendet erstmals den statistischen Ansatz des Strukturgleichungsmodells (SEM) zur Erfassung der verschiedenen Response-Shift-Parameter. Die Vorteile dieses Ansatzes liegen einerseits in der Überprüfung der Ähnlichkeiten der gesamten Varianz-Kovarianz-Matrizen zu (wenigstens) zwei Messzeitpunkten sowie einer Differenzierung von Neukonzeptualisierung (Gamma-Change) und Rekalibrierung (Beta-Change) durch die Anwendung eines einzigen statistischen Verfahrens. Außerdem beruht der Ansatz der SEM lediglich auf zwei Messungen der Outcome-Kriterien, ohne zusätzlich Idealskalen oder Retrospektiveinschätzungen zu erfassen. Die Definition der Neukonzeptualisierung entspricht dem Ansatz der Invarianz der Faktoren-Muster (Messmodelle) zu unterschiedlichen Messzeitpunkten. Als Rekalibrierung (Beta-Change) bezeichnet Schmitt einerseits die Invarianz der Faktorladungen. Homogenität der Ladungen der manifesten Variablen auf der latenten Variable über die Zeit bedeutet eine Stabilität der Skalierungen der einzelnen manifesten Variablen, da diese zu verschiedenen Zeitpunkten das gleiche statistische Gewicht für das latente Konstrukt besitzen. Zusätzlich definiert der Autor eine Veränderung der Varianz eines latenten Faktors über die Zeit als Rekalibrierung, wobei diese Veränderung nicht auf eine Intervention zurückzuführen sein darf (dann wäre es ein *Alpha-Change* bzw. ein angestrebter Interventionseffekt). Als weitere Rekalibrierungsvariante bezeichnet Schmitt die Veränderung der Messfehlervarianzen einer manifesten Variablen über die Zeit. Da der Messfehler statistisch in direktem Zusammenhang mit der Reliabilität der gemessenen manifesten Variablen steht, wird die Varianz der Messfehler als Variation in der Skalierung der manifesten Variablen angesehen, wobei die Voraussetzung erfüllt sein muss, dass die Varianz der entsprechenden latenten Variable invariant über die Zeit ist. Der Autor schlägt folgendes Vorgehen mittels SEM zur Erfassung von Rekalibrierung und Neukonzeptualisierung vor: (1) allgemeiner Test auf Homogenität der Varianz-Kovarianz-Matrizen, (2) Definition invarianter Faktorstrukturen zu verschiedenen Messzeitpunkten, (3) Definition invarianter Faktorladungen sowie Faktorvarianzen zu verschiedenen Messzeitpunkten und (4) Testung der Invarianz der Messfehlereinflüsse. Nach Schmitt und Kollegen (1984) führt die Response-Shift-Überprüfung mittels SEM zu vergleichbaren Ergebnissen wie die Strategie nach Terborg und Kollegen (1980). Da der SEM-Ansatz jedoch die Durchführung zusätzlicher Datenerhebungen (in diesem Fall die Retrospektivmessung) überflüssig macht, ist dieser bei der Response-Shift-Erfassung zu bevorzugen. Weiterhin resultiert aus der Anwendung des SEM die Möglichkeit, einen statistischen

Cut-off-Wert für die Invarianztestungen und der somit verglichenen Parameter zu bestimmen. Der  $\chi^2$ -Differenztest ermöglicht die Überprüfung der Modellpassung im Rahmen der Invarianztestungen, so dass Unterschiede zwischen geschachtelten Modellen auf ihre statistische Bedeutsamkeit getestet werden können. Kritisch am Vorgehen nach Schmitt ist anzumerken, dass die Parameterfreisetzungen nicht entsprechend ihrer Bedeutung für die Modellpassung geprüft werden, sondern gemäß dem Ansatz nach Invarianztestung bei unterschiedlichen Gruppen (im Querschnitt). Dieses Vorgehen ist möglicherweise weniger sensitiv für die Aufdeckung von Response-Shift-Prozessen. Außerdem wird Response-Shift nach dieser Methode nur auf Ebene der Kovarianzen geprüft. Mittelwerte werden nicht berücksichtigt.

Millsap und Hartog (1988) veröffentlichten eine konzeptionelle Erweiterung des SEM-Ansatzes nach Schmitt (1982). Zur genauen Erfassung eines Response-Shifts postulieren die Autoren die Anwendung eines Prä-Post-Kontrollgruppen-Designs. Neukonzeptualisierung im Sinne eines Gamma-Change muss nach Ansicht der Autoren nicht nur Resultat einer Intervention sein, sondern kann auch unabhängig von einer geplanten Maßnahme in einer Kontrollgruppe stattfinden (beispielsweise durch andere Umweltfaktoren). Es wird unterschieden zwischen einer differentiellen und einer parallelen Neukonzeptualisierung. Ausgangspunkt dieser Überlegungen ist die Annahme identischer Faktorenstrukturen der zu prüfenden Messmodelle in der Interventions- und der Kontrollgruppe zum Messzeitpunkt 1. *Differentielle Neukonzeptualisierung* liegt dann vor, wenn die Faktorenstruktur des Messmodells zum Postmesszeitpunkt zwischen den Gruppen statistisch signifikant unterschiedlich ist. *Parallele Neukonzeptualisierung* ist dann gegeben, wenn sich die Messmodelle zum Postmesszeitpunkt nicht zwischen den Gruppen unterscheiden, wohl aber in beiden Gruppen statistisch signifikante Unterschiede zu den Messmodellen der Prämessung bestehen. Das Ergebnis einer differentiellen Neukonzeptualisierung ist im Sinne eines Interventionseffektes interpretierbar, die parallele Form jedoch nicht. Wenn die Faktorenstrukturen beider Gruppen zum Prämesszeitpunkt statistisch signifikant voneinander verschieden sind, ist zwar eine Neukonzeptualisierung nachweisbar, aber ein Zusammenhang mit der Intervention nicht.

Die bisher beschriebenen Response-Shift-Erfassungen mittels Faktorenanalyse oder SEM fokussieren lediglich auf die zugrunde liegenden Varianzen und Kovarianzen und lassen die Vergleiche von Mittelwerten außer Acht. Oort (2005a) bezieht diese Strategie in seine Response-Shift-Erfassung mit ein, was Auswirkungen auf die Operationalisierung der jeweiligen Response-Shift-Prozesse hat. Dabei handelt es sich bei der Ablehnung der Hypothese der In-

varianz der Faktorladungen zu zwei Messzeitpunkten entweder um eine Neukonzeptualisierung oder eine Repriorisierung. Eine *Neukonzeptualisierung* ist dann gegeben, wenn die Faktorladung einer manifesten Variablen zu einem der beiden Messzeitpunkte bei (nahe) Null liegt und zum anderen Messzeitpunkt eine für die Annahme des Messmodells angemessene Größe besitzt. Inhaltlich bedeutet eine solche Veränderung einer Faktorladung, dass die dazugehörige manifeste Variable lediglich zu einem Messzeitpunkt (k)eine Bedeutung besitzt und somit das latente Konstrukt sich *qualitativ* verändert hat. Eine *Repriorisierung* ist dann anzunehmen, wenn sich die Faktorladungen zu beiden Messzeitpunkten stark unterscheiden, ohne dass eine Ladung zu einem der beiden Messzeitpunkte einen Wert nahe oder gleich Null annimmt. Inhaltlich bedeutet diese Veränderung eine Abnahme oder Zunahme der Bedeutung der entsprechenden manifesten Variablen für das latente Konstrukt.

Nach Oort (2004, 2005a) würde bei der Non-Äquivalenz der Achsenschnittpunkte dann von einem Response-Shift im Sinne einer *uniformen Rekalibrierung* gesprochen, wenn sich die Achsenschnittpunkte einer manifesten Variablen unabhängig vom Mittelwert des latenten Konstrukts ändern. Der Wert einer Skala besitzt demnach zu beiden Messzeitpunkten eine unterschiedliche (quantitative) Bedeutung. Die uniforme Rekalibrierung ist ein Prozess, in dem alle Stufen einer Skala in der Stichprobe eine andere Bedeutung erhalten, das heißt die befragten Personen skalieren die gesamte Skala einer beobachtbaren Variable zu beiden Messzeitpunkten unterschiedlich, wobei die Differenz zwischen den Skalenabstufungen gleich bleibt. In diesem Sinne kann von einer homogenen ‚Verschiebung‘ aller Skalenabstufungen gesprochen werden. Die *non-uniforme Rekalibrierung* wird operationalisiert durch die Veränderung der Varianz der beobachtbaren Variablen. Diese Form der Rekalibrierung liegt dann vor, wenn zwischen zwei Messzeitpunkten eine Varianzänderung der manifesten Variablen im Messmodell stattfindet und diese nicht im Zusammenhang mit einer Varianzveränderung der latenten Variable steht. Demgemäß muss bei der Überprüfung der non-uniformen Rekalibrierung der Varianzeinfluss der latenten Variablen auf die Varianz der manifesten Variablen kontrolliert werden. Inhaltlich bedeutet diese Form der Rekalibrierung, dass die Personen die Skalierung der manifesten Variablen zum zweiten Messzeitpunkt uneinheitlich (neu) adjustieren, also nur einigen Skalenpunkten eine andere (quantitative) Bedeutung beimessen. Dem Autor zufolge zeigt sich eine uniforme Rekalibrierung darin, dass *lediglich* eine Interceptfreisetzung vorzunehmen ist. Die Messfehlervarianz der manifesten Variablen bleibt in diesem Fall restringiert. Eine non-uniforme Rekalibrierung kann sich *entweder* in der Form



der Freisetzung der Messfehlervarianz *oder* in der Freisetzung der Kombination aus Messfehlervarianz und Intercept einer Skala äußern.

Oort (2005a) erklärt die statistische Operationalisierung der beiden Typen der Rekalibrierung ausgehend von der Invarianz des Messfehleranteils. Dieser setzt sich per Definition aus einem so genannten spezifischen (oder *uniquen*) Anteil ( $\eta$ ) und einem reinen Fehleranteil ( $\varepsilon$ ) zusammen. Im spezifischen Anteil des Messfehlers sind alle Einflüsse auf die entsprechende Skala enthalten, die im Zusammenhang mit der gemessenen Variable stehen, jedoch nicht durch die latente Variable erklärt werden. Operationalisiert wird dieser spezifische Anteil der Skala jedoch nicht mehr im Messfehleranteil, sondern im Achsenschnittpunkt. Er stellt die Ausprägung der Skala, die erreicht wird, wenn die latente Variable keinen Einfluss auf die Skala hat. Messfehlervarianz operationalisiert nach Oort (ebd.) die Varianz dieser spezifischen Faktoren. Auf diesen Zusammenhang soll im folgenden Exkurs nochmals ausführlicher eingegangen werden.

#### *Exkurs: Linear Latent Variable Model nach Oort*

Die Response-Shift-Prozesse nach Sprangers und Schwartz (1999) werden von Oort (2004) anhand des sog. *Linear Latent Variable Model* (LLVM) operationalisiert, das sich aus der Item-Response-Theorie (IRT) und dem Ansatz der Faktorenanalyse zusammensetzt. Auf eine ausführliche Darstellung der Grundlagen der IRT wird hier verzichtet und auf entsprechende Literatur verwiesen (bspw. Steyer & Eid 2001). Nach der Item-Response-Theorie hängt die Beantwortung eines Items von Item- und Personenparametern ab. Dabei sind die Merkmale der Items per definitionem Parameter, die sich *nur* auf Itemcharakteristika beziehen wie beispielsweise die Itemschwierigkeit. Merkmale der Personen sind demgemäß Parameter, die sich *ausschließlich* auf personale Variablen beziehen (bspw. die Ausprägung der physischen Funktionsfähigkeit). Das Linear Latent Variable Model geht von folgender Gleichung für eine Person (i) aus:

$$x_i = \tau + \Gamma \xi_i + \lambda \eta_i + \varepsilon_i \quad (5)$$

$x_i$  = Wert / Vektor einer beobachteten Variable

$\xi_i$  = Wert / Vektor eines latenten Konstrukts

$\eta_i$  = Wert / Vektor eines (latenten) spezifischen (*uniquen*) Wertes des Faktors

$\varepsilon_i$  = Wert / Vektor eines (latenten) Fehlerwertes des Faktors

$\tau$  = Intercept

$\Gamma$  = Regressionskoeffizient des latenten Konstrukts

$\lambda$  = Regressionskoeffizient des spezifischen (uniquen) Faktorwerts

Die Personenparameter sind nach der IRT beziehungsweise dem LLVM (Gleichung 5) das latente Konstrukt ( $\xi$ ), der spezifische Faktorwert ( $\eta$ ) und der Fehlerwert ( $\varepsilon$ ). Die Itemparameter sind entsprechend die Koeffizienten des Intercepts ( $\tau$ ), des Regressionskoeffizienten des latenten Konstrukts ( $\Gamma$ ) und der Regressionskoeffizient des spezifischen (uniquen) Faktorwerts ( $\lambda$ ). Diese Itemparameter können in zweifacher Hinsicht interpretiert werden: einmal im Sinne des faktorenanalytischen Ansatzes, der ein Intercept ( $\tau$ ) als Achsenschnittpunkt und die Regressionskoeffizienten ( $\Gamma$ ) als Ladungen des Faktors sowie ( $\lambda$ ) des spezifischen Anteils des Faktors definiert. In der IRT stellt der Parameter  $\tau$  die Itemschwierigkeit dar und der Regressionskoeffizient  $\Gamma$  den Diskriminationsparameter. Die Itemschwierigkeit gibt an, wie einfach es ist, auf einem Item einen hohen Wert zu erzielen. Der Diskriminationsparameter gibt an, wie genau die Ausprägung der beobachteten Variablen Unterschiede in der Ausprägung des latenten Konstrukts abbilden kann.

Der spezifische Faktorwert sowie dessen Regressionskoeffizient sind in der allgemeinen IRT sowie in der KTT nicht enthalten. Oort (2004) geht jedoch davon aus, dass diese genauso interpretierbar sind, wie das latente Konstrukt und dessen Regressionskoeffizienten. Zum spezifischen Faktor gehören nach Oort alle Aspekte und Inhalte der beobachteten (manifesten) Variablen, die nicht mit dem latenten Faktor zusammenhängen, jedoch auch nicht auf Messfehlerinflüsse zurückzuführen sind (siehe dazu auch Meredith & Teresi 2006). Der Autor geht weiterhin von verschiedenen Annahmen aus: Das latente Konstrukt ( $\xi$ ) ist unkorreliert mit dem spezifischen Faktorwert ( $\eta$ ) und dem Messfehler ( $\varepsilon$ ). Der spezifische Faktorwert ist unkorreliert mit dem Messfehler und die Messfehler sind untereinander unkorreliert. Der Mittelwert eines Messfehlers ergibt Null. Dementsprechend lässt sich folgende Gleichung auf Mittelwertsebene beschreiben:

$$\text{MEAN}(y) = \mu = \tau + \Gamma\alpha + \Lambda\beta \quad (6)$$

$\alpha$  = Mittelwert des latenten Konstrukts

$\beta$  = Mittelwert des spezifischen Faktors

Die Kovarianz der beobachteten Variablen lässt sich entsprechend folgender Gleichung darstellen:

$$\text{COV}(y, y') = \Sigma = \Gamma\Phi\Gamma' + \Lambda\Psi\Lambda' + \Theta \quad (7)$$

$\Phi = \text{COV}(\xi, \xi')$  = Varianz-Kovarianzmatrix der latenten Konstrukte

$\Psi = \text{COV}(\eta, \eta')$  = Varianz-Kovarianzmatrix der spezifischen Faktoren

$\Theta = \text{COV}(\varepsilon, \varepsilon')$  = Varianz-Kovarianzmatrix der Messfehler

Die einzelnen Varianz-Kovarianzmatrizen lassen sich noch weiter zerlegen in Produkte aus Diagonalmatrizen und der entsprechenden Korrelationsmatrix des jeweiligen Konstrukts beziehungsweise deren Subparameter. Für  $\Phi$  und  $\Psi$  gilt:

$$\Phi = \sqrt{\text{Diag}(\Phi)} \Phi^* \sqrt{\text{Diag}(\Phi)} \quad (8)$$

und

$$\Psi = \sqrt{\text{Diag}(\Psi)} \Psi^* \sqrt{\text{Diag}(\Psi)} \quad (9)$$

Nach dem LLVM sind vier unterschiedliche Parameter beschreibbar: *Personenparameter*, *Itemparameter*, *Mittelwerte* und *Kovarianzen*. Auf die Darstellung der Verknüpfung der Parameter über zwei Messzeitpunkte wird hier verzichtet und auf einschlägige Literatur verwiesen (Oort 2004).

Die Grundvoraussetzung für die Identifikation des Messmodells bilden die folgenden, an allgemeinen Konventionen orientierten Parameterdefinitionen: Der Mittelwert des latenten Konstrukts wird auf 0 gesetzt und die Varianz des latenten Konstrukts auf 1. Das gleiche gilt für alle anderen latenten Variablen (spezifischen Faktoren und Messfehler). Die entsprechenden Mittelwerte und Varianzen des MZP 2 werden an den Werten zum MZP 1 standardisiert. Nach Oort (2004) können nun vier verschiedene Response-Shift-Prozesse anhand des LLVM erfasst werden<sup>1</sup>: Die statistische Überprüfung der beiden Response-Shift-Prozesse Rekonzeptualisierung und der Repriorisierung wurde bereits im vorangegangenen Abschnitt

---

<sup>1</sup> Oort (2004) unterscheidet insgesamt 11 verschiedene Möglichkeiten, unterschiedliche Formen des Response-Shift nachweisen zu können, wobei auf eine ausführliche Darstellung der statistischen Hintergründe an dieser Stelle verzichtet und auf die angegebene Literatur verwiesen wird.

ausführlich dargestellt. An dieser Stelle soll das Vorgehen zur Unterscheidung zwischen uniformer und non-uniformer Rekalibrierung beschrieben werden.

Rekalibrierung nach Sprangers und Schwartz (1999) wird definiert als die Veränderung interner Standards, die als Grundlage für die Bewertung von Items beziehungsweise Skalen anzusehen sind. Diese Veränderung kann alle Antwortoptionen gleichermaßen betreffen (uniforme Rekalibrierung), so dass sich diese lediglich auf den Mittelwert des gemessenen Items beziehungsweise der Skala auswirkt (die Varianz bleibt gleich). Verändert sich jedoch nicht die Bewertung aller Skalenstufen gleichermaßen in dieselbe Richtung, so findet eine uneinheitliche Veränderung der Skalenstufen oder der Ausgangsskalierung statt (non-uniforme Rekalibrierung). Dies soll an folgendem Beispiel verdeutlicht werden: Wenn zwischen zwei Messzeitpunkten eine Veränderung des Mittelwertes in der Skala Soziale Funktionsfähigkeit stattfindet, ohne dass diese Veränderung auf die Veränderung des latenten Konstrukts GLQ zurückzuführen ist, dann sprechen Sprangers und Schwartz (1999) davon, dass sich die internen Standards zur Bewertung der Items beziehungsweise der Skala verändert haben. Möglicherweise haben die befragten Personen zwischen den MZPen besondere Erfahrungen in ihrem sozialen Umfeld gemacht oder sich an der Bewertung anderer orientiert, so dass die vormals festgelegte Einteilung der Sozialen Funktionsfähigkeit (bzw. deren Items) in beispielsweise ‚mäßig‘ oder ‚schwer‘ mit anderen quantitativen (oder auch inhaltlichen) Zuständen verknüpft worden sind. Werden alle Skalenstufen in derselben Art und Weise verändert (ändert sich also auch nichts an deren Differenzen), ist die Rekalibrierung uniform. Ändern die befragten Personen jedoch nur *einige* der Skalenstufen (bspw. die beiden gerade genannten, alle anderen Abstufungen bleiben jedoch beim Niveau des ersten MZP), dann ist die Rekalibrierung non-uniform.

Wie oben anhand der testtheoretischen Darstellungen beschrieben, ist der Parameter  $\tau$  ein Intercept (KTT) beziehungsweise die Itemschwierigkeit (IRT). Für die Operationalisierung der Interceptveränderungen als uniforme Rekalibrierung nach Ort spricht die Definition dieses Parameters als Itemparameter beziehungsweise Parameter des Messinstruments. Objektiv hat sich jedoch zwischen den Messungen an dem Item beziehungsweise dem Messinstrument nichts geändert, so dass die Veränderung im Intercept (bzw. der Itemschwierigkeit), die nicht auf die Veränderung der latenten Variable zurückzuführen ist, eine Veränderung eines personenbezogenen Merkmals ausdrückt. Ort sieht darin die Veränderung des spezifischen Faktors der gemessenen Funktion, die nicht auf die GLQ zurückgeführt werden kann. Zur Erfas-

sung der uniformen Rekalibrierung ist nach Angaben von Oort (2004) die Faktorenanalyse in der Lage. Zwar wird bei der Faktorenanalyse nicht zwischen einem spezifischen Faktor der manifesten Variable und dem Messfehler unterschieden ( $\zeta = \Lambda\eta + \varepsilon$ ). Dennoch lässt sich nach dem LLVM auf der Mittelwertebene die uniforme Rekalibrierung erfassen. Dem liegt die Annahme der KTT zugrunde, dass der Mittelwert eines Messfehlers Null ist. Somit bildet sich im Intercept der Einfluss des spezifischen Faktors ab. Formal nimmt der Autor die Invarianz der Intercepts (als Instrumentenparameter) an, um die Invarianz des spezifischen Faktors zu messen. Diese formale Definition ist notwendig, um die Messmodelle nach der Gleichung (6) identifizierbar zu machen.

Die non-uniforme Rekalibrierung wird nach Oort operationalisiert durch die Varianz des spezifischen Faktors (Diag  $\Psi$ ). Da davon ausgegangen wird, dass reine Messfehleranteile miteinander unkorreliert sind (KTT) und somit keinen Einfluss auf die Mittelwerte der beobachtbaren Variablen haben (siehe Gleichung 6), kommen nach dem LLVM nur die Ladung oder die Varianz des spezifischen Faktors als Operationalisierungen für die non-uniforme Rekalibrierung in Frage. Nach der Definition der LLVM ist die Ladung des spezifischen Faktors jedoch ein Itemparameter beziehungsweise ein Parameter des Messinstruments, wobei das Instrument zu beiden MZPen identisch ist. Demzufolge wird die non-uniforme Rekalibrierung durch die Varianzveränderung des spezifischen Faktors operationalisiert. Auch zur Erfassung der non-uniformen Rekalibrierung kann die Faktorenanalyse herangezogen werden, obwohl sie formal den spezifischen Faktor nicht operationalisiert. Hier gilt die oben genannte Differenzierung des Messfehlers  $\zeta (= \Lambda\eta + \varepsilon)$  ebenso. Nach der Klassischen Testtheorie ist davon auszugehen, dass die Varianz-Kovarianzmatrix (COV  $\zeta\zeta'$ ) sich zusammensetzt aus der Varianz-Kovarianzmatrix des spezifischen Faktors ( $\Lambda\Psi\Lambda'$ ) und der Varianz-Kovarianzmatrix des Messfehlers ( $\Theta$ ), die jedoch Null ergibt. Somit reduziert sich die Varianz des Messfehlers  $\zeta$  rein theoretisch auf  $\Lambda\Psi\Lambda'$ . Somit sind nach Oort (2004) alle Response-Shift-Parameter mit Hilfe der Konfirmatorischen Faktorenanalyse überprüfbar.

#### *Weitere statistische Verfahren zur Response-Shift-Analyse*

Es wurden mittlerweile auch andere statistische Auswertungsstrategien zur Response-Shift-Erfassung eingesetzt. Dazu zählen Wachstumsanalysen, Residualanalysen und Hierarchisch Lineare Modelle. Diese werden im Folgenden kurz beschrieben.

Der Einsatz von Wachstumsanalysen wurde bereits von Schwartz (1999) beschrieben. Auch Brossart und Kollegen (2002) verwenden diesen Ansatz. Dabei werden über die Kalkulation individueller Verläufe von Variablen personenbezogene Steigungskoeffizienten berechnet und individuelle Veränderungsverläufe abgebildet. Anschließend können verschiedene Variablen miteinander verglichen werden. Schwartz (1999) setzt dieses Verfahren im Rahmen der Analyse von Veränderungen der physischen Funktionsfähigkeit und dem subjektiven Wohlbefinden ein. Brossart und Kollegen (2002) untersuchen Krankheitsverläufe und die Ausprägung der negativen Affektivität im Zusammenhang mit Response-Shift. Differenzen in den Verläufen der untersuchten Variablen können als Hinweis auf einen Response-Shift dienen, wenn alternative Erklärungen ausgeschlossen werden können. So wäre beispielsweise denkbar, dass der körperliche Zustand einer Person sich im Verlauf verschlechtert, die psychische Befindlichkeit jedoch stabil bleibt oder positive Tendenzen aufweist. Dies wäre ein möglicher Hinweis auf einen Response-Shift. Der Nachteil dieser Methode liegt darin, dass aufgrund der Differenzen in den Verläufen der verschiedenen Parameter keine Aussage über die Art eines Response-Shifts möglich ist. Ob die Unterschiede in den Verläufen auf eine Rekalibrierung, Repriorisierung oder Neukonzeptualisierung zurückzuführen sind muss durch den Einsatz (wenigstens) eines weiteren Verfahrens geklärt werden (Schwartz & Sprangers 2000).

Mit Hilfe von Residualanalysen, einer weiteren statistischen Methode, versuchen Mayo und Kollegen (2008) einen Response-Shift bei Schlaganfallpatienten nachzuweisen. Sie ermittelten eine berichtete und vorhergesagte GLQ und Differenzwerte (individuelle Residualwerte) aus den beiden GLQ-Einschätzungen. Die berichtete GLQ wurde mittels der EQ-5D VAS erfasst (EuroQol Group 1990) und Prädiktoren der GLQ aus einer Reihe von Lebensqualitäts- und Schlaganfalldiagnostikinstrumenten mit Angaben zu Funktionsfähigkeiten und Symptomen ermittelt. Außerdem wurden Angaben zur Krankheitsschwere sowie soziodemographische Variablen als Prädiktoren aufgenommen. Prädiktoren wurden einmalig zu Studienbeginn und GLQ jeweils zu allen drei MZPen erfasst. In einem ersten Schritt wurde aufgrund der Prädiktoren der GLQ für jede Person ein individueller Vorhersagewert der GLQ für den jeweils nächsten MZP berechnet. Anschließend wurde geprüft, inwieweit der Vorhersagewert vom tatsächlichen Wert abweicht. Bei großer Abweichung (hohem, am Ausgangswert zentrierten Residualwert) wurde ein Response-Shift angenommen. Es wurde speziell die Repriorisierung erfasst, indem geprüft wurde, ob die Bedeutung der Prädiktoren für die berichtete GLQ zu den verschiedenen MZPen variieren. Anschließend wurden noch Analysen mit

latentem Klassen je nach Response-Shift-Zugehörigkeit durchgeführt und mit den Werten einer Then-Test-Analyse verglichen. Die Autoren kommen zu dem Ergebnis, dass einerseits bei Schlaganfallpatienten nur bei einer Minderheit ein Response-Shift auftritt und andererseits die Methode der Residualanalyse geeignet ist, einen Response-Shift zu erfassen. Es konnte eine hohe Übereinstimmung zwischen den Ergebnissen der Residualanalyse und der Then-Test-Messung festgestellt werden. Kritisch bleibt anzumerken, dass die Methode der Residualanalyse eine Response-Shift-Erfassung auf individueller Ebene ermöglicht, jedoch keine Aussage über die Art des Response-Shifts zulässt. Erst durch den Einsatz zusätzlicher Verfahren beziehungsweise zusätzlicher Prädiktorvariablen zu den MZPen kann auf eine Repriorisierung geschlossen werden.

Eine ähnliche Form der Erfassung von Response-Shift ist die Verwendung von Hierarchisch Linearen Modellen (HLM, Lowy & Bernhard 2004). Nach dieser Methode wird die GLQ mittels eines entsprechenden Instrumentes erfragt, so dass sie als Kriteriumsvariable vorliegt. Weiterhin wurden Prädiktoren der GLQ ermittelt, die zur Erklärung der GLQ herangezogen wurden. Die Messungen der einzelnen Personen zu drei verschiedenen MZPen bilden die Ebene 1 des HLM, die Personen als übergeordnete Einheit die Ebene 2. Mit Hilfe der HLM ist es somit möglich, personenbezogene Auswertungen über Veränderungen der Beziehungen zwischen der GLQ und den Prädiktoren zu erfassen und personenbezogene Wachstumskurven abzuleiten. Zur Analyse der Repriorisierung auf Personenebene wurden die Intercepts und Steigungskoeffizienten auf Ebene 2 nicht fixiert, um die Veränderung der GLQ pro Person erfassen zu können. Die Parameter auf Ebene 1 wiederum wurden fixiert, um die Hypothese zu testen, dass die Bedeutung der Prädiktoren zwischen den MZPen identisch ist. Interaktionseffekte zur Testung des Einflusses von MZP und Prädiktor wurden ebenfalls getestet. Als wesentliche Ergebnisse konnten die Autoren nachweisen, dass die Bedeutung des Prädiktors ‚Functional Performance‘ vom MZP 1 zum MZP 3 statistisch signifikant zunahm und die des Prädiktors ‚Appetite and Energy‘ statistisch signifikant abnahm (Bernhard et al. 2004). Eine Rekalibrierung kann mit dem Verfahren nicht erfasst werden.

Als wesentliche Kritikpunkte an den gerade beschriebenen Verfahren lassen sich zwei Aspekte anführen: Die Verfahren erfassen nicht alle drei Response-Shift-Prozesse gleichzeitig. Aufgrund der Ergebnisse ist entweder nur eine Form des Response-Shift genau erfasst oder es kann nur eine Aussage darüber getroffen werden, ob überhaupt ein Response-Shift stattgefunden hat. Somit ist zweitens für die vollständige Erfassung der Response-Shift-Pro-

zesse nach Sprangers und Schwartz (1999) die Anwendung (wenigstens) eines weiteren Verfahrens der Response-Shift-Erfassung notwendig. Die genannten statistischen Verfahren (Wachstumsanalysen, Residualanalysen, HLM) sind in der bisherigen Anwendung unökonomischer als die KFA-Methode nach Oort.

#### 3.2.4 Exkurs: Veränderungsmessung und Response-Shift

Response-Shift stellt eine Veränderung interner Standards, Werte oder Konzepte einer Person oder Gruppe von Personen durch die Auseinandersetzung mit einem kritischen Lebensereignis dar. Somit ist das Phänomen für die Veränderungsmessung von zentraler Bedeutung. Nach Gollwitzer und Jäger (2007) ist Response-Shift als ein methodisches Problem bei der Durchführung von Evaluationsstudien mit Veränderungsmessungen zu betrachten. Response-Shift wird von den Autoren explizit als Messwiederholungsartefakt bezeichnet. Im Folgenden werden verschiedene Strategien der Veränderungsmessung skizziert. Dabei werden die drei Ansätze der direkten, indirekten und quasi-indirekten Veränderungsmessung vorgestellt. In diesem Zusammenhang werden Bezüge zur statistischen Signifikanz und klinischen Bedeutsamkeit thematisiert, da Response-Shift für die Interpretation üblicher Effektstärkeparameter Probleme aufwirft. Anschließend wird auf die Änderungssensitivität als Eigenschaft von Instrumenten zur Erfassung von Veränderungen im Rahmen evaluativer Verlaufsstudien eingegangen.

##### *Methoden der Veränderungsmessung*

Allgemein werden drei unterschiedliche Methoden zur Erfassung von Veränderungen diskutiert. Dabei handelt es sich um die direkte, indirekte und quasi-indirekte Veränderungsmessung (Nübling et al. 2004, Michalak et al. 2003, Schulte 1993, Steyer et al. 1997). Die Unterscheidung zwischen den Ansätzen der direkten und indirekten Veränderungsmessung ist auch für die Interpretation der Ergebnisse zu Lebensqualitätsbefragungen von Bedeutung. Die direkte Veränderungsmessung ist ein sog. ankerbasierter, die indirekte Veränderungsmessung hingegen ein verteilungsbasierter Ansatz zur Interpretation von Lebensqualitätsveränderungen.

Ziel der *direkten Veränderungsmessung* ist die Erfassung der subjektiven Einschätzung einer Person, (1) ob, und wenn ja, (2) in welchem Ausmaß eine Veränderung der GLQ über einen definierten Zeitraum stattgefunden hat. Der bekannteste Ansatz dieser Art der Verände-



rungsmessung ist die Erfassung der *Minimalen Klinisch-Bedeutsamen Differenz* (minimal clinically important difference, *MCID*) nach Jaeschke und Kollegen (1989). Die Autoren definieren die MCID als „... the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management“ (ebd., S 408). Zur Erfassung der MCID wird zuerst eine globale Einschätzung der Veränderung (in einem definierten Zeitraum) erfragt. Der befragten Person stehen die Antwortalternativen *Verschlechterung*, *keine Veränderung* und *Verbesserung* zur Verfügung. Hat eine positive oder negative Veränderung stattgefunden, soll die Person auf einer siebenstufigen Likertskala das Ausmaß der Veränderung einschätzen. Diese Skala ist von einer ‚geringfügigen Verschlechterung/Verbesserung‘ (-/+1) bis zur ‚massiven Verschlechterung/Verbesserung‘ (-/+7) kalibriert, so dass eine Veränderung insgesamt auf einer 15-stufigen Ratingskala (massive Verschlechterung (-7) bis massive Verbesserung (+7)) abgebildet wird. Die Autoren gehen weiter davon aus, dass der Wertebereich der Skala von -3 bis -1 beziehungsweise +1 bis +3 eine *kleine*, der Bereich von -5 bis -4 sowie von +4 bis +5 eine *mittlere* und der von -7 bis -6 und von +6 bis +7 eine *große* Veränderung anzeigt (siehe dazu auch Guyatt et al. 2002). Diese Form der Veränderungsmessung wird häufig in der klinischen Praxis verwendet, wobei vor allem die Einfachheit der Erfassung sowie die Ökonomie der Auswertung Vorteile dieser Methode sind (Fischer et al. 1999). Der MCID-Ansatz zählt zu den *ankerbasierten Ansätzen* der Erfassung von Lebensqualitätsveränderungen. Es wird davon ausgegangen, dass die befragte Person ihren aktuellen Zustand mit dem zu Beginn des vorgegebenen Zeitraums in Beziehung setzt, die Veränderung demnach anhand der retrospektiven Einschätzung „ankert“. Es existieren jedoch empirische Belege dafür, dass diese zuletzt genannte Annahme kritisch zu bewerten ist. Die Einschätzung der Veränderung korreliert oftmals nur mit der Zustandsmessung zum aktuellen Messzeitpunkt, nicht aber mit dem Zustand zu Beginn des Zeitraums, der für die Beurteilung der Veränderung relevant ist (Wyrwich & Tardino 2006, Norman et al. 1997, Cella et al. 2002). Daher ist anzunehmen, dass das aktuelle Befinden den Veränderungsscore unverhältnismäßig deutlicher beeinflusst als das zurückliegende. Ein möglicher Grund dafür ist darin zu sehen, dass zurückliegendes Befinden nicht mehr gut erinnert werden kann (Norman et al. 2001, Guyatt et al. 2002). Wyrwich und Kollegen (2005) kritisieren an dieser Methode weiterhin, dass es sich um Ein-Item-Ansätze handelt, die keine weitere psychometrische Differenzierung ermöglichen, dass die Abstufung in kleine, mittlere und große Veränderung eine künstliche Einteilung der MCID-Entwickler ist und dass die Erfassung nicht an klinischen Kriterien ausgerichtet ist, die eigentlich als Ankerpunkte dienen. Krause und Jay (1994) verweisen zusätz-

lich darauf, dass der Bewertungshintergrund für die Beantwortung der Items zwischen den befragten Personen enorm variiert und somit kaum inhaltliche Aussagen über die Veränderung gemacht werden können.

Die *indirekte Veränderungsmessung* entspricht der klassischen Prä-Post-Messung. Zu zwei Messzeitpunkten werden Zustandserfassungen mit demselben Instrument durchgeführt und aus den jeweiligen Ergebnissen ein Differenzwert gebildet. Indirekt ist diese Methode, da aus keiner der erfassten Zustandsmessungen allein eine Aussage über eine Veränderung abgeleitet werden kann. Die Differenz gilt jedoch als Maß der Veränderung. Diese Strategie ist weit verbreitet in wissenschaftlichen Wirksamkeitsstudien und bildet die Grundlage für die sog. *verteilungsbasierten Ansätze* der Veränderungsmessung (Lydick & Epstein 1993, Wyrwich et al. 2005). Crosby und Kollegen (2003) geben einen Überblick über verteilungsbasierte Ansätze der Veränderungsmessung. Dazu zählen der t-Test für abhängige Stichproben, Latent Growth Curve Analysen, Effektstärken, der Reliable Change Index und der Standard Error of the Mean (SEM). Hier soll auf das Konzept der Effektstärkeberechnung eingegangen werden, da es für die vorliegende Arbeit von Bedeutung ist. Drei Effektstärkevarianten werden im Zusammenhang mit der indirekten Veränderungsmessung häufig diskutiert: die *Standardized Effect Size* (SES, oder auch ES), der *Standardized Response Mean* (SRM) und *Guyatt's Responsiveness Index* (GRI). Grundlage dieser Effektstärkeberechnungen bildet das Prinzip, dass die Mittelwertsdifferenz aus zwei Messungen an einem Streuungsmaß relativiert wird (Igl et al. 2005). Der SES/ES verwendet als Streuungsmaß die Standardabweichung der Messwerte zum Messzeitpunkt 1 ( $SD_{t1}$ ). Beim SRM wird die Differenz zwischen den beiden Mittelwerten der Messungen an der Streuung der Messwertdifferenzen standardisiert ( $SD_{t2-t1}$ ). Der GRI relativiert die Differenz zwischen zwei gemessenen Mittelwerten ( $t_1, t_2$ ) im Falle eines Ein-Gruppen-Prä-Post-Designs an der Standardabweichung der Differenzwerte einer Baselinephase zu einem vorgezogenen Messzeitpunkt  $t_0$  und der Prämessung  $t_1$  ( $SD_{t1-t0}$ ). Die Besonderheit des GRI ist darin zu sehen, dass die gemessene Mittelwertsdifferenz an zufälligen Schwankungen relativiert wird. Zur Berechnung des GRI sind drei Messzeitpunkte notwendig, wobei eine davon eine Baselinemessung ( $t_0$ ) sein muss. Der SES/ES ist die am häufigsten verwendete Methode zur Einschätzung der Größe eines Effektes. Insgesamt hat sich für die Bewertung von Effektstärken eine Einteilung nach Cohen (1992) etabliert, die ursprünglich zur Kalkulation von Stichprobenumfängen entwickelt wurde. Derzufolge wird ein Effekt um 0.2 als klein, um 0.5 als mittel und um 0.8 als groß bezeichnet. Nach Norman und Kollegen (2003) kann angenommen werden, dass ab einer Effektstärke  $ES \geq 0.5$  davon aus-

zugehen ist, dass es sich bei dem gefundenen Effekt um einen klinisch bedeutsamen handelt. Crosby und Kollegen (2003) kommen zu einer vergleichbaren Einschätzung.

Die *quasi-indirekte* Veränderungsmessung wurde bereits als Thentest ausführlich beschrieben. Anstatt der üblichen Prä-Messung wird bei dieser Strategie der Prä-Zustand *retrospektiv* eingeschätzt, um die Differenzbildung zwischen aktuellem Zustand (Post-Messung) und früherem Zustand (retrospektive Prämessung) vor demselben Bewertungshintergrund vorzunehmen. Obwohl es Hinweise auf eine hohe Übereinstimmung der verschiedenen Methoden der Veränderungsmessung gibt (Nübling et al. 2004, Stieglitz 1990), zeigen sich in der Mehrheit der Vergleiche eher geringe bis mäßige korrelative Übereinstimmungen (Blessmann et al. 2001, Jelitte et al. 2006, Kohlmann & Raspe 1998, Fischer et al. 1999).

Response-Shift ist für die Methoden der Veränderungsmessung ein bedeutsamer Faktor, da davon auszugehen ist, dass die indirekte Veränderungsmessung response-shift-beeinflusst sein kann. Die direkte und quasi-indirekte Veränderungsmessung kontrollieren den Response-Shift dadurch, dass die Einschätzung(en) nicht vor unterschiedlichen Bewertungshintergründen stattfinden. Korrelative Studien zur Validität der direkten Veränderungsmessung stellen diese jedoch eher in Frage. Bei der direkten VM wird eher von einer Überschätzung der Effekte ausgegangen (Lam & Bengo 2003). In der Vergangenheit wurden viele Studien publiziert, die einen Vergleich zwischen dem Ansatz der indirekten Erfassung von Lebensqualitätsveränderungen mit der quasi-indirekten Erfassung durchführten. Unabhängig von der Grunderkrankung oder der Behandlungsphase zeigt sich dabei häufig der Befund, dass die retrospektive Messung größere Effekte produziert als die herkömmliche Art der Prospektivmessung (bspw. Adang et al. 1998, Bernhard et al. 1999, Jelitte et al. 2006, Joore et al. 2002, Kohlmann & Raspe 1998, Nübling et al. 2004, Rees et al. 2003). Dennoch sagt diese einseitige Befundlage nichts darüber aus, welche Erfassungsstrategie Veränderungen genauer erfasst. Es existiert jedoch auch eine nennenswerte Anzahl an Studien, die eine gute Übereinstimmung zwischen den Ergebnissen der quasi-indirekten Veränderungsmessung und objektiven / klinischen Kriterien nahe legen (Bray et al. 1984, Hoogstraten 1982, Hoogstraten 1985, Howard et al. 1979, Nieuwkerk et al. 2007, Pohl 1982, Sprangers & Hoogstraten 1989).

## Änderungssensitivität

Nach Guyatt und Kollegen (1992) ist ein wichtiger Aspekt im Rahmen evaluativer Studien, dass die Messinstrumente in der Lage sind, über Differenzwerte intrapersonale wahre Veränderungen abzubilden und durch Veränderungen, die nicht auf das Zielkonstrukt zurückzuführen sind, möglichst nicht beeinflusst sein sollten. Allgemein ausgedrückt, sprechen die Autoren von einem Verhältnis zwischen Signal und Rauschen, wobei die ‚wahre Veränderung‘ als zu entdeckendes Signal und das Rauschen als Anteil der Fehlervarianz verstanden wird. Deshalb wird neben den klassischen Gütekriterien der Objektivität, Reliabilität und Validität im Zusammenhang mit Veränderungsmessungen die Überprüfung der *Änderungssensitivität* eines Instruments als zusätzliches Gütekriterium diskutiert (beispielsweise Epstein 2000, Igl et al. 2005, Schuck 2000). Änderungssensitivität bezieht sich auf die Fähigkeit eines Instruments, eine Veränderung im zu messenden Konstrukt auch tatsächlich abzubilden. Im angloamerikanischen Sprachraum wird zwischen den Konstrukten *Responsiveness* und *Sensitivity to change* unterschieden, wobei mit Ersterem die Fähigkeit zur Erfassung klinisch bedeutsamer Veränderungen gemeint ist, mit dem zweiten Begriff die Kalibrierung der Skala in möglichst gut differenzierende Abschnitte (Testa 2000).

Ausgangspunkt der Diskussion zur Änderungssensitivität stellt die Unterscheidung von drei verschiedenen Funktionen von Messinstrumenten nach Kirshner und Guyatt (1985, sowie Guyatt et al. 1992) dar. Besitzen Messinstrumente eine *diskriminative Funktion*, werden diese beispielsweise zur Zuweisung von Personen in bestimmte Gruppen eingesetzt (behandlungsbedürftig vs. nicht behandlungsbedürftig, epidemiologische Studien zur Prävalenz und Inzidenz von Erkrankungen), ist es ausreichend, dass die Instrumente den klassischen Gütekriterien genügen. Dasselbe gilt für die *Prädiktion* anhand der Ergebnisse eines Messinstruments. Dabei zielen Instrumente darauf ab, Vorhersagen über den zukünftigen Verlauf bestimmter Variablen zu machen (Prognoseeinschätzungen von Erkrankungen, Mortalitätsberechnungen). Sollen Instrumente jedoch zu *evaluativen Zwecken* eingesetzt werden, dann sollten diese in der Lage sein, Veränderungen in der zu messenden Zielvariable genau zu erfassen. Die Fähigkeit dazu wird als *Änderungssensitivität* bezeichnet. Strittig ist noch immer, ob es sich bei der Änderungssensitivität um ein eigenständiges Gütekriterium handelt oder inhaltlich der Validität, konkret der longitudinalen Validität, zugeordnet werden kann (Erickson 2000, Hays & Hadorn 1992, Liang 2000, Schuck 2000).

Terwee und Kollegen (2003) leiten aus einer Literaturübersicht zur Operationalisierung und Messung von Änderungssensitivität drei unterschiedlich weite Definitionen des Konstrukts ab. Die erste, sehr allgemeine Definition fasst unter Änderungssensitivität die Fähigkeit zur Erfassung von *Veränderungen im Allgemeinen* zusammen. Dazu zählen die Autoren die Fähigkeit, Veränderungen über die Zeit, klinische Veränderungen oder allgemeine Behandlungseffekte (beispielsweise Streiner & Norman 1989) zu erfassen. Von dieser Auslegung des Begriffs lässt sich eine Gruppe von Definitionen abgrenzen, die unter der Änderungssensitivität die Fähigkeit eines Instrumentes verstehen, eine *klinisch bedeutsame Veränderung* zu erfassen. Hierbei geht es darum, dass die erfassten Veränderungen für die Person von (bewusst) wahrnehmbarer Wichtigkeit oder für den weiteren Behandlungsverlauf von Bedeutung sind (beispielsweise Jaeschke et al. 1989). Die dritte Definition von Änderungssensitivität lässt sich zusammenfassen als Fähigkeit eines Instruments, eine Veränderung des zugrunde liegenden *true scores* abzubilden (beispielsweise Garratt et al. 1996). Die verschiedenen Definitionen von Änderungssensitivität erfordern unterschiedliche Erfassungsstrategien von Veränderungen. Während Studien der erstgenannten Gruppe unter anderem statistische Signifikanzmaße anhand von allgemeinen t-Tests beziehungsweise varianzanalytischen Verfahren oder einfache Effektstärkeparameter (ES) verwenden, zieht die zweite Gruppe spezifischere Maße heran (Number Needed to Treat (NNT), selektive Effektstärkeparameter (Standardized Response Mean (SRM), Gyatt's Responsiveness statistic (GRS)). Die dritte Gruppe bezieht ein weiteres Außenkriterium in die Kalkulation ein. Statistische Ansätze sind Berechnungen der Sensitivität/Spezifität, ROC-Kurven oder korrelative Maße.

Die Darstellung zum Konstrukt Änderungssensitivität ist für die vorliegende Arbeit insofern von Bedeutung, da mittels des statistischen Verfahrens der Konfirmatorischen Faktorenanalyse (KFA) berechenbar ist, welcher Anteil der Mittelwertsdifferenzen auf den einzelnen Skalen zwischen zwei Messzeitpunkten auf eine Veränderung im latenten Konstrukt ‚GLQ‘ zurückgeführt werden kann. Oort (2005a) stellt diesen Ansatz zur Effektstärkeberechnung von Response-Shift-Parametern vor, der am ehesten der Definition der dritten Gruppe der Änderungssensitivität zugeordnet werden kann. Die Differenz der beobachteten Skalenmittelwerte  $\mu_1$  und  $\mu_2$  lässt sich zerlegen in verschiedene mathematische Terme, die den Einfluss der Rekalibrierung und der Repriorisierung kalkulierbar machen.

Zur Bestimmung der Effektstärke dient üblicherweise die Differenz der beobachteten Mittelwerte relativiert an der Streuung der Werte in einer der beiden Stichproben, einer aus

beiden Stichproben gemittelten Streuung, einer gewichteten Streuung oder der Streuung der Prä-Post-Differenzen (Igl et al. 2006). In der vorliegenden Arbeit wird eine Berechnung der Effektstärke nach Oort (2005a) vorgenommen. Mit Hilfe dieser Formel lassen sich die berechneten Mittelwertsveränderungen zerlegen in einen Anteil einer ‚wahren Veränderung‘ sowie den einzelnen Response-Shift-Einflüssen. Als Streuungsmaß wird die Summe aus den Varianzen der beiden zu vergleichenden Variablen (Prä-Post-Messungen) abzüglich der zweifachen Kovarianz dieser beiden Variablen berücksichtigt.

Die Differenz der Mittelwerte im Zähler wird beschrieben anhand der Formeln zur Bestimmung des Skalenmittelwerts aufgrund des Achsenschnittpunkts und dem Einfluss des latenten Faktors entsprechend einer Regressionsgleichung.

$$\mu_1 = \tau_1 + \Gamma_1\alpha_1 \quad (10)$$

ebenso gilt

$$\mu_2 = \tau_2 + \Gamma_2\alpha_2 \quad (11)$$

$\mu$  = Mittelwert der manifesten Variable

$\tau$  = Intercept

$\Gamma$  = Kovarianz

$\alpha$  = Mittelwert des latenten Faktors

$_1$  = Messzeitpunkt 1

$_2$  = Messzeitpunkt 2

Werden die Mittelwerte durch die beiden Gleichungen ersetzt, beinhaltet der Zähler die entsprechenden Terme.

Es gilt

$$\mu_2 - \mu_1 = (\tau_2 + \Gamma_2\alpha_2) - (\tau_1 + \Gamma_1\alpha_1) \quad (12)$$

Da aufgrund der Identifikation des Messmodells der Mittelwert des latenten Faktors zum Messzeitpunkt 1 auf Null gesetzt wird, verkürzt sich die entsprechende Formel wie folgt:

$$\mu_2 - \mu_1 = (\tau_2 + \Gamma_2\alpha_2) - \tau_1 \quad (13)$$

Oder umgestellt:

$$\mu_2 - \mu_1 = \tau_2 - \tau_1 + \Gamma_2\alpha_2 \quad (14)$$

Die drei verschiedenen Response-Shift-Prozesse können berechnet werden, wenn der Term  $\Gamma_2\alpha_2$ , der den Anteil der Mittelwertsdifferenz zwischen den beobachteten Mittelwerten darstellt, der auf das latente Konstrukt gewichtet an der Faktorladung zum MZP 2 zurückzuführen ist, weiter differenziert wird. Um den Anteil der Repriorisierung beziehungsweise Neukonzeptualisierung sowie den der wahren Veränderung zu berücksichtigen, wird der Term  $\Gamma_2\alpha_2$  in entsprechend zwei Anteile zerlegt, wobei für die Faktorladung  $\Gamma_2$  gilt:

$$\Gamma_2 = (\Gamma_2 - \Gamma_1) + \Gamma_1 \quad (15)$$

Ausgehend von der Zerlegung der Faktorladung zum MZP 2 ergibt sich für die beobachtete Mittelwertsdifferenz die Formel:

$$\mu_2 - \mu_1 = (\tau_2 - \tau_1) + (\Gamma_2 - \Gamma_1)\alpha_2 + \Gamma_1\alpha_2 \quad (16)$$

Die in Gleichung 11 enthaltenen mathematischen Terme lassen sich den drei Response-Shift-Prozessen und der ‚wahren Veränderung‘ zuordnen. Der Anteil der Rekalibrierung an der beobachteten Mittelwertsdifferenz resultiert aus der Differenz der Intercepts:

$$\text{Rekalibrierung} = (\tau_2 - \tau_1) \quad (17)$$

Die Repriorisierung beziehungsweise Neukonzeptualisierung ergibt sich aus dem Quotienten:

$$\text{Repriorisierung bzw. Neukonzeptualisierung} = (\Gamma_2 - \Gamma_1)\alpha_2 \quad (18)$$

Der Anteil der beobachteten Mittelwertsdifferenz, der auf eine ‚wahre Veränderung‘ des latenten Konstrukts zurückzuführen ist, läßt sich anhand des folgenden Quotienten berechnen:

$$\text{Wahre Veränderung} = \Gamma_1\alpha_2 \quad (19)$$

Die Effekte lassen sich wie folgt differenzieren: Die drei Einzelquotienten für die Rekalibrierung, Repriorisierung und Neukonzeptualisierung sowie der wahren Veränderung werden an der Summe der Varianz der Werte zum MZP 1 und der zum MZP 2 sowie an der zweifachen Kovarianz der Werte relativiert. Es ergeben sich drei Formeln zur Kalkulation der Größe der Effekte:

$$\text{Rekalibrierung} = \frac{(\tau_2 - \tau_1)}{\sqrt{\sigma_{j1j1} + \sigma_{j2j2} - 2\sigma_{j2j1}}} \quad (20)$$

Der Quotient zur Berechnung des Effektes der Rekalibrierung ist nur dann sinnvoll lösbar, wenn die Intercepts einer Skala ungleich sind. Findet keine Rekalibrierung auf der entsprechenden Skala statt, erhält der Quotient den Wert Null.

$$\begin{array}{l} \text{Repriorisierung bzw.} \\ \text{Neukonzeptualisierung} \end{array} = \frac{(\Gamma_2 - \Gamma_1)\alpha_2}{\sqrt{\sigma_{j1j1} + \sigma_{j2j2} - 2\sigma_{j2j1}}} \quad (21)$$

Eine Repriorisierung oder Neukonzeptualisierung ist nur dann berechenbar, wenn die Faktorladungen der Skala zu beiden MZPen unterschiedlich sind.

Der Effekt durch die Veränderung des latenten Konstrukts resultiert entsprechend aus dem Anteil der wahren Veränderung relativiert an den genannten Varianzverhältnissen.

$$\text{wahre Veränderung} = \frac{\Gamma_1\alpha_2}{\sqrt{\sigma_{j1j1} + \sigma_{j2j2} - 2\sigma_{j2j1}}} \quad (22)$$

Findet weder eine Rekalibrierung noch eine Repriorisierung beziehungsweise Neukonzeptualisierung auf der zugrundeliegenden Skala statt, entspricht der Effekt für die wahre Veränderung dem Wert der beobachteten Mittelwertsdifferenz.

### 3.2.5 Bewertung der Methoden zur Response-Shift-Erfassung

Im Folgenden sollen einige Kritikpunkte an den im vorangegangenen Abschnitt vorgestellten Methoden der Response-Shift-Erfassung benannt werden. Besonderes Gewicht wird



dabei auf das am häufigsten verwandte Verfahren, den Thentest, sowie auf den in dieser Arbeit angewandten KFA-Ansatz gelegt.

Die individuellen Ansätze haben den großen Vorteil, dass sie es den befragten Personen ermöglichen, einerseits eine persönliche Bestimmung gesundheitsbezogener Lebensqualität vorzunehmen und andererseits dadurch eine auf die einzelne Person gerichtete Response-Shift-Erfassung zu ermöglichen. Zur einzelfallorientierten Arbeit sind diese Verfahren demnach gut geeignet. Verschiedene Aspekte sind jedoch kritisch anzumerken (Schwartz & Sprangers 2000). So ist es mit den individuellen Ansätzen nicht möglich, alle drei Komponenten (Rekalibrierung, Repriorisierung und Neukonzeptualisierung) differenziert zu erfassen. Eine Ausnahme bildet hier der qualitative Ansatz nach Sprangers und Schwartz, wobei es sich bei der vorgestellten Strategie bereits um eine umfangreiche Methodenkombination handelt. Ohne eine Kombination mit wenigstens einem anderen Verfahren (zumeist dem Thentest) bleibt die Gesamterfassung eines Response-Shift unvollständig. Daraus resultiert ein weiterer Kritikpunkt, die zeitaufwändige Durchführung und Auswertung. Ohne softwaregestützte Auswertung sind die individuellen Ansätze für größere Stichproben nicht geeignet. Das gilt für die Durchführung der Verfahren an sich und in noch größerem Maße für die Kombination mit anderen, ergänzenden Strategien. Außerdem ist bei diesem Ansatz ebenfalls nicht auszuschließen, dass Veränderungen durch beispielsweise Erinnerungsverzerrungen oder implizite Veränderungstheorien beeinflusst werden. Diese Einflussfaktoren werden im Zusammenhang mit der Kritik am Thentest noch genauer beschrieben (siehe unten).

Das am häufigsten angewandte Verfahren zur Response-Shift-Erfassung ist der Thentest. Seine Vorteile liegen in der Ökonomie der Durchführung und Auswertung. Er kann im Einzelfall ebenso gut verwendet werden wie in der Analyse von Gruppen. Nachteilig an diesem Vorgehen ist die unvollständige Erfassung des Response-Shifts, da hiermit lediglich eine Rekalibrierung aufgedeckt werden kann. Mittlerweile existiert jedoch auch eine umfangreiche inhaltliche Diskussion zum Thentest-Ansatz, die im Folgenden skizziert wird.

Insbesondere beim Thentest sind verschiedene andere Einflussfaktoren außer einem Response-Shift auf die relevanten Differenzwerte denkbar. Kritisch ist anzumerken, dass Retrospektivmessungen einem Recall Bias unterliegen (Ahmed et al. 2004). Die Aufforderung, einen relativ weit in der Vergangenheit liegenden Zustand genau zu erinnern, ist möglicherweise durch Erinnerungsschwierigkeiten verzerrt. Je ungenauer die Erinnerung, desto invali-

der ist die rückblickende Beurteilung. Gegen das Argument des Recall Bias spricht jedoch, dass bei dem Auftreten reiner Erinnerungsverzerrungen nicht mit einem gerichteten Trend im Sinne eines Response-Shift zu rechnen ist (Jansen et al. 2001). Zu erwarten ist bei einer reinen Erinnerungsverzerrung eine allgemeine Streuung der geschätzten Retrospektivwerte um den prospektiv eingeschätzten Prä-Wert (i. S. eines klassischen Messfehlereinflusses (Rost 2004)). Howard und Dailey (1979) konnten bereits in einer der ersten Studien zum Response-Shift, in der ein Erinnerungsbias kontrolliert wurde, feststellen, dass Teilnehmer einer Schulungsveranstaltung zur Verbesserung ihrer Kommunikation nach einem Semester ihren Ausgangswert gut erinnerten und im retrospektiven Prätest dennoch einen davon abweichenden Wert angaben.

Ein weiteres methodisches Argument, das die Strategie des Thentest in Frage stellt, ist die Möglichkeit der Befragten, sozial erwünscht zu antworten (Griner-Hill & Betz 2005). Aus der speziellen Situation heraus, dass die Befragten dasselbe Instrument zweimal hintereinander mit unterschiedlichen Instruktionen vorgelegt bekommen, ist der Zweck der Strategie – Veränderungen zu erfassen – für den Befragten auch bewusst beeinflussbar. Während die Transparenz der Strategie auch bei der prospektiven Erfassung der Kriterien gegeben ist, bleibt die Einflussnahme durch die befragte Person durch sozial erwünschtes Verhalten begrenzter. Griner-Hill und Betz (2005) schlagen deshalb vor, zur Einschätzung der Effekte von Interventionen auf den herkömmlichen Prä-Post-Test zurückzugreifen. Steht die Analyse subjektiver Veränderungen der Interventionsteilnehmer im Vordergrund, ist die Thentest-Strategie anwendbar.

Außerdem wird in der – eher unbewussten – Neigung zur Reduktion kognitiver Dissonanz ein nicht unwesentlicher Einflussfaktor auf die Thentest-Methodik gesehen. Befragte, die in der Vergangenheit an einer Maßnahme teilgenommen haben (beispielsweise einer Rehabilitation) und in diesem Zusammenhang einen Thentest zur Erfassung der GLQ durchführen, werden möglicherweise eher dazu neigen, eine Veränderung zu produzieren, auch wenn diese in Wirklichkeit gar nicht vorhanden ist. Dieser Prozess wird in der Response-Shift-Diskussion unter der Bezeichnung *Implizite Veränderungstheorien* (Norman 2003) diskutiert. Der Grund für ein solches Verhalten ist in der persönlichen Rechtfertigung zu sehen, eine Maßnahme für sich als wirksam zu erklären, da sonst die Teilnahme sinnlos gewesen wäre (i.S. einer *justification of effort*, Cooper & Axson 1982). Die individuelle Rechtfertigungsprozedur beeinflusst die Retrospektivmessung wiederum stärker als die Prospektivmessung.

Von Bedeutung in diesem Zusammenhang sind auch Erwartungen kranker Personen an Interventionen, an denen sie teilnehmen. So konnten Faller und Kollegen (2000) feststellen, dass Patienten in der onkologischen Rehabilitation höhere Prozess- und Ergebniserwartungen an die Maßnahme stellten, als Patienten in der orthopädischen Rehabilitation. In Kombination mit den Grundlagen der impliziten Veränderungstheorie ist nicht auszuschließen, dass die befragten Personen in einem Thentest Einschätzungen beziehungsweise Selbstbewertungen vornehmen würden, die einen Effekt produzieren, der in dieser Form nicht vorhanden wäre.

Mittlerweile existieren einige Studien zur Überprüfung des Response-Shifts anhand von Veränderungen objektiver Kriterien. Nieuwkerk und Kollegen (2007) kamen beim Vergleich der Veränderungsmessungen mittels Retrospektiv- und Prospektivmessung bei HIV-Patienten zu dem Ergebnis, dass die über die retrospektive Prämessung erfassten Veränderungen in der GLQ engere Korrelationen mit selbsteingeschätzten klinischen Parametern (BMI) aufzeigen als mit objektiven Gesundheitsparametern.

Hinsichtlich des KFA-Ansatzes steht den großen Vorteilen der differenzierten Response-Shift-Erfassung und der ökonomischen Erfassung der Lebensqualitätsdaten zu lediglich zwei Messzeitpunkten ohne Retrospektivmessung die Komplexität der statistischen Auswertung entgegen. Diese ist durch Fortschritte in der Softwareentwicklung (v.a. AMOS, Arbuckle 2005) deutlich vereinfacht worden. Ein Aspekt, der kritisch anzumerken ist, bezieht sich auf die Abhängigkeit des KFA-Ansatzes von großen Stichproben. Einzelfallanalysen sind mit diesem Ansatz nicht möglich. Mit der Anwendung der KFA-Strategie zur Response-Shift-Erfassung begann eine Diskussion über die Vor- und Nachteile der Methode. Eine Kritik äußert Donaldson (2005), der die Annahme eines Response-Shift lediglich als eine mögliche Erklärung für eine Invarianz der Parameter zwischen zwei Messzeitpunkten ansieht. Die exakte Operationalisierung dieses Prozesses sei nicht gelungen. Eine Veränderung der Faktorladung beispielsweise ist nicht zwangsläufig ein Response-Shift, sondern möglicherweise auch ein Interventionseffekt, der sich auf einer Variablen (Skala) abzeichnet, die dadurch ihr relatives Gewicht zu anderen Skalen verändert. Weiterhin betont Donaldson die Bedeutung der theoretischen Modellannahmen mit einem Verweis auf die Unterscheidung von Variablen in Indikator- und Kausalvariablen der GLQ. Eine von den Modellannahmen nach Oort und Ahmed abweichende Modelldefinition – beispielsweise dass die Physische Lebensqualität nur einen tatsächlichen Indikator besitzt, nämlich die Physische Funktionsfähigkeit, während die anderen Variablen eher die latente Variable kausal beeinflussen (und nicht wie angenommen, von

ihr beeinflusst werden) – führt im Verlauf der Modellspezifikationen zu gänzlich abweichenden Aussagen zu Response-Shift-Parameterschätzungen. Weiterhin wesentlich ist die unterschiedliche Vorgehensweise der Invarianztestung bei den Autoren Oort und Ahmed, da Ahmed nicht die Invarianztestung der Faktorladungen vorzieht, sondern die der Äquivalenz der Faktorvarianzen. Im Sinne einer stringenten Invarianztestung ist jedoch mit der Testung der Faktorladungen zu beginnen. Außerdem berücksichtigt Ahmed nicht die Invarianztestung der Intercepts. Abschließend stellt Donaldson die Vernachlässigung der Invarianztestung der Intercepts beim Vorgehen nach Ahmed in Frage, da diese zur Abschätzung der wahren Veränderung in den Mittelwerten der latenten Variablen vorauszusetzen sind.

Weitere Fragen, die sich aus den KFA-Ansätzen nach Oort und Ahmed ergeben, beziehen sich auf die Güte der Modelle, die als Modell 1 (Ahmed) beziehungsweise Nullmodell (Oort) als Ausgangsmodelle dienen. Bei Ahmed zeigt sich eine grenzwertig akzeptable Passung des Messmodells mit der Kovarianzstruktur der Daten (RMSEA = .08, CFI = .94, die genannten Gütemaße für SEM werden in Kapitel 6.3.3 ausführlich beschrieben). Borsboom und Kollegen (2007) sehen in dieser kritischen Passung eine mögliche Ursache für das Ergebnis von Ahmed, dass der KFA-Ansatz keinen Response-Shift aufgedeckt hat. Bei der Verwendung von Strukturgleichungsmodellen, und damit auch von Messmodellen im Rahmen der Konfirmatorischen Faktorenanalyse, ist immer davon auszugehen, dass andere theoretische Modelle existieren, die besser zu den empirischen Daten passen als das bei der jeweiligen Stichprobe gefundene. Grundlegende Modellspezifikationen, die auf theoretischer Basis basieren, bilden die Voraussetzung für das Ausgangsmodell zur Response-Shift-Erfassung. Unzureichende Modellspezifikationen können jedoch zu Verzerrungen in der Kovarianzmatrix führen und so eine Aufdeckung eines Response-Shifts verhindern. Diese Annahme ist umso begründeter, je weniger gut die Passung des Ausgangsmodells ist.

### **3.3 Empirische Befunde zum Response-Shift**

Nachdem im vorangegangenen Kapitel die Methoden zur Erfassung von Response-Shift dargestellt wurden, folgt ein Überblick über empirische Befunde zum Phänomen. In einem ersten Abschnitt werden Studien vorgestellt, die auf den Vergleich verschiedener Response-Shift-Erfassungsmethoden ausgerichtet waren. Ergebnisse der einzelnen Strategien werden ebenfalls berichtet. In einem weiteren Abschnitt werden Studien zur Untersuchung an homogenen und heterogenen Stichproben mit Krebspatienten präsentiert. Vorab werden allgemeine

Angaben zur Response-Shift-Untersuchung dargestellt. Dazu gehört auch die Vorstellung der bisher einzigen Metaanalyse zum Phänomen.

Bisher liegt eine Vielzahl an Einzelbefunden vor, die Response-Shift-Einflüsse bei verschiedenen Grunderkrankungen untersuchen. Einige wurden bereits im vorangegangenen Kapitel zur Bewertung der methodischen Ansätze dargestellt. In den vergangenen Jahren stieg die Anzahl der Publikationen zum Thema deutlich. Mittlerweile existieren empirische Untersuchungen zum Phänomen bei unterschiedlichen Grundkrankheiten und durchgeführten Behandlungen. Dies gilt beispielsweise für das Schlafapnoe-Syndrom (Chin et al. 2004), Schlaganfall (Ahmed et al. 2004, 2005c), Hörschädigungen (Joore et al. 2002, Yardley & Dibb 2007), Knie-Arthroplastie (Razmjou et al. 2006, Razmjou et al. 2009), Zahnimplantate (Ring et al. 2005), HIV-Erkrankungen (Nieuwkerk et al. 2007, Rapkin 2000), Bluthochdruck (Bar-on et al. 2000) und Multiple Sklerose (Schwartz et al. 2004). Echteld und Kollegen (2005, 2007) veröffentlichten eine Studie zur Response-Shift-Erfassung bei Patienten auf einer Palliativstation. Des Weiteren existieren Arbeiten zu Bauchspeicheldrüsentransplantation (Adang et al. 1998). Im pädiatrischen Bereich sind Publikationen mit empirischen Untersuchungen zum Thema erschienen, unter anderem zur chronischen Mittelohrentzündung (Timmerman et al. 2003), Juvenilen Rheumatischen Arthritis (Brossart et al. 2002) und Diabetes mellitus (Wagner 2005).

Schwartz und Kollegen (2006) legten eine Metaanalyse mit 19 Studien zum Response-Shift im Rahmen von Lebensqualitätserfassungen vor. Es wurden fünf Lebensqualitätskomponenten untersucht: Allgemeine Lebensqualität, Fatigue, psychische Befindlichkeit, Schmerzen und Physische Rollenfunktion. Unterschiedliche Moderatorvariablen wurden auf ihre Bedeutung für Response-Shift-Prozesse untersucht. Unter anderem waren dies die angegebenen Stichprobenumfänge, Studiendesign, Grunderkrankung und die Response-Shift-Erfassungsmethode. Es wurden kleine bis Null-Effekte für Response-Shift in den untersuchten Lebensqualitätskomponenten berechnet ( $ES = 0.31$  für Fatigue bis  $0.07$  für Schmerzen). Die Berechnung ungewichteter Effektstärken zeigt das Ergebnis, dass in zwei Domänen (Fatigue und Allgemeine Lebensqualität) kleine Effekte, in den weiteren drei Bereichen Nulleffekte bestanden. Die ‚Response-Shift-Erfassungsmethode‘ war die einzige Moderatorvariable, für die statistische Bedeutsamkeit festgestellt wurde. Jedoch basiert dieses Ergebnis lediglich auf einer Substichprobe von sieben Studien, wobei sechs davon den Thentest als Methode durchführten. Die Aussagekraft dieser Variablen als Moderator ist daher sehr kritisch zu bewerten.

Weiterhin ist kritisch anzumerken, dass in den aufgenommenen Studien zum Teil nicht erkennbar war, ob es sich bei dem gefundenen Response-Shift um eine Über- oder Unterschätzung der wahren Veränderung handelte. Dies wird als Grund für die geringen Effektstärken bei Verwendung der ungewichteten Studienparameter gewertet. Weiterhin wurden aus den metaanalytischen Berechnungen solche Studien ausgeschlossen, die mit dem methodischen Ansatz der konfirmatorischen Faktorenanalyse arbeiteten. Mögliche Ursachen dafür können in den fehlenden Angaben zur Berechnung der Effektstärken in den publizierten Arbeiten beziehungsweise in der mangelnden Vergleichbarkeit der publizierten Effektstärkeparameter gesehen werden. Studien mit KFA-Ansatz zur Berechnung von Response-Shift wurden nur beschreibend in die Darstellung der Ergebnisse der Metaanalyse aufgenommen.

### 3.3.1 Studien zum Vergleich methodischer Ansätze zur Erfassung von Response-Shift

Es existiert eine breite Palette an Methoden zur Response-Shift-Erfassung. Diese unterscheiden sich z. T. darin, welche Response-Shift-Prozesse diese aufdecken. Von daher stellt sich die Frage nach der konvergenten Validität der verschiedenen Strategien, die auf die Erfassung von Response-Shift ausgerichtet sind.

Nach einer Studie von Visser und Kollegen (2005) zeigt sich eine hohe Übereinstimmung zwischen Thentest und dem Ansatz der Konfirmatorischen Faktorenanalyse (KFA) nach Oort (2005a). In acht von neun Skalen des SF-36 identifizieren beide Strategien einen Response-Shift. Dieses Ergebnis wird als Argument gegen den Recall-Bias als Verzerrungstendenz in der Thentest-Strategie gewertet. Kritisch ist anzumerken, dass die Strategien per definitionem nicht die gleichen Response-Shift-Prozesse erfassen. Der Thentest erfasst eine Rekalibrierung und der KFA-Ansatz alle drei Prozesse. Nach den Ergebnissen der Studie von Oort (2005b) findet auf der Skala Soziale Funktionsfähigkeit eine Repriorisierung statt, die im Thentest nicht erfasst werden kann. Auf der Skala Allgemeine Gesundheitswahrnehmung decken beide Verfahren einen Response-Shift auf, wobei der Trend jedoch gegenläufig ist. Diese Diskrepanz wird von den Autoren auf die mittels KFA-Methode aufgedeckte Rekonzeptualisierung der GLQ zurückgeführt. Ein zusätzlich durchgeführter Ansatz zur individuellen Erfassung von Ankerpunkten der Lebensqualität zeigte vom Thentest und KFA-Ansatz deutlich abweichende Befunde.

Die Ergebnisse von Visser und Kollegen (2005) stützen Aussagen einer Arbeit von Schmitt und Kollegen (1984), die ebenfalls eine hohe Übereinstimmung zwischen Thentest-

und SEM-Ansatz und eine mangelnde Übereinstimmung zwischen diesen beiden Methoden und dem Ansatz der Idealskalen postuliert. Auch wenn der Ansatz mittels KFA nach Schmitt in mancher Hinsicht nicht mit dem KFA-Ansatz von Oort gleichzusetzen ist, weisen beide Studien auf eine gewisse Konvergenz zwischen Thentest und Kovarianz-Struktur-Ansätzen hin.

Gegensätzlich zu den berichteten Konvergenzen zwischen Thentest und KFA- beziehungsweise SEM-Methode berichten Ahmed und Kollegen (2005c) eine deutliche Divergenz zwischen diesen beiden Methoden sowie eine hohe Übereinstimmung zwischen dem Thentest und einem individuellen Ansatz, dem Patient Generated Index nach Ruta und Kollegen (1994). Die Untersuchung wurde an Schlaganfallpatienten durchgeführt, wobei eine Kontrollgruppe in das Studiendesign integriert wurde. Diese zeigte über einen Zeitverlauf von sechs Monaten keinen Response-Shift, wohingegen mit dem Thentest bei ca. 50% sowie mit dem PGI bei ca. 80% der befragten Schlaganfallpatienten ein Response-Shift nachgewiesen wurde. Beide Ansätze erfassen verschiedene Response-Shift-Prozesse, so dass die Konvergenz lediglich für die Anwesenheit oder Abwesenheit des Phänomens, nicht aber für die genauere Differenzierung gilt. Ahmed und Kollegen (2005a) konnten nachweisen, dass die Güte der Erinnerung an den Prämesszeitpunkt bedeutsam ist für die Erfassung des Response-Shift mit dem Thentest. Mit dem KFA-Ansatz konnte kein Response-Shift nachgewiesen werden. In einer weiteren Auswertung zum Zusammenhang von Response-Shift-Prozessen und Veränderungen in klinischen Parametern konnten Ahmed und Kollegen – im Gegensatz zu Nieuwkirk und Kollegen (2007) – keine statistisch bedeutsame Korrelation zwischen beiden feststellen.

Eine Vergleichsstudie nach Ahmed und Kollegen (2009) zur Prüfung der Vergleichbarkeit zwischen der KFA-Methode und dem SEM-Ansatz nach Schmitt erbrachte deutliche Unterschiede zwischen diesen Strategien. Mittels KFA-Ansatz nach Oort konnten verschiedene Response-Shift-Prozesse identifiziert werden. Die Überprüfung nach dem Ansatz von Schmitt führte zum gegenteiligen Ergebnis. Auf keiner der überprüften Ebenen waren Parameter freizusetzen. Der Grund dafür wird von den Autoren einerseits in dem unterschiedlichen Vorgehen gesehen. Der Oort-Ansatz berücksichtigt Varianzen, Kovarianzen und Intercepts beziehungsweise Mittelwerte. Der Schmitt-Ansatz bezieht sich ausschließlich auf die Varianzen und Kovarianzen. Andererseits ist nach dem Vorgehen nach Schmitt auch die Reihenfolge der Ebenen der Parametertestungen definiert, die sich an dem allgemeinen Vorgehen der Invarianztestung orientiert.

Nolte und Kollegen (2009) untersuchten den Thentest mit Hilfe des KFA-Ansatzes nach Ort. Die Autoren prüften, ob die Ausprägung der Response-Shift-Parameter der Posttestmessung mit denen der Messungen anhand des retrospektiven Prätest überein stimmten. Da beide Einschätzungen zeitnah erfolgen, wurde angenommen, dass von Invarianz der Messungen ausgegangen werden kann. Outcome Kriterium der Untersuchung war nicht die GLQ, sondern Faktoren wie Selbstmanagement, Compliance und Empowerment. Im Ergebnis zeigte sich, dass aufgrund der Invarianztestung der Messmodelle der Posttestmessung und der retrospektiven Prämessung Response-Shift nachgewiesen wurde. Dieses Ergebnis weist darauf hin, dass der Thentest nicht grundsätzlich als response-shift-bereinigt angesehen werden darf.

### 3.3.2 Response-Shift bei Krebserkrankungen

Empirische Untersuchungen zum Einfluss eines Response-Shift auf die subjektive Lebensqualität von Krebspatienten liegen für verschiedene Krebserkrankungen vor. Die größere Anzahl der Studien erfasst homogene Stichproben von Krebserkrankungen, um Aussagen über das Phänomen bei spezifischen Patientengruppen abzuleiten. Andere Studien berichten Ergebnisse zum Response-Shift bei heterogenen Patientenstichproben, die sich aus Personen mit unterschiedlichen Krebserkrankungen zusammensetzen.

#### *Response-Shift bei Stichproben mit homogener Krebsdiagnose*

Bernhard und Kollegen (1999, 2001, 2004) untersuchten den Einfluss von Response-Shift auf die Einschätzung der Lebensqualität zu verschiedenen Messzeitpunkten bei 187 Darmkrebspatienten. Es wurden Retrospektiverfassungen mittels Thentest durchgeführt. Die Retrospektivmessungen unterschieden sich statistisch (hoch)signifikant von den entsprechenden Prospektivmessungen (akut und Nachsorge), so dass von einer Rekalibrierung auszugehen ist. Die Art der Therapie (und damit zusammenhängender Nebenwirkungen) beeinflusst den Response-Shift nicht.

Jansen und Kollegen (2001) untersuchten die GLQ von 46 Brustkrebspatientinnen und erfassten eine Hierarchie persönlich relevanter Werte zu Beginn und nach Abschluss einer postoperativen Strahlentherapie. Hinweise auf eine Repriorisierung wurde anhand von Veränderungen in den Wertehierarchien gefunden. Im Rahmen einer Thentest-Erfassung wurde zusätzlich ein retrospektiver Prätest mit den Instrumenten SF-36 (Ware & Sherbourne 1992) sowie der Rotterdam Symptom Checkliste (RSCL, Haes & Olschewski 1998) durchgeführt.



Die Ergebnisse der Thentest-Analyse erbrachten statistisch signifikante Response-Shift-Effekte auf der Skala Allgemeine Lebensqualität. Die Richtung eines Response-Shifts wurde durch den Trend der allgemeinen Entwicklung im untersuchten Konstrukt mit bedingt: verbesserte sich die wahrgenommene Lebensqualität zwischen beiden Messzeitpunkten, zeigte sich im Thentest eine noch deutlichere Verbesserung (positive Rekalibrierung). Verschlechterte sich die wahrgenommene Lebensqualität zwischen den beiden Messzeitpunkten, resultierte im Thentest eine noch deutlichere Verschlechterung (negative Rekalibrierung).

Untersuchungen zum Response-Shift bei Lungenkrebspatienten liegen von Westerman und Kollegen (2006, 2007) sowie von Broberger und Kollegen (2006) vor. Im Rahmen eines qualitativen Ansatzes untersuchten Westerman und Kollegen (2006) die Fatigue-Einschätzungen von Lungenkrebspatienten im Verlauf der Chemotherapie. Die Autoren berichten von Rekalibrierung und Repriorisierung bei den Befragten. Broberger und Kollegen (2006) untersuchten eine mögliche Rekalibrierung der GLQ bei Lungenkrebspatienten mit Hilfe des Thentests. Es zeigte sich insgesamt ein sehr heterogenes Bild an Response-Shift-Effekten. Eine eindeutige Aussage über das Auftreten einer Rekalibrierung auf Gruppenebene ist anhand der Befunde nicht möglich.

Zur Untersuchung von Response-Shift bei PCa-Patienten wurden verschiedene Studien durchgeführt. Lepore und Eton (2000) untersuchten Response-Shift bei PCa-Patienten anhand eines retrospektiven Prätests und einer Wertehierarchie. Repriorisierung hat den Autoren zu Folge Auswirkungen auf die Bewertung der Lebensqualität insbesondere im Falle einer Verschlechterung des Kontinenzzustands. Bei den Personen, deren Inkontinenz sich verschlechterte, kam es trotzdem zu einer positiveren Einschätzung der Lebensqualität, wenn sie ihre persönlich relevanten Werte veränderten. Weiterhin konnten die Ergebnisse zur positiven und negativen Rekalibrierung von Jansen und Kollegen (2000) von den Autoren bestätigt werden: Zeigten Männer eine Verbesserung der GLQ über zwei Messzeitpunkte, wird der Ausgangswert im Thentest ‚noch‘ schlechter dargestellt als im prospektiv erfassten Prätest. Verschlechterte sich die GLQ mit der Zeit, führte die retrospektive Beurteilung des Ausgangswerts zu einer noch größeren Differenz zwischen den Messzeitpunkten. Zusätzlich zeigte sich, dass bei Männern mit einer positiven Rekalibrierung Verschlechterungen im Kontinenzstatus weniger Einfluss auf die Bewertung der GLQ hatte als bei Männern mit negativer Rekalibrierung.

Rees und Kollegen (2003) untersuchten Response-Shift-Effekte bei einer Stichprobe mit fortgeschrittenem PCa und konnten über drei Messzeitpunkte anhand eines Thentests eine Rekalibrierung des Bewertungshintergrundes zur Einschätzung der Lebensqualität nachweisen. Zwischen den ersten beiden Messzeitpunkten konnte im herkömmlichen Prä-Post-Vergleich eine signifikante Verbesserung der GLQ festgestellt werden, die mittels der Thentest-Methodik nochmals statistisch signifikant deutlicher ausfiel. Zusätzlich wurde eine gesunde Kontrollgruppe in die Studie aufgenommen, in der keine Response-Shift-Effekte nachgewiesen wurden.

Rees und Kollegen (2005) konnten anhand der Thentest-Strategie auch bei Angehörigen von PCa-Patienten im fortgeschrittenen Stadium einen Response-Shift nachweisen. Partnerinnen der Krebspatienten gaben ein statistisch signifikant höheres Belastungsniveau durch die Erkrankung an als die Betroffenen selbst. Es zeigten beide Gruppen über einen Zeitraum von sechs Monaten einen deutlichen Response-Shift anhand des Thentest, wobei der Response-Shift bei den Partnern statistisch signifikant größer ausfiel als bei den PCa-Patienten.

Korfage und Kollegen (2007) fanden mit Hilfe des Thentests Belege, dass Männer retrospektiv nach Diagnosestellung eines lokalen PCa ihre GLQ vor der Diagnosemitteilung statistisch signifikant besser einschätzten. Zu drei Messzeitpunkten wurde die GLQ erfasst. Vor der Diagnosestellung, einen Monat sowie sieben Monate nach Diagnosestellung. Es zeigte sich ein Response-Shift Effekt in Höhe eines mittleren Effekts ( $ES = -0.43$ ) in einem Zeitraum bis sieben Monate nach Diagnosemitteilung. Der Vergleich der Prospektiverfassung zum Zeitpunkt 1 Monat nach Diagnosemitteilung mit der Retrospektiverfassung für diesen Messzeitpunkt sieben Monate nach der Diagnosemitteilung ergibt jedoch keinen Hinweis auf einen Response-Shift. Weiterhin war die Größe eines möglichen Response-Shift in dieser Studie abhängig von den Fragebögen, die verwendet wurden.

Studien zum Thema Response-Shift von Postulart und Kollegen (2000) und Tsevat und Kollegen (1995), die einen Vergleich aus Prospektivmessung zu zwei Messzeitpunkten und den Thentest zum Postmesszeitpunkt durchführen, weisen darauf hin, dass ein Response-Shift nicht nur dadurch zustande kommen kann, dass im Verlauf der Bewältigung einer kritischen Lebenssituation ein Umdenken stattfindet, sondern dass bereits der prospektiv erhobene Prä-Wert bewältigungsorientiert verzerrt sein kann. So haben Postulart und Kollegen (2000) festgestellt, dass unabhängige, nicht-betroffene Beurteiler, die Einschätzungen der GLQ von Menschen vor einer Bauchspeicheldrüsen-Transplantation abgeben sollten, die GLQ vor der

Transplantation in vergleichbarer Weise beurteilen wie tatsächlich Betroffene dieser Transplantation in ihren Retrospektivmessungen nach 5, 12 und 18 Monaten Posttransplantation. Der prospektiv erhobene Prä-Wert der an Bauchspeicheldrüsenkrebs Erkrankten unterschied sich jedoch statistisch signifikant im Sinne einer positiven Verzerrung von den Beurteilungen der gesunden Vergleichsgruppe.

### *Response-Shift bei Stichproben mit heterogenen Krebsdiagnosen*

Eine Studie zum Response-Shift bei einer heterogenen Stichprobe mit Krebspatienten (Lungen- und Colorektalkrebs) wurde von Sharpe und Kollegen (2005) durchgeführt. Sie setzten zur Response-Shift-Erfassung die SEIQoL-Methode ein. Response-Shift hatte einen Einfluss auf die Einschätzung der GLQ. Bei Personen, die zu beiden Messzeitpunkten verschiedene Lebensbereiche als am bedeutsamsten für die eigene Lebensqualität beurteilten (bei geringen Ausgangswerten), resultierte eine Verbesserung der Lebensqualität im Bereich eines mittleren Effektes nach Cohen. Im Vergleich dazu ergab sich für die Personen, die zu beiden Messzeitpunkten ein und denselben Lebensbereich an erster Stelle angaben, eine signifikante Verschlechterung der Lebensqualität im Bereich eines großen Effektes, wenn die ‚Wichtigkeit‘ des Lebensbereichs trotz Erstplatzierung zu beiden Messzeitpunkten als relativ gering eingeschätzt wurde.

Eine Arbeit von Visser und Kollegen (2000) untersuchte den Einfluss von Response-Shift auf die Beurteilung von Fatigue an einer heterogenen Stichprobe von Krebspatienten, die einer Strahlentherapie unterzogen wurden. Mit Hilfe der Thentest-Messung zeigte sich ein Response-Shift-Effekt. Die Differenz zwischen dem Then- und dem Posttest war statistisch signifikant, wohingegen die Werte der normalen Prä-Testung zum Postmesswert keinen statistisch signifikanten Unterschied aufwiesen.

Die Response-Shift-Erfassung mittels KFA von Oort (2005b) wurde bereits an anderer Stelle berichtet, soll hier aber der Vollständigkeit halber nochmals kurz vorgestellt werden. Ein Response-Shift wurde auf unterschiedlichen Ebenen und in verschiedenen Dimensionen der GLQ aufgedeckt. Zwischen den beiden Messzeitpunkten kam es zu einer Neukonzeption der GLQ. Die *Allgemeine Gesundheitswahrnehmung* (Skala des SF-36) war zum Zeitpunkt der ersten Erfassung dem Faktor *Allgemeine Fitness* zugeordnet (gemeinsam mit der SF-36 Skala *Vitalität* sowie einer *Fatigueskala*). Zum zweiten Messzeitpunkt lud diese Skala auf

dem Faktor *Allgemeine Psychische Befindlichkeit*. Eine Repriorisierung zeigte sich im Bereich der *Sozialen Funktionsfähigkeit*, deren Bedeutung für das latente Konstrukt zum zweiten Messzeitpunkt statistisch signifikant zunahm. Eine Rekalibrierung zeigte sich in den Bereichen (Skalen) *Rollenfunktion*, *Schmerzen* und *Vitalität*.

Sharpe und Kollegen (2005) konnten nachweisen, dass eine Repriorisierung/Neukonzeptualisierung der GLQ nur bei etwa 50 % einer untersuchten Stichprobe von Krebspatienten mit Metastasierungen auftrat. Response-Shift ist aus Sicht der Autoren nur dann hilfreich für die Auseinandersetzung mit der Krankheitssituation, wenn die Werte und Ziele, die Menschen vor der Diagnosestellung hatten, für die Auseinandersetzung mit der neuen Lebenssituation unpassend sind. Nur die Patienten, die vor der Diagnosestellung Werte als lebensbedeutsam nannten, die für die Krankheitsbewältigung als kritisch anzusehen waren, zeigte sich eine signifikante Verbesserung der Lebensqualität, nachdem sie neue Werte für sich definierten. Diejenigen, die von Beginn an bewältigungsrelevante Werte definierten, blieben in ihrer Definition gesundheitsbezogener Lebensqualität und auf hohem Niveau der Lebensqualität stabil.

### **3.4 Zusammenfassung**

Response-Shift ist seit fast 40 Jahren ein Thema in der Evaluationsforschung, wobei es erst in den vergangenen 10 Jahren zunehmend an Bedeutung gewonnen hat. Mit ein Grund dafür ist die Weiterentwicklung der Lebensqualitätsforschung, die Response-Shift als Einflussfaktor zur Erklärung paradoxer Befunde heranzog. Eine Integration verschiedener Response-Shift-Prozesse (Rekalibrierung nach Howard et al. 1979 sowie Gamma-Change nach Golembiewski et al. 1976) und deren konzeptuelle Weiterentwicklung bietet das Modell nach Sprangers und Schwartz (2000), in dem Response-Shift definiert wird als Veränderung interner Standards (Rekalibrierung), persönlicher Werte (Repriorisierung) und Neudefinition (Neukonzeptualisierung) eines zugrunde liegenden Konstrukts (beispielsweise der GLQ). Die Autorinnen formulieren Annahmen darüber, welche Bedingungen zum Auftreten eines Response-Shifts beitragen. Bestimmte Ereignisse können als Katalysatoren wirken (z.B. Veränderung des Gesundheitszustandes), was zu einem Response-Shift führen kann, wenn bestimmte Vorbedingungen (u.a. Persönlichkeitsfaktoren, soziodemographische Aspekte) und Mechanismen (u.a. Bewältigungsstile) in der Person miteinander interagieren.

Der differenzierten Konzeptionalisierung von Response-Shift steht eine Vielfalt an methodischen Ansätzen gegenüber, die zur Erfassung einzelner Response-Shift-Prozesse geeig-

net sind. Grundsätzlich lassen sich individuell-orientierte, Design- und statistische Ansätze unterscheiden. Am häufigsten wird der so genannte Thentest verwendet, der zu den Design-Ansätzen zählt. Dabei wird im Rahmen einer Prä-Post-Messung zum Postmesszeitpunkt eine retrospektive Beurteilung des Zustandes zum Zeitpunkt der Prämessung erhoben. Die Differenz zwischen der prospektiv und retrospektiv beurteilten Prämessung wird als Rekalibrierung bezeichnet. Ein statistisches Verfahren zur Response-Shift Erfassung wurde von Oort (2005a) vorgestellt. Mittels Konfirmatorischer Faktorenanalyse (KFA) werden Invarianztestungen bezüglich der zu verschiedenen Messzeitpunkten abgeleiteten Messmodelle überprüft. Besondere Relevanz hat die Invarianz des gesamten Messmodells, der Faktorladungen, der Achsen-schnittpunkte sowie der Messfehlervarianzen. Kritisch an dem KFA-Ansatz nach Oort ist, dass er Response-Shift nur dann abbilden kann, wenn bei einer genügend großen Anzahl von Personen diese Prozesse stattfinden. Dafür ist er in der Erhebung jedoch ökonomischer, da lediglich zwei Prospektivmessungen zu verschiedenen Messzeitpunkten benötigt werden. Die Auswertung mittels Softwareprogrammen wie AMOS (Arbuckle 2005) erleichtert die statistische Verarbeitung der Daten.

Obwohl die Anzahl empirischer Arbeiten zum Response-Shift in den vergangenen Jahren deutlich zugenommen hat, existiert bisher eine überschaubare Menge an Einzelbefunden. Häufig wurde Response-Shift bei krebserkrankten Personen untersucht. Die Resultate der Untersuchungen weisen auf die Bedeutung des Phänomens für die Evaluation von Interventionen hin. Für die Veränderungsmessung ist es von großer Bedeutung, da mittlerweile genügend Belege dafür existieren, dass herkömmliche Effektstärkeberechnungen möglicherweise keine validen Operationalisierungen von Interventionsergebnissen liefern. Der KFA-Ansatz nach Oort bietet auch dafür einen differenzierten Ansatz zur Trennung von Response-Shift Effekten und wahren Veränderungen.

## **4. Response-Shift und Veränderungsmessung von gesundheitsbezogener Lebensqualität bei Prostatakrebspatienten in der onkologischen Rehabilitation**

Im folgenden Kapitel werden die Grundlagen und Befunde zum Response-Shift im Rahmen der Veränderungsmessung gesundheitsbezogener Lebensqualität zusammengefasst, um die Ableitung der Fragestellungen vorzubereiten. Das Prozessmodell von Response-Shift und Lebensqualität nach Sprangers und Schwartz (1999) stellt den Ausgangspunkt dar für die Ableitung eines spezifischen Wirkmodells zur Beschreibung und Erfassung von Response-Shift bei PCa-Patienten im Rahmen der AHB. Dieses spezifische Wirkmodell enthält die wesentlichen Zusammenhänge und Annahmen, die der vorliegenden Arbeit zugrunde liegen. Die Moderatorvariablen Benefit Finding, Selbstwirksamkeit, positive und negative Affektivität werden in das Wirkmodell integriert. Der Darstellung der inhaltlichen Perspektive folgt die Zusammenfassung der methodischen Vorgehensweise der Response-Shift-Erfassung mittels Konfirmatorischer Faktorenanalyse.

GLQ ist ein bedeutsames Outcomekriterium in der Evaluation von Interventionen bei Krebspatienten. Es wird als latentes Konstrukt bezeichnet, das sich in Indikatorvariablen manifestiert und in erster Linie durch Selbstbeurteilungen der erkrankten Person zu erheben ist (Sonn et al. 2009). GLQ ist ein Teilbereich der allgemeinen Lebensqualität, der sich auf Aspekte des Befindens und Handelns bezieht, die mit körperlichen Einschränkungen oder (chronischen) Krankheiten zusammenhängen (Bullinger 2000). In der vorliegenden Arbeit wird ein Modell von Böhmer und Luszczynska (2006) übernommen, in dem GLQ als latentes Konstrukt bezeichnet wird, das sich aus der subjektiven Einschätzung der Komponenten Physische, Kognitive, Soziale, Emotionale und Rollenfunktionsfähigkeit erschließen lässt. Mit Hilfe des krebsspezifischen Fragebogens EORTC-QLQ-C30 (Fayers et al. 1995) werden diese Komponenten gemessen. Kritisch wird diskutiert, ob spezifische Symptombelastungen oder Behandlungsnebenwirkungen ebenfalls als Indikatoren der GLQ zu verstehen sind (Fayers & Hand 2002). In der vorliegenden Arbeit wird davon ausgegangen, dass diese Faktoren so genannte Kausalfaktoren für die GLQ sind. Diese bilden die GLQ nicht ab, sondern beeinflussen diese gerichtet und werden daher nicht in das GLQ-Modell aufgenommen.

Es existiert ein umfangreicher Bestand an Literatur zur Behandlung, der gesundheitlichen Beeinträchtigungen und der GLQ von PCa-Patienten mit Bezug zum Verlauf der Erkrankung und der Anwendung unterschiedlicher Therapiemethoden. Folgendes ist für PCa-Patienten anzunehmen: Je nach durchgeführter Akuttherapie stehen unterschiedliche Beeinträchtigungen und Belastungen im Vordergrund. Aufgrund der dargestellten Studienergebnisse ist von einer zumindest kurzfristigen Beeinträchtigung in den Bereichen der sexuellen Funktionsfähigkeit, Inkontinenz sowie Darm- und Verdauungsbeschwerden auszugehen, wobei sich die Beeinträchtigungen und Nebenwirkungen bei keiner Therapieform konsistent nachweisen lassen (Gomela et al. 2009). Es ist davon auszugehen, dass bei einer (radikalen) Prostatektomie eher Inkontinenz und sexuelle Beeinträchtigung nachweisbar sind. Bei der Strahlentherapie sind es eher Verdauungsbeschwerden. Bezüglich der Lebensqualität bieten die Befunde ebenfalls ein unklares Bild. Einerseits wird postuliert, die Lebensqualität von PCa-Patienten unterscheide sich nicht von Personen in der gesunden Allgemeinbevölkerung, jedoch sind auch diesbezüglich anderslautende Studienergebnisse veröffentlicht. Nach Litwin und Talcott (2005) lassen sich aus den Befunden zur Lebensqualität bei PCa-Patienten folgende Annahmen ableiten: Die sexuelle Funktionsfähigkeit, der Urogenitaltrakt und Darmbereich sind nach der Behandlung von Prostatakrebs im Frühstadium häufiger als bisher angenommen beeinträchtigt. Krankheits- und krebsspezifische Fragebögen sind eher sensitiv, in einer solchen Phase Veränderungen in der GLQ abzubilden, als generische Instrumente. Inter- und intraindividuelle Faktoren spielen für die Beziehung zwischen organspezifischen Dysfunktionen und aufgetretenen Belastungen und Beeinträchtigungen eine wichtige Rolle. Beeinträchtigungen variieren über die Zeit und je nach Behandlungsmaßnahme. GLQ ist ein subjektives Konstrukt, das sich im Verlauf der Erkrankung verändern kann.

Aus diesem Grund wird für die vorliegende Arbeit davon ausgegangen, dass sich ein großer Teil der PCa-Patienten zum Zeitpunkt des Beginns der AHB mit körperlichen und psychischen Beeinträchtigungen und Belastungen auseinandersetzt. Es wird weiterhin angenommen, dass durch diese Beeinträchtigungen die Bewertung der GLQ beeinflusst wird. Weiterhin wird eine mögliche Erklärung für die zum Teil widersprüchlichen Befunde hinsichtlich der Beeinträchtigung oder der Veränderung der GLQ bei Männern mit PCa in dem Auftreten eines Response-Shift gesehen, der mit der vorliegenden Arbeit untersucht werden soll.

Belastungen von Patienten im Rahmen der onkologischen Rehabilitation lassen sich nach Lübke (1998) beschreiben durch die Ungewissheit über den weiteren Krankheitsverlauf, die

Belastungen durch die Therapie, die Notwendigkeit der kontinuierlichen diagnostischen Überwachung und die möglicherweise lang anhaltende psychische Belastung für die erkrankte Person. Daher sind die wesentlichen Ziele der onkologischen Rehabilitation die Reduktion von körperlichen, psychischen und sozialen Beeinträchtigungen, die durch die Krebserkrankung oder deren Behandlung entstehen, sowie die Verbesserung der Lebensqualität und die soziale Integration (AWMF 2002). Um diese Ziele zu erreichen, werden verschiedene therapeutische Bausteine eingesetzt: Allgemeine (bspw. Physiotherapie) und spezifische therapeutische Maßnahmen (bspw. Schmerztherapie), Ernährungs-, Schulungs- und Informationsangebote, psychologische Interventionen und Psychotherapie sowie soziale Hilfeleistungen und gegebenenfalls berufliche Maßnahmen. Der EORTC-QLQ-C30 dient nach Heim und Schwerte (2006) in der onkologischen Rehabilitation als Selbstbeurteilungsinstrument zur Messung der GLQ.

Ein Phänomen, das in den vergangenen Jahren häufiger mit der Messung von Lebensqualität in Zusammenhang gebracht wurde, ist der so genannte *Response-Shift*. Dabei handelt es sich um eine Veränderung der (individuellen) Bewertungsgrundlage für die Einschätzung zentraler Konzepte, wie beispielsweise die GLQ, bedingt durch massive oder chronische Belastungen. Response-Shift kann unterschiedliche Formen annehmen (Sprangers & Schwartz 2000): Bei einer *Rekalibrierung* verändert sich die interne subjektive Skalierung, die eine Person für die Bewertung eines Items beziehungsweise der GLQ verwendet. Dabei verschieben sich entweder die Ankerpunkte einer Skala (Pole) oder die subjektive Bewertung der einzelnen Skalenabstufungen ändert sich. Bei einer *Repriorisierung* unterscheidet sich die quantitative Bedeutung einzelner Komponenten der GLQ zu verschiedenen Zeitpunkten. Bei der *Neukonzeptualisierung* wird das Konstrukt GLQ zu unterschiedlichen Zeitpunkten qualitativ anders definiert. Eine Veränderung des Bewertungshintergrunds durch eine Rekalibrierung, Repriorisierung oder Neukonzeptualisierung ist aus Sicht der Veränderungsmessung für die Auswertung von Fragen in Selbstbeurteilungsinstrumenten von Bedeutung.

Das *Prozessmodell von Response-Shift und Lebensqualität* nach Sprangers und Schwartz (1999) bildet den theoretischen Ausgangspunkt der Arbeit. Das Modell setzt sich aus verschiedenen Faktoren zusammen. Der zentrale Prozess ist die Veränderung des Bewertungshintergrunds für die Einschätzung der GLQ. Einflussfaktoren auf diesen Prozess sind ereignisbezogene, personale, soziale sowie zeitbezogene Variablen. Auslöser für eine Umbewertung der GLQ werden im Modell als *Katalysatoren* bezeichnet. Genaue Merkmale eines sol-



chen Katalysators sind bisher nicht untersucht. Es wird jedoch angenommen, dass eine kritische Veränderung des Gesundheitszustands durch Auftreten einer bedrohlichen oder gravierenden Erkrankung sowie der Beginn oder Wechsel der Behandlung einer solchen Krankheit als Katalysator wirken kann. Im weiteren Verlauf der Auseinandersetzung mit dieser Erkrankung wird die Bewertung der GLQ durch Vorbedingungen beziehungsweise *spezifische Persönlichkeitsmerkmale* einerseits sowie Mechanismusvariablen beziehungsweise *Bewältigungsfaktoren* andererseits beeinflusst. Ausgehend von diesen Annahmen wurde ein spezifisches Wirkmodell abgeleitet. Dieses ist in Abbildung 5 dargestellt.

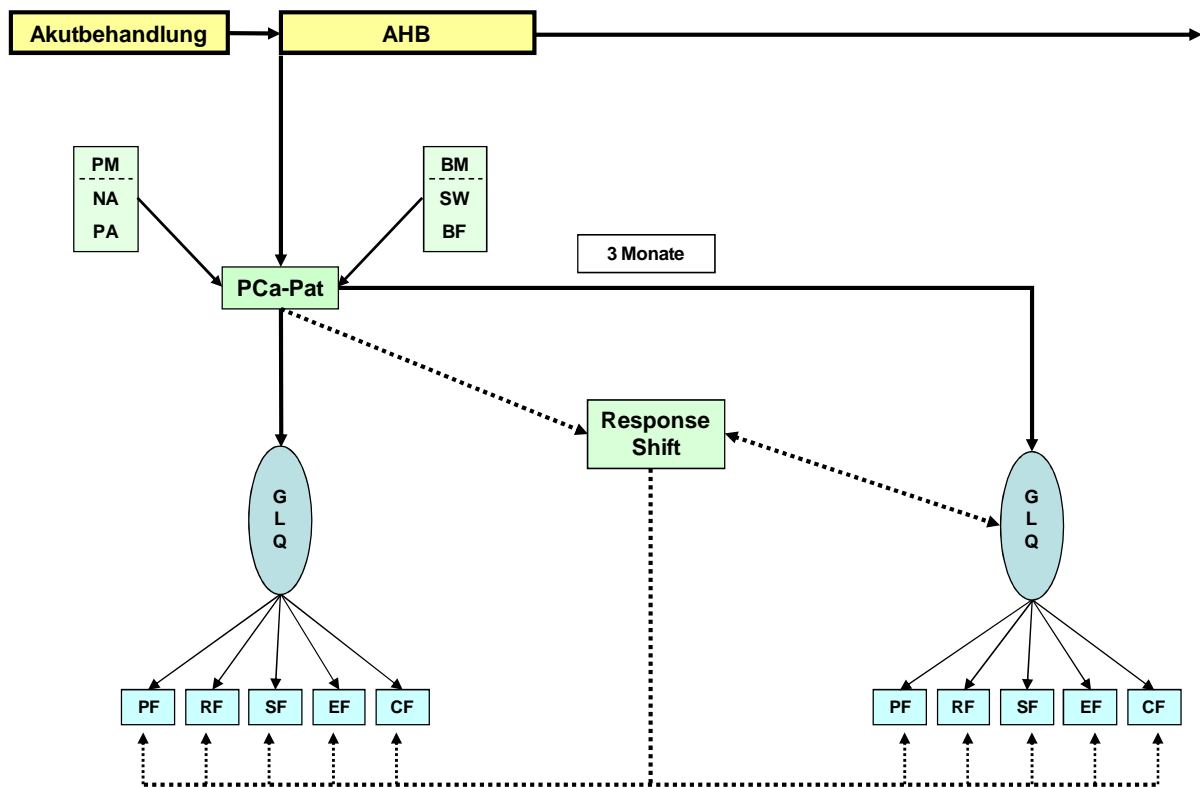


Abbildung 5: Wirkmodell der Veränderung gesundheitsbezogener Lebensqualität und Response-Shift bei PCa-Patienten mit Anschlussheilbehandlung

Anmerkungen: PM = Persönlichkeitsmerkmale, NA = Negative Affektivität, PA = Positive Affektivität, PCa-Pat. = PCa-Patienten, GLQ = Gesundheitsbezogene Lebensqualität, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, BM = Bewältigungsmechanismen, SW = Selbstwirksamkeit, BF = Benefit Finding, AHB = Anschlussheilbehandlung

In der vorliegenden Arbeit werden Patienten mit einem PCa zu Beginn der Anschlussheilbehandlung sowie drei Monate nach dieser Ersterhebung katamnestisch hinsichtlich der GLQ mit dem EORTC-QLQ-C30 untersucht. Patienten mit einer AHB beginnen diese rehabilitative Maßnahme in der Regel in unmittelbarem Anschluss an die Akutbehandlung. Akut-

medizinisch werden die Patienten entweder chirurgisch, mit Strahlentherapie, Chemotherapie oder einer Kombination aus diesen Verfahren behandelt. Aufgrund der zeitlichen Nähe zur Akutbehandlung wird angenommen, dass sich die Auseinandersetzung mit dem PCa sowie dessen Behandlung in einem frühen, akuten Stadium, in einer kritischen Phase befindet. Patienten in diesem Stadium der medizinisch-rehabilitativen Versorgung sind in der Regel durch die Grunderkrankung, die Folgen der Akutbehandlung sowie verschiedene damit zusammenhängende Nebenwirkungen beeinträchtigt. Dem Modell zufolge wird davon ausgegangen, dass der Wechsel von der Akutversorgung in die AHB relevant für die Beurteilung der GLQ ist. Steht im Rahmen der Akutbehandlung in erster Linie die medizinisch-organische Versorgung des Patienten im Vordergrund, geht mit der Teilnahme an einer AHB eine stärker ganzheitliche Orientierung, im Sinne eines bio-psycho-sozialen Modells, hinsichtlich Behandlung und Bewältigung der Erkrankung einher. Außerdem wird angenommen, dass durch die AHB der soziale Vergleich mit anderen PCa-Patienten einen besonderen Einfluss auf die Bewertung des bisherigen und weiteren Krankheitsverlaufs und auf die GLQ der einzelnen Betroffenen erhält. Es wird weiterhin davon ausgegangen, dass der Beginn der AHB ein Ereignis ist, das entweder einen Response-Shift auslösen oder diesen zumindest deutlich beeinflussen kann. Als *Katalysatoren* wird in der vorliegenden Arbeit neben der Erkrankung an sich der Wechsel der Therapie von Akut- zur Rehabilitationsbehandlung definiert.

Studien zu *Einflussfaktoren* auf Response-Shift gibt es bisher nur in Ausnahmefällen (Lepore & Eton 2000, Schwartz et al. 2006). Die vorliegende Arbeit soll einen Beitrag zum Schließen dieser Lücke leisten. Lepore und Eton (2000) kamen zu dem Ergebnis, dass das Ausmaß an Lebensqualitätsveränderung ein Prädiktor für das Ausmaß an Response-Shift ist. Je deutlicher die Veränderung der Lebensqualität, umso ausgeprägter fand eine Rekalibrierung bei den untersuchten PCa-Patienten statt. Im Rahmen dieser Arbeit soll der Einfluss der Lebensqualitätsveränderung auf einen Response-Shift ebenfalls untersucht werden. Dabei wird jedoch nicht nur die Rekalibrierung berücksichtigt, sondern auch die Repriorisierung sowie die Neukonzeptualisierung. Weiterhin werden zwei unterschiedliche personale Variablen als Einflussfaktoren auf einen Response-Shift bei Patienten mit PCa in der Anschlussheilbehandlung berücksichtigt: *Bewältigungsmechanismen* und *Persönlichkeitsmerkmale*. Die untersuchten Bewältigungsmechanismen in der vorliegenden Arbeit sind Benefit Finding und Selbstwirksamkeitserwartung. Bei den Persönlichkeitsmerkmalen handelt es sich um positive und negative Affektivität.

*Benefit Finding* (Affleck & Tennen 1996) ist ein kognitiver Bewältigungsprozess, der durch eine bewusstere Wahrnehmung des eigenen Lebens und der Umwelt, das Setzen neuer relevanter Ziele und Werte, einer Zunahme an persönlicher Stärke und Kraft und Ähnliches gekennzeichnet ist (Tedeschi et al. 1998). Insbesondere das Setzen neuer Ziele und die Orientierung an neuen Werten ist ein Prozess, der für Response-Shift im Zusammenhang mit einer Repriorisierung beziehungsweise Neukonzeptualisierung angenommen wird. Beachtenswert für die vorliegende Arbeit ist der Befund, dass Benefit Finding einerseits negativ korreliert mit Depressivität und positiv mit intrusiven Gedanken (Helgeson et al 2006). Im Rahmen der vorliegenden Arbeit wird davon ausgegangen, dass intrusive Gedanken ein Zeichen der kognitiven Verarbeitung der als belastend erlebten PCa-Erkrankung beziehungsweise deren Behandlung ist, und dass es sich dabei um einen Hinweis auf eine kognitive Umstrukturierung relevanter persönlicher Lebenskonzepte und -ziele handelt. Die Studienlage zur Beziehung zwischen Benefit Finding und GLQ bei PCa-Patienten ist unklar (Kinsinger et al. 2006).

Die *Allgemeine Selbstwirksamkeit* umfasst die Beurteilung eigener Kompetenzen, die hilfreich sind für die Bewältigung oder Kontrolle von schwierigen, neuen oder anforderungsreichen Aufgaben (Bandura 1997, Luszczynska et al. 2005). In der Auseinandersetzung mit Krebserkrankungen konnte nachgewiesen werden, dass eine positive Selbstwirksamkeitserwartung ein protektiver Faktor für die Bewältigung der Erkrankung ist (Lev 1997) und in einem positiven Zusammenhang mit Lebensqualität steht (Böhmer et al. 2007). In der vorliegenden Arbeit wird Selbstwirksamkeitserwartung als kognitiver Bewältigungsmechanismus verstanden, der die positive Neubewertung einer kritischen Lebenssituation entweder dadurch unterstützt, dass eine Situation umbewertet wird (externale Perspektive) oder die eigenen Möglichkeiten unter einer neuen, jedoch weiterhin positiven Perspektive betrachtet werden (internale Perspektive). Beide Prozesse können als Einflussfaktor auf einen Response-Shift gewertet werden.

Es werden zwei verschiedene Persönlichkeitsmerkmale in Bezug auf ihren Einfluss auf einen Response-Shift untersucht: Die *Positive* und *Negative Affektivität*. Nach Watson und Kollegen (1988b) umfasst positive Affektivität als persönliche Disposition eine Stabilität in Befindensmerkmalen wie Energie, Engagement und Konzentration. Negative Affektivität beinhaltet das überwiegende Erleben von Gefühlszuständen wie Gereiztheit, Nervosität und Ängstlichkeit. Sprangers und Schwartz (2000) nehmen an, dass positive Affektivität und damit verbundene Eigenschaften wie beispielsweise Optimismus und ein hohes Selbstwertge-

fühl eher response-shift-förderlich sind, wohingegen negative Affektivität und nahe liegende Konstrukte wie Depressivität oder externale Kontrollüberzeugungen das Auftreten eines Response-Shift vermindern. Diese Annahmen sollen anhand der vorliegenden Arbeit geprüft werden.

Es existieren verschiedene Verfahren, um Response-Shift zu erfassen. Das am häufigsten verwendete Verfahren ist der sog. Thentest, bei dem zum Postmesszeitpunkt neben der auf den aktuellen Zeitpunkt bezogenen Beurteilung des untersuchten Konstrukts eine zusätzliche Retrospektivbeurteilung des Konstrukts für den Zeitpunkt der Prämessung durchgeführt wird. Vertreter dieser Methode gehen davon aus, dass die Differenz zwischen der damaligen (zum Prämesszeitpunkt erhobenen) und der retrospektiven (zum Postmesszeitpunkt erhobenen) Prämessung einen Response-Shift abbildet. Mit der Methode des Thentests wird jedoch nur die Rekalibrierung erfasst. Ein Ansatz zur Erfassung eines Response-Shift mit Hilfe des statistischen Verfahrens der Konfirmatorischen Faktorenanalyse (KFA) stammt von Oort (2005a). Diese Variante der Response-Shift-Messung benötigt keine Retrospektiverfassung und liefert aufgrund statistischer Parameter Informationen zu allen drei Response-Shift-Prozessen. Die Rekalibrierung lässt sich nach diesem gruppenbezogenen Ansatz nochmals in einer uniforme und eine non-uniforme Form unterteilen. Eine *uniformen Rekalibrierung* liegt dann vor, wenn sich die Achsenschnittpunkte einer manifesten Variablen unabhängig vom Mittelwert des latenten Konstrukts zwischen den MZPen ändern. Der Wert einer Skala besitzt demnach zu beiden MZPen eine unterschiedliche (quantitative) Bedeutung, unabhängig von der latenten Variable GLQ. Inhaltlich ist bei einer uniformen Rekalibrierung davon auszugehen, dass alle Abstufungen einer Skala eine andere quantitative Bedeutung erlangen beziehungsweise die befragten Personen die gesamte Skala einer beobachtbaren Variable zu beiden MZPen unterschiedlich skalieren, wobei die Differenz zwischen den Skalenstufen gleich bleibt. Von einer *non-uniformen Rekalibrierung* wird dann gesprochen, wenn zwischen zwei MZPen eine Änderung der Varianz der manifesten Variablen im Messmodell stattfindet, wobei diese nicht im Zusammenhang mit einer Varianzveränderung der latenten Variable steht. Inhaltlich bedeutet diese Form der Rekalibrierung, dass die Personen die Skalierung der manifesten Variablen zum zweiten MZP uneinheitlich (neu) adjustieren, also nur einigen Skalenpunkten eine andere (quantitative) Bedeutung beigemessen wird. Eine uniforme Rekalibrierung ist daran zu erkennen, dass *lediglich* eine Interceptfreisetzung auf einem Indikator der GLQ stattfindet. Die Messfehlervarianz dieser manifesten Variablen bleibt jedoch restringiert. Eine non-uniforme Rekalibrierung kann sich *entweder* nur in der Form der Freisetzungen der

Messfehlervarianz *oder* in der Freisetzung der Kombination aus Messfehlervarianz und Intercept eines Indikators äußern. Eine *Repriorisierung* ist dann anzunehmen, wenn sich die Faktorladungen zu beiden MZPe stark unterscheiden, ohne dass eine Ladung zu einem der beiden MZPe einen Wert nahe oder gleich Null annimmt. Inhaltlich bedeutet diese Veränderung eine Ab- oder Zunahme der Bedeutung der entsprechenden manifesten Variablen für das latente Konstrukt. Eine *Neukonzeptualisierung* ist dann gegeben, wenn die Faktorladung einer manifesten Variablen zu einem der beiden MZPe bei (nahe) Null liegt und zum anderen MZP eine für die Annahme des Messmodells angemessene Größe besitzt. Inhaltlich bedeutet eine solche Veränderung einer Faktorladung, dass die dazugehörige manifeste Variable lediglich zu einem MZP eine Bedeutung besitzt und somit das latente Konstrukt sich *qualitativ* verändert hat.

Oort (2004) legt seinem Vorgehen zur Response-Shift-Testung mittels KFA das *Linear Latent Variable Model* (LLVM) zugrunde. Das LLVM geht von der Annahme aus, dass sich der Messwert einer Person auf einer Indikatorvariable der GLQ zusammensetzt aus einem Anteil eines spezifischen Werts dieses Indikators und einem Anteil des latenten Konstrukts GLQ. Weiterhin wird davon ausgegangen, dass auch klassische Messfehlereinflüsse für die Ausprägung dieses beobachteten Werts relevant sein können, diese aber eher zu vernachlässigen sind. Die Repriorisierung und die Neukonzeptualisierung lassen sich über die Veränderungen der Regressionskoeffizienten der latenten Variable GLQ berechnen, was einem faktoranalytischen Ansatz entspricht. Die uniforme und non-uniforme Rekalibrierung wird nach dem LLVM über die Intercepts beziehungsweise Messfehler operationalisiert. Im Sinne der IRT wird davon ausgegangen, dass es sich bei diesen Parametern um Personenparameter handelt, die den spezifischen Faktor repräsentieren, also den Restanteil der Varianz beziehungsweise des Wertes der beobachteten Indikatorausprägung, der unabhängig vom Einfluss der GLQ ist, aber dennoch inhaltlich für den Indikator Relevanz besitzt. Der spezifische Faktorwert sowie dessen Regressionskoeffizient sind in der allgemeinen IRT sowie in der KTT nicht enthalten. Es wird jedoch davon ausgegangen, dass diese genauso interpretierbar sind, wie das latente Konstrukt und dessen Regressionskoeffizienten. Ausgehend von den Berechnungen der verschiedenen Response-Shift-Parameter, dem Wert für die latente Variable GLQ sowie den Varianzen/Kovarianzen der Indikatoren zu zwei MZPen lassen sich die beobachteten Effekte auf den jeweiligen Indikatorvariablen unterteilen in Response-Shift-Effekte und einen Effekt bedingt durch die wahre Veränderung der GLQ. Für die Response-Shifts-Prozesse Rekalibrierung, Repriorisierung und Neukonzeptualisierung lassen sich auf Mittelwertebene

wiederum Einzeleffekte berechnen, so dass eine differenzierte Aussage über alle Response-Shift-Prozesse und die Veränderung durch die GLQ möglich wird. Der Nachteil des Ansatzes besteht darin, dass Response-Shift nur auf Gruppenebene erfasst werden kann und sich keine Aussagen über die individuellen Unterschiede zwischen den Gruppenmitgliedern oder einem individuellen Auftreten dieses Phänomens ableiten lassen. Es ist eine entscheidende Voraussetzung für die Verwendung dieser Response-Shift Erfassungsstrategie, dass Response-Shift bei einem großen Teil der Stichprobe auftritt. In der vorliegenden Arbeit wird der statistische Ansatz der KFA eingesetzt, um Response-Shift im Zusammenhang mit der Veränderung der GLQ bei PCa-Patienten im Rahmen der AHB zu erfassen.

## 5. Fragestellungen der vorliegenden Arbeit

Mit der vorliegenden Arbeit soll Response-Shift im Rahmen der Erfassung von GLQ bei PCa-Patienten während und nach der Anschlussheilbehandlung (AHB) anhand des statistischen Verfahrens der Konfirmatorischen Faktorenanalyse untersucht werden. In einem ersten Schritt soll geprüft werden, ob ein Response-Shift als Einflussfaktor auf die Lebensqualitätserfassung in der vorliegenden Stichprobe auftritt (Fragestellung 1). In einem weiteren Schritt soll geprüft werden, ob eine deutliche Veränderung der Globalen Lebensqualität zwischen den MZPen zu mehr und größeren Response-Shift-Effekten beiträgt als geringe Veränderungen in der Globalen Lebensqualität der PCa-Patienten. (Fragestellung 2). Abschließend sollen potenzielle Moderatorvariablen auf ihre Einflüsse auf einen Response-Shift untersucht werden (Fragestellung 3). Die Fragestellung 1 stellt die Hauptfragestellung der vorliegenden Arbeit dar. Die Fragestellungen 2 und 3 zu den Analysen der Einflussfaktoren werden als Nebenfragestellungen bearbeitet.

### *Fragestellung 1*

Hat Response-Shift einen Einfluss auf die Erfassung der gesundheitsbezogenen Lebensqualität in der onkologischen Rehabilitation von PCa-Patienten im Rahmen einer AHB und falls ja, in welcher Form?

Die dargestellten Untersuchungsergebnisse zur Erfassung von Response-Shift im Zusammenhang mit der Erfassung von GLQ bei Menschen mit Krebserkrankungen und deren Behandlung im Allgemeinen und Prostatakrebs im Speziellen weisen darauf hin, dass auch bei PCa-Patienten in der Anschlussheilbehandlung vom Auftreten eines Response-Shifts ausgegangen werden kann. Entsprechend dem Modell von Sprangers und Schwartz (1999) wird angenommen, dass in der untersuchten Stichprobe ein Response-Shift im Rahmen der Auseinandersetzung mit einer lebensbedrohlichen Erkrankung stattfindet. Wissenschaftliche Befunde zur Untersuchung des Phänomens im Zusammenhang mit Behandlungen von Krebserkrankungen in der Akutversorgung bestätigen diese These (z.B. Oort 2005b). Studien zum Response-Shift im Rahmen stationärer Rehabilitationsmaßnahmen weisen ebenfalls auf diesen Zusammenhang hin (Nübling et al. 2004, Kohlmann & Raspe 1998). Die vorliegende Arbeit untersucht erstmals das Auftreten eines Response-Shift bei PCa-Patienten in der An-

schlussheilbehandlung und geht aufgrund der theoretischen Annahmen, der bisherigen Befunde sowie der methodischen Vorgehensweise von folgenden Hypothesen aus:

Aufgrund der zeitlichen Nähe zur Akutbehandlung und der relativ kurzen Zeit nach Beginn der Behandlung des Prostatakarzinoms sowie der Intervention im Rahmen der Anschlussheilbehandlung tritt in der untersuchten Stichprobe von 212 PCa-Patienten ein Response-Shift im Zeitraum von drei Monaten von Beginn der AHB bis drei Monate nach Beginn der AHB auf.

Dieser Response-Shift tritt auf verschiedenen Parameterebenen auf, die mittels der Methode der Konfirmatorischen Faktorenanalyse erfasst werden können. Es findet ein Response-Shift in Form einer uniformen Rekalibrierung im Sinne einer Veränderung von Skalenintercepts statt. Ebenso wird erwartet, dass eine non-uniforme Rekalibrierung auf einzelnen Skalen in Form einer Invarianz der Messfehlervarianzen identifiziert werden kann. Diese tritt entweder lediglich auf der Ebene der Messfehlervarianz auf oder auf der Ebene der Messfehlervarianz und des Intercepts einer Skala. Eine Repriorisierung wird auf der Ebene der Faktorladungen erwartet, die in massiver Ausprägung in eine Neukonzeptualisierung übergeht. Dies ist dann der Fall, wenn eine der Faktorladungen eines GLQ-Indikators lediglich zu einem MZP einen Wert um Null und zum anderen MZP einen Wert mit substantieller Höhe aufweist. Die Faktorladungen der Skalen des EORTC-QLQ-C30, die zur Modellbildung des latenten Faktors GLQ verwendet werden, unterliegen einem Response-Shift. Überprüft werden die Hypothesen mithilfe der Konfirmatorischen Faktorenanalysen als statistischem Verfahren und den Veränderungen der Modellpassung je nach Freisetzung der genannten Parameter.

Aus diesen inhaltlichen Hypothesen lassen sich verschiedene statistische Hypothesen ableiten.

Die Nullhypothese bezüglich der **uniformen Rekalibrierungen** lautet:

Die Intercepts der Skalen (1) Physische, (2) Rollen-, (3) Soziale, (4) Emotionale und (5) Kognitive Funktionsfähigkeit zum Messzeitpunkt 1 ( $\tau_{x1}$ ) sind gleich den Intercepts dieser Skalen zum Messzeitpunkt 2 ( $\tau_{x2}$ ).



Daraus ergeben sich im Einzelnen folgende Null- und Alternativhypothesen für die uniforme Rekalibrierung:

Tabelle 11: Hypothesen zur uniformen Rekalibrierung

Nullhypothesen	Alternativhypothesen
$\tau_{11} = \tau_{12}$	$\tau_{11} \neq \tau_{12}$
$\tau_{21} = \tau_{22}$	$\tau_{21} \neq \tau_{22}$
$\tau_{31} = \tau_{32}$	$\tau_{31} \neq \tau_{32}$
$\tau_{41} = \tau_{42}$	$\tau_{41} \neq \tau_{42}$
$\tau_{51} = \tau_{52}$	$\tau_{51} \neq \tau_{52}$

Anmerkungen:  $\tau$  = Intercepts,  $1x-5x$  = Lebensqualitätsskalen des EORTC-QLQ-C30,  $x_1$  = Messzeitpunkt 1,  $x_2$  = Messzeitpunkt 2

Die Nullhypothese gilt als widerlegt, wenn auf wenigstens einer Skala eine uniforme Rekalibrierung auftritt und gleichzeitig die Messfehlervarianz dieser Skala restringiert bleibt.

Die Nullhypothese bezüglich der **non-uniformen Rekalibrierungen** lautet:

Die Messfehlervarianzen der (1) Physischen, (2) Rollen-, (3) Sozialen, (4) Emotionalen und (5) Kognitiven Funktionsfähigkeit zum Messzeitpunkt 1 ( $\delta_{x1}$ ) sind gleich den Messfehlervarianzen dieser Skalen zum Messzeitpunkt 2 ( $\delta_{x2}$ ).

Daraus ergeben sich im Einzelnen folgende Null- und Alternativhypothesen hinsichtlich der non-uniformen Rekalibrierung:

Tabelle 12: Hypothesen zur non-uniformen Rekalibrierung

Nullhypothesen	Alternativhypothesen
$\delta_{11} = \delta_{12}$	$\delta_{11} \neq \delta_{12}$
$\delta_{21} = \delta_{22}$	$\delta_{21} \neq \delta_{22}$
$\delta_{31} = \delta_{32}$	$\delta_{31} \neq \delta_{32}$
$\delta_{41} = \delta_{42}$	$\delta_{41} \neq \delta_{42}$
$\delta_{51} = \delta_{52}$	$\delta_{51} \neq \delta_{52}$

Anmerkungen:  $\delta$  = Messfehlervarianzen,  $1x-5x$  = Lebensqualitätsskalen des EORTC-QLQ-C30,  $x_1$  = Messzeitpunkt 1,  $x_2$  = Messzeitpunkt 2

Die Nullhypothese gilt als widerlegt, wenn auf wenigstens einer Skala eine non-uniforme Rekalibrierung auftritt.

Die Nullhypothese bezüglich der **Repriorisierung** lautet:

Die Faktorladungen der Skalen (1) Physische, (2) Rollen-, (3) Soziale, (4) Emotionale und (5) Kognitive Funktionsfähigkeit zum Messzeitpunkt 1 ( $\lambda_{x1}$ ) sind gleich den Faktorladungen dieser Skalen zum Messzeitpunkt 2 ( $\lambda_{x2}$ ).

Daraus ergeben sich im Einzelnen folgende Null- und Alternativhypothesen hinsichtlich der Repriorisierung:

Tabelle 13: Hypothesen zur Repriorisierung

Nullhypothesen	Alternativhypothesen
$\lambda_{11} = \lambda_{12}$	$\lambda_{11} \neq \lambda_{12}$
$\lambda_{21} = \lambda_{22}$	$\lambda_{21} \neq \lambda_{22}$
$\lambda_{31} = \lambda_{32}$	$\lambda_{31} \neq \lambda_{32}$
$\lambda_{41} = \lambda_{42}$	$\lambda_{41} \neq \lambda_{42}$
$\lambda_{51} = \lambda_{52}$	$\lambda_{51} \neq \lambda_{52}$

Anmerkungen:  $\lambda$  = Faktorladung,  $1x - 5x$  = Lebensqualitätsskalen des EORTC-QLQ-C30,  $x_1$  = Messzeitpunkt 1,  $x_2$  = Messzeitpunkt 2

Die Nullhypothese gilt als widerlegt, wenn auf wenigstens einer Skala eine Repriorisierung auftritt.

Die Nullhypothese bezüglich der **Neukonzeptualisierung** lautet:

Die Faktorladungen der Skalen (1) Physische, (2) Rollen-, (3) Soziale, (4) Emotionale und (5) Kognitive Funktionsfähigkeit sind zu beiden MZPen signifikant von Null verschieden und laden zu beiden MZPen auf einem Faktor. Daraus ergibt sich folgende Null- und Alternativhypothese hinsichtlich der Neukonzeptualisierung:

Tabelle 14: Hypothesen zur Neukonzeptualisierung

Nullhypothese	Alternativhypothese
$\Sigma_{t1} = \Sigma_{t2}$	$\Sigma_{t1} \neq \Sigma_{t2}$

Anmerkungen:  $\Sigma$  = Kovarianzmatrix,  $t_1$  = Messzeitpunkt 1,  $t_2$  = Messzeitpunkt 2

Die Nullhypothese gilt als widerlegt, wenn eine Skala zu einem MZP eine Faktorladung von (nahe) Null aufweist und diese Faktorladung zum anderen MZP deutlich von Null ab-

weicht. Der Unterschied in der Höhe der Faktorladungen dieser Skala muss statistisch signifikant sein.

### *Fragestellung 2*

Hat das Ausmaß der Veränderung der Globalen Lebensqualität einen Einfluss auf den Response-Shift und falls ja, in welcher Form?

Nach dem Modell von Sprangers und Schwartz (1999) führt die Veränderung des Gesundheitszustands als Katalysator über ein Zusammenwirken aus Vorbedingungen (z. B. Persönlichkeitsvariablen) und Mechanismen (z. B. Bewältigungsprozesse) zu einem Response-Shift, der wiederum die wahrgenommene Lebensqualität beeinflusst. Dem Modell zufolge findet zusätzlich ein Rückkopplungsprozess statt von der wahrgenommenen Lebensqualität auf die Verarbeitung des Katalysators.

Wie die Response-Shift-Prozesse Repriorisierung, Neukonzeptualisierung, uniforme und non-uniforme Rekalibrierung im Zusammenhang stehen mit der wahrgenommenen Lebensqualität, wurde bisher nur wenig untersucht. Lepore und Eton (2000) untersuchten die Prädiktionskraft verschiedener Variablen auf eine Rekalibrierung. Anhand einer Hierarchischen Regressionsanalyse erwies sich lediglich das Ausmaß der Lebensqualitätsveränderung als signifikanter Prädiktor. Die Rekalibrierung wurde mit der Thenteststrategie erfasst. Eine Untersuchung des Einflusses der Veränderung der GLQ auf einen Response-Shift, gemessen mit Hilfe der Methode nach Oort, wurde bisher nicht untersucht. Nach Studien von Jansen und Kollegen (2001) zur Beziehung zwischen der Veränderung von GLQ und Response-Shift können zwei verschiedene Rekalibrierungsprozesse auftreten, wobei die Richtung des Response-Shifts durch die Richtung der Lebensqualitätsveränderung mitbestimmt wird: Eine positive Rekalibrierung liegt dann vor, wenn eine positive Veränderung der allgemeinen Lebensqualität zwischen zwei Messzeitpunkten (gemessen mittels gängiger Prä-Post-Mittelwertsdifferenzen) durch Rekalibrierungsprozesse (gemessen mithilfe des Thentests) nochmals übertroffen wird. Eine negative Rekalibrierung zeigt sich darin, dass eine Verschlechterung der allgemeinen Lebensqualität zwischen zwei Messzeitpunkten (gemessen anhand eines normalen Prä-Post-Vergleichs) unter Einbeziehung von Rekalibrierungsprozessen (mittels Thentest erfasst) nochmals deutlicher ausfällt. In beiden Fällen wird – ausgedrückt in Effekten – ein beobachteter Effekt durch eine Rekalibrierung nochmals vergrößert.

In der vorliegenden Arbeit soll untersucht werden, ob es Unterschiede im Ausmaß des Response-Shifts in Abhängigkeit vom Ausmaß der Veränderung der allgemeinen Lebensqualität gibt. Aufgrund der Ergebnisse von Lepore und Eton (2000) ist anzunehmen, dass das Ausmaß der Veränderung der allgemeinen Lebensqualität nicht nur auf eine Rekalibrierung, sondern auch auf eine Repriorisierung oder eine Neukonzeptualisierung Einfluss nehmen kann. Nach Jansen und Kollegen (2001) wird von der Annahme ausgegangen, dass positive und negative Rekalibrierungen zwei Prozesse sind, die einem Kontinuum zuzuordnen, das heißt diese nicht voneinander unabhängig sind, sondern ineinander ‚übergehen‘. Findet ein Response-Shift statt und kommt es zu einer Verbesserung der allgemeinen Lebensqualität, führt das zum Auftreten eines positiven Response-Shifts im Sinne einer Überschätzung der herkömmlich erfassten Prä-Post-Differenz. Finden eine Verschlechterung der wahrgenommenen Lebensqualität und ein Response-Shift statt, resultiert daraus ein negativer Response-Shift ebenfalls im Sinne einer Überschätzung der herkömmlich erfassten Prä-Post-Differenz. Findet keine Veränderung der wahrgenommenen Lebensqualität statt, tritt auch kein Response-Shift auf. Daraus wird weiterhin abgeleitet, dass unterschiedliche positive beziehungsweise negative Differenzen in der wahrgenommenen Lebensqualität mit unterschiedlichen Response-Shift-Ausmaßen in Beziehung stehen und diese Beziehung linear verläuft: Je größer die Verbesserung beziehungsweise Verschlechterung der Globalen Lebensqualität, desto größer der Response-Shift.

Die zuvor abgeleiteten Annahmen zur Fragestellung 2 führen zu folgendem Vorgehen: Die Gesamtstichprobe der 212 PCa-Patienten wird aufgrund der Veränderungen in der allgemeinen Lebensqualität anhand des Medians (der Veränderung) dichotomisiert. Gemessen wird die allgemeine Lebensqualität mit der Skala ‚Globale Lebensqualität‘ des EORTC-QLQ-C30. Die individuellen Differenzwerte werden mittels Prä-Post-Vergleich ermittelt. Entsprechend der formulierten Annahmen wird davon ausgegangen, dass Patienten mit einer deutlichen Verbesserung der Globalen Lebensqualität einen größeren Response-Shift zeigen als Personen mit weniger deutlichen Verbesserungen der Globalen Lebensqualität. Zwei verschiedene Kriterien lassen sich zur Überprüfung dieser Hypothese heranziehen. Das erste Kriterium bezieht sich auf die Anzahl der Response-Shift-Prozesse beziehungsweise Parameterfreisetzungen: Innerhalb der Gruppe mit einer deutlicheren Verbesserung der Globalen Lebensqualität finden mehr Parameterfreisetzungen statt und werden mehr Response-Shift-Prozesse erwartet. Zum anderen bezieht sich der Unterschied auf die Höhe der aufgedeckten Response-Shift-Effekte. Dabei wird angenommen, dass sich in der Gruppe mit deutlicheren

Verbesserungen der Globalen Lebensqualität größere Response-Shift-Effekte zeigen als in der entsprechenden Vergleichsgruppe.

Diese inhaltlichen Hypothesen lassen sich statistisch wie folgt formulieren:

Tabelle 15: Hypothesen zur Veränderung der Globalen Lebensqualität und Response-Shift

Nullhypothesen	Alternativhypothesen
$X_{RS \text{ Globale LQ-Differenz} \geq Md} \leq X_{RS \text{ Globale LQ-Differenz} \leq Md}$	$X_{RS \text{ Globale LQ-Differenz} \geq Md} > X_{RS \text{ Globale LQ-Differenz} \leq Md}$
$ES_{RS \text{ Globale LQ-Differenz} \geq Md} \leq ES_{RS \text{ Globale LQ-Differenz} \leq Md}$	$ES_{RS \text{ Globale LQ-Differenz} \geq Md} > ES_{RS \text{ Globale LQ-Differenz} \leq Md}$

$X_{RS}$  = Anzahl der Response-Shift-Prozesse

$ES_{RS}$  = Response-Shift-Effekt

Globale LQ-Differenz  $\geq Md$  = Gruppe der Personen mit deutlicher Veränderung der Globalen Lebensqualität

Globale LQ-Differenz  $\leq Md$  = Gruppe der Personen mit geringer Veränderung der Globalen Lebensqualität

Die Überprüfung der Hypothese hinsichtlich der Anzahl der Response-Shift-Prozesse lässt sich aufgrund der getrennten Durchführung der Response-Shift-Ableitungen nicht mittels eines statistischen Verfahrens auf statistische Signifikanz prüfen. Die Ergebnisse haben somit lediglich deskriptiven Charakter. Trotzdem wird in der vorliegenden Arbeit davon ausgegangen, dass wenn in der Gruppe ‚Globale LQ-Differenz  $\geq Md$ ‘ mehr Response-Shift-Prozesse stattfinden, dies als Hinweis auf die Ablehnung der Nullhypothese gewertet wird. Die Hypothese bezüglich der Unterschiede in den Response-Shift-Effekten wird anhand der Differenzen zwischen den aufgedeckten Response-Shift-Effekten zwischen beiden Gruppen überprüft. Zeigt sich in der Gruppe ‚Globale LQ-Differenz  $\geq Md$ ‘ auf einer Skala ein Response-Shift-Effekt, der den Response-Shift-Effekt in der Gruppe ‚Globale LQ-Differenz  $\leq Md$ ‘ um das Ausmaß eines kleinen Effektes nach Cohen überschreitet, wird dies als Hinweis auf die Ablehnung der Nullhypothese gewertet.

### Fragestellung 3

Welchen Einfluss haben Benefit Finding, die wahrgenommene Selbstwirksamkeit und die Persönlichkeitsmerkmale positive und negative Affektivität auf einen Response-Shift bei PCa-Patienten in der onkologischen Rehabilitation im Rahmen der AHB?

Nach dem Response-Shift-Modell von Sprangers und Schwartz (1999) gibt es verschiedene Faktoren, die einen Response-Shift beeinflussen können. Dazu zählen die Art des auslösenden Ereignisses und dessen Folgen als Katalysator, Persönlichkeitsaspekte, die eine personale Grundlage bezüglich der Art des Umgangs mit Anforderungen und Belastungen bildet und im Modell als *Vorbedingungen* bezeichnet werden, sowie Verarbeitungsstrategien und -prozesse, die eine Bewältigung eines Ereignisses in bestimmter Weise fördern. Letztgenannte Aspekte werden im Modell *Mechanismen* genannt.

Auch zur Beantwortung der Fragestellungen 3.1 bis 3.4 werden die Patienten der Gesamtstichprobe in zwei Gruppen anhand des Medians der entsprechenden Moderatorvariablen eingeteilt. Anschließend werden mit Hilfe der KFA für jede Gruppe separate Response-Shift-Testungen durchgeführt.

Da mit Ausnahme der Studie von Lepore und Eton (2000) und der Metaanalyse von Schwartz und Kollegen (2006) bisher keine empirischen Arbeiten zur Bedeutsamkeit von Einflussfaktoren auf einen Response-Shift existieren, wird für die Überprüfung der vorliegenden Fragestellungen 3.1 bis 3.4 ein hypothesengenerierender Ansatz verfolgt. Trotzdem werden im Folgenden zu jeder Fragestellung einige inhaltliche Annahmen formuliert, die im Rahmen der Diskussion der Ergebnisse zur Fragestellung 3 aufgegriffen werden sollen.

### *Fragestellung 3.1*

Welchen Einfluss hat *Benefit Finding* auf einen Response-Shift im Zusammenhang mit der Erfassung der gesundheitsbezogenen Lebensqualität bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Benefit Finding bezeichnet einen kognitiven Prozess, im Rahmen dessen Personen, die sich mit widrigen Umständen auseinandersetzen müssen, Möglichkeiten für positive persönliche Entwicklungen sehen und diese umzusetzen versuchen (Affleck und Tennen 1996). Das Setzen neuer relevanter Ziele und Werte ist ein für das Benefit Finding zentraler Prozess. Die überwiegende Mehrzahl der Studien zum Thema Benefit Finding und GLQ finden keinen Zusammenhang zwischen diesen beiden Variablen (bspw. Helgeson et al. 2006, Kinsinger et al. 2006)

Für die vorliegende Fragestellung ist bedeutsam, dass Benefit Finding mit Veränderungen interner Standards, Werte und Konzepte einhergeht. Relevant ist zudem der Befund, dass Benefit Finding positiv korreliert mit intrusiven Gedanken (Helgeson et al. 2006). Eine Erklärung dafür ist, dass eine kognitive Verarbeitung belastender Lebensereignisse durch intrusive Gedanken gekennzeichnet sein kann. Die intrusiven Gedanken gehören demnach zu einer angemessenen Bewältigung des Ereignisses beziehungsweise zu einer Veränderung interner Standards, da sich die Person gedanklich mit der gegenwärtigen Lebenssituation und den (früheren) internen Standards auseinandersetzt. Dieses ist ein konflikthafter Prozess. Im weiteren Verlauf führt Benefit Finding dann zu einer positiven Umdeutung der nicht mehr aufrechtzuerhaltenden Ziele und Werte sowie der gegebenen Lebenssituation beziehungsweise des belastenden Ereignisses durch kognitive Umstrukturierung. Diese kognitive Umstrukturierung beinhaltet die Veränderung interner Standards, die im Modell nach Sprangers und Schwartz (1999) als Response-Shifts bezeichnet werden. Somit wird für die Auswertung der vorliegenden Fragestellung davon ausgegangen, dass ein hohes Ausmaß an Benefit Finding mit mehr Response-Shift-Prozessen und größeren Response-Shift-Effekten einhergeht.

### *Fragestellung 3.2*

Welchen Einfluss hat die wahrgenommene *Selbstwirksamkeit* auf einen Response-Shift im Zusammenhang mit der Erfassung der gesundheitsbezogenen Lebensqualität bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Selbstwirksamkeit wird definiert als Kompetenzerwartung, schwierige, neue oder anforderungsreiche Aufgaben zu kontrollieren und zu bewältigen (Bandura 1997). Für die vorliegende Fragestellung ist von Bedeutung, dass von einem positiven Zusammenhang zwischen Selbstwirksamkeit und Lebensqualität auszugehen ist (Cunnigham et al. 1991). Nach Lev (1997) gilt eine hohe Selbstwirksamkeit als protektiver Faktor in der Auseinandersetzung mit Krebserkrankungen. Weber und Kollegen (2004) fanden einen negativen Zusammenhang zwischen Selbstwirksamkeit und Depressivität bei PCa-Patienten. Böhmer und Kollegen (2007) wiesen eine positive direkte Beziehung zwischen wahrgenommener Selbstwirksamkeit und physischer, sozialer und emotionaler Befindlichkeit als Komponenten der Lebensqualität bei einer heterogenen Stichprobe von Krebspatienten nach chirurgischen Eingriffen nach.

Der Einfluss der wahrgenommenen Selbstwirksamkeit auf einen Response-Shift ist bisher nicht bekannt. Grundsätzlich kann sich der Einfluss von Selbstwirksamkeit in zweierlei Hinsicht auswirken: Einerseits kann ein hohes Maß an Selbstwirksamkeit einen Response-Shift unterdrücken, da aufgrund assimilativer Bewältigungsstrategien eine positive Haltung zur Auseinandersetzung mit der Erkrankung und Behandlung eingenommen wird, indem Optimismus und Engagement zur Bewältigung der Situation aus dem Beibehalten an bisherigen Zielen und internen Standards geschöpft wird. Ziel ist in diesem Fall die Veränderung der Situation unter Beibehaltung interner Standards, Werte und persönlich relevanter Konzepte. Andererseits kann ein hohes Maß an Selbstwirksamkeit mit einer positiven Haltung einhergehen, indem von (aktuell) nicht erreichbaren internen Standards und Zielen abgewichen und auf die wesentlichen Lebensbereiche fokussiert wird (akkomodative Bewältigung), um in der neuen Lebenssituation besser zu Recht zu kommen. Diesem Fall liegt definitionsgemäß ein Response-Shift zugrunde. Das Ziel der Person ist die Adjustierung der internen Standards, Werte und Konzepte, wobei ein hohes Maß an Selbstwirksamkeit die Erwartung fördert, die geänderten Standards zu erreichen.

In der vorliegenden Studie mit PCa-Patienten im Rahmen der Anschlussheilbehandlung werden Personen untersucht, die in unmittelbarer zeitlichen Nähe zur Akutbehandlung und somit auch in einer frühen Phase der Auseinandersetzung und Bewältigung der Erkrankung stehen, weshalb nach dem Modell von Sprangers und Schwartz (1999) von einem Response-Shift auszugehen ist. Sollte demnach ein Response-Shift auftreten (Fragestellung 1), dann wird von der Annahme ausgegangen, dass Selbstwirksamkeit diesen Prozess fördert, d.h. dass Personen mit einem hohen Ausmaß an Selbstwirksamkeit mehr Response-Shift-Prozesse und größere Response-Shift-Effekte zeigen als Personen mit einer geringen Ausprägung der Selbstwirksamkeit.

### *Fragestellung 3.3*

Welchen Einfluss hat die *positive Affektivität* auf einen Response-Shift im Zusammenhang mit der Erfassung der gesundheitsbezogenen Lebensqualität bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Studien zur Untersuchung der Beziehung zwischen positiver Affektivität und Response-Shift liegen bisher nicht vor. Sprangers und Schwartz (1999) gehen aufgrund ihres Modells



davon aus, dass Response-Shift durch Persönlichkeitsvariablen wie Optimismus oder ein hohes Maß an Selbstwert beeinflusst wird. Die Autorinnen nehmen an, dass positive Affektivität einen Response-Shift fördert.

Für positive Affektivität wird ebenfalls, wie für die wahrgenommene Selbstwirksamkeit, angenommen, dass diese einen Response-Shift unterschiedlich beeinflussen kann. Einerseits wird angenommen, dass positive Affektivität einen Response-Shift unterdrücken kann. Das sollte dann der Fall sein, wenn sich die Person mit einem kritischen Lebensereignis auseinandersetzt, diese Auseinandersetzung jedoch eher auf das Ziel ausgerichtet ist, den Status Quo vor der Erkrankung zu stabilisieren. In einem solchen Fall kann im Rahmen des Umgangs mit einer lebensbedrohlichen Erkrankung ein hohes Ausmaß an positiver Affektivität die Folge einer Abwehr der anstehenden Auseinandersetzung mit der neuen Lebenssituation sein. Hierbei kann davon ausgegangen werden, dass eine bewusste Verarbeitung der gesamten Problemsituation (noch) nicht erfolgt ist. Für einen solchen Fall wird angenommen, dass kein Response-Shift stattfindet oder bisher stattgefunden hat. Andererseits kann positive Affektivität einen Response-Shift fördern beziehungsweise das Ergebnis eines Response-Shift sein. Fördern kann es diesen Prozess, da positive Affektivität durch verschiedene Aspekte operationalisiert wird, die für eine Veränderung interner Standards, Werte und Ziele relevant sein können. Dazu zählen Variablen wie „Stärke“, „Entschlossenheit“ oder „Aktivsein“. Wenn bei einer Person ein Response-Shift stattfindet, kann dieser durch ein hohes Ausmaß an positiver Affektivität gefördert werden. Weiterhin kann ein hohes Maß an positiver Affektivität das Ergebnis einer angemessenen Veränderung interner Standards, Werte und wichtiger persönlicher Konzepte sein, da ein neues Gleichgewicht zwischen diesen und der gegebenen Lebenssituation hergestellt werden konnte. Außerdem ist dispositioneller Optimismus nach Wrosch und Scheier (2003) ein wesentlicher Einflussfaktor auf Lebensqualität, und zwar nicht nur weil dadurch problemorientiertes Coping in Belastungssituationen gefördert wird, sondern auch weil es im Zusammenhang mit emotionsorientiertem Coping zu einer Änderungen von persönlichen Zielen und Werten (i. S. einer Repriorisierung) beitragen kann. Krohne und Kollegen (1996) gehen für die Skala Positive Affektivität davon aus, dass ein hohes Ausmaß an positiver Affektivität gekennzeichnet ist durch Energie, Konzentration und freudiges Engagement. Eine geringe Ausprägung läßt sich durch Traurigkeit und Lethargie beschreiben. Auch in diesem Sinne ist davon auszugehen, dass ein hohes Ausmaß an positiver Affektivität eher response-shift-förderlich ist als ein geringes Ausmaß.

Da die Messung der positiven Affektivität drei Monate nach Beginn der AHB durchgeführt wurde, und unter der Voraussetzung, dass in der vorliegenden Stichprobe ein Response-Shift stattgefunden hat, wird davon ausgegangen, dass positive Affektivität einen response-shift-fördernden Charakter besitzt. Ein hohes Ausmaß an positiver Affektivität sollte demnach mit mehr Response-Shift-Prozessen sowie größeren Response-Shift-Effekten einhergehen als eine geringe Ausprägung dieser Variable.

#### *Fragestellung 3.4*

Welchen Einfluss hat die *negative Affektivität* auf einen Response-Shift im Zusammenhang mit der Erfassung der gesundheitsbezogenen Lebensqualität bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Zur Beziehung zwischen der Persönlichkeitsvariablen negative Affektivität und Response-Shift liegen bisher ebenfalls keine empirischen Arbeiten vor. Sprangers und Schwartz (1999) gehen in ihrem Modell zwar davon aus, dass Depressivität in einem negativen Zusammenhang mit Response-Shift steht, legen dafür jedoch keinen empirischen Nachweis vor.

Die Beziehung zwischen negativer Affektivität und Response-Shift kann ebenfalls in zweifacher Richtung angenommen werden. Einerseits kann negative Affektivität das Auftreten eines Response-Shift unterdrücken, da davon ausgegangen werden kann, dass eine allgemeine negative Befindlichkeit die Veränderung interner Standards, Ziele und Werte eher erschwert, wenn der Response-Shift-Prozess im Sinne einer konstruktiven Umstrukturierung gesehen wird. Im Sinne emotionskongruenter Kognitionen ist in diesem Fall ein Festhalten an früheren Vorstellungen und einer damit einhergehenden Verstärkung einer negativen Befindlichkeit vorstellbar. Weiterhin spricht für diese Annahme, dass negative Affektivität nach der vorliegenden Definition eine Persönlichkeitsvariable ist, die durch Aspekte wie „Ängstlichkeit“, „Ärger“, „Bekümmern“ oder „Nervosität“ gekennzeichnet ist. Es wird angenommen, dass diese Einstellungen – als stabile Persönlichkeitseigenschaften – eine konstruktive Umstrukturierung von Zielen und Werten nach einem kritischen Lebensereignis eher erschweren als fördern. Dementsprechend wäre weiter davon auszugehen, dass eine Person mit einer hohen Ausprägung an negativer Affektivität auch in Situationen großer Belastungen weniger adaptiv – im Sinne einer Akkomodation oder Anpassung innerer Einstellungen – auf die neue Lebenssituation reagiert. Außerdem gehen Watson und Pennebaker (1989) von der so ge-

nannten *Symptom-Perception-Hypothese* aus, die annimmt dass ein höheres Ausmaß an negativer Affektivität mit einer gesteigerten Aufmerksamkeit für körperliche Beschwerden einhergeht. In einem solchen Fall ist nicht grundsätzlich auszuschließen, dass dies zu einem Response-Shift beitragen kann, indem durch die ständige Auseinandersetzung mit körperlichen Beschwerden eine Anpassung von Zielen und Werten erfolgt. Allerdings ist anzunehmen, dass die vermehrte Beschäftigung mit körperlichen Beschwerden einen Response-Shift eher verhindert, weil die Krankheitsverarbeitung weniger auf der kognitiven Ebene stattfindet und nicht vorrangig auf die Zukunftsperspektive ausgerichtet ist. Im Vordergrund stehen stattdessen die aktuellen Beschwerden und körperlichen Belastungen.

Andererseits kann ein hohes Ausmaß an negativer Affektivität als Zeichen der Auseinandersetzung mit neuen Werten, Zielen und Konzepten angesehen werden. Vor allem im Rahmen der Bewältigung einer lebensbedrohlichen Erkrankung sind die oben beschriebenen Affektzustände der negativen Affektivität als normale Phänomene zu erwarten, die somit auch im Zusammenhang mit einem Response-Shift auftreten könnten. Jedoch wird in der vorliegenden Arbeit negative Affektivität nicht unter dem Bewältigungsaspekt, sondern als Persönlichkeitsmerkmal untersucht. Von daher ist in Frage zu stellen, ob ein höheres Ausmaß an negativer Affektivität als stabiles Persönlichkeitsmerkmal änderungsorientierte Kognitionen hervorruft, die zu einer Verschiebung des Bewertungshintergrunds der GLQ beitragen. Die Ergebnisse zur Untersuchung von Benefit Finding weisen darauf hin, dass intrusive Gedanken ein Zeichen des konflikthaften Erlebens der Umstrukturierung eigener Standards, Ziele und Werte sind. Diese intrusiven Gedanken, die im weiteren Verlauf zu einer Neuadjustierung persönlicher Konzepte führen, können zwar mit negativen Emotionen einhergehen, jedoch korreliert Benefit Finding negativ mit Depressivität (Helgeson et al. 2006). Krohne und Kollegen (1996) gehen davon aus, dass ein geringes Ausmaß an negativer Affektivität mit Ruhe und Gelassenheit einhergeht, wobei diese Gefühlskomponenten zwar nicht während der Auseinandersetzung mit einer Krebserkrankung oder einer Umstrukturierung persönlicher Werte, Ziele und Konzepte anzunehmen sind. Jedoch ist unter der Perspektive, diese Zustände als eher persönlichkeitsinhärent anzunehmen davon auszugehen, dass ein solches Erleben eine günstigere Basis für einen Response-Shift darstellt, als eine ausgeprägte negative Affektivität.

Abschließend ist zusammenzufassen, dass Zustände wie Ruhe und Gelassenheit, die für ein geringes Ausmaß an negativer Affektivität stehen, als Persönlichkeitsmerkmale eher förderlich für eine konstruktive Auseinandersetzung mit einer belastenden Lebenssituation ange-

sehen werden. In solchen Zuständen sollte es einer Person eher möglich sein, eine lösungsorientierte Perspektive zu entwickeln und einzunehmen. Ein hohes Ausmaß an negativer Affektivität kann eine konstruktive Auseinandersetzung mit einer lebensbedrohlichen Erkrankung erschweren. Der Zustand der Person wäre gekennzeichnet durch Ängstlichkeit, Nervosität und Bekümmern. Auch wenn anzunehmen ist, dass diese Gefühle im Zusammenhang mit intrusiven Gedanken auftreten, werden diese – wiederum im Sinne eines stabilen Persönlichkeitsmerkmals – nicht als response-shift-förderlich angesehen.

In der vorliegenden Arbeit wird negative Affektivität als Persönlichkeits- und nicht als Bewältigungsmerkmal im Zusammenhang mit Response-Shift untersucht. Daher wird angenommen, dass ein geringes Ausmaß an negativer Affektivität eher mit Response-Shift einhergeht als ein hohes Ausmaß.

## **B. Empirischer Teil**

## 6. Methoden

Im Folgenden werden Angaben zur Stichprobe, Datenerhebung und Modellbildung vorgestellt, wobei das am Ende des Kapitels abgeleitete Messmodell die Grundlage für die Response-Shift-Erfassung im Ergebniskapitel der vorliegenden Arbeit darstellt.

### 6.1 Allgemeines Vorgehen

Die *Klinik Bad Oexen – Klinik für onkologische Rehabilitation und Anschlussheilbehandlung* stellte für die vorliegende Arbeit Daten von Brustkrebspatientinnen und PCa-Patienten zur Verfügung. Diese Daten wurden im Rahmen klinikinterner Datensammlungen erfasst. Im Zeitraum von März bis August 2005 wurden die Patientinnen und Patienten, die zur Rehabilitation in der Klinik Bad Oexen behandelt wurden, zu zwei Messzeitpunkten mittels Fragebogenverfahren zu ihrer GLQ und weitere Variablen befragt. Der erste Erhebungszeitpunkt (MZP 1) war zu Beginn der Rehabilitation in der Klinik. Der zweite Erhebungszeitpunkt (MZP 2) lag drei Monate nach Beginn der Rehabilitation.

Aus projektbezogenen Gründen wurden einige Instrumente erst nach einer weiteren Planungsphase in die Erhebungen aufgenommen, so dass Daten zur Beantwortung der Fragestellung 3 der vorliegenden Arbeit nur von einem kleineren Teil der Gesamtstichprobe vorliegen. Auf diese Stichprobencharakteristika wird weiter unten genauer eingegangen.

Die Auswahl der Patientinnen und Patienten erfolgt ohne Eingrenzung. Die Verteilung der Fragebögen zum Messzeitpunkt 1 erfolgte über die Mitarbeiterinnen und Mitarbeiter der Klinik Bad Oexen, die diese mit anderen Unterlagen zu Rehabeginn an die Rehabilitandinnen und Rehabilitanden ausgaben. Die Befragung zum MZP 2 fand auf postalischem Wege statt. Allen Patientinnen und Patienten, von denen Daten zu Beginn der Rehabilitation vorlagen, wurden drei Monate nach Rehabeginn angeschrieben und um die nochmalige Beantwortung der entsprechenden Instrumente – sowie der zusätzlich aufgenommenen Fragebögen – gebeten. Der Rücklauf der Befragung ging zu beiden Messzeitpunkten an die Klinik Bad Oexen, in der die Daten in Excel-Dateien eingegeben und anonymisiert wurden. Eine Kontrolle der Dateneingabe fand in der Klinik statt. Die so kontrollierten und anonymisierten Daten wurden dem Autor der vorliegenden Arbeit zur Verfügung gestellt. Anschließend wurde eine nochmalige Plausibilitätskontrolle der Daten durchgeführt. Zu Beginn der Befragung wurden die Patientinnen und Patienten über ein Schreiben der Klinik sowie deren Mitarbeiter über die Be-

fragung informiert und aufgeklärt. Es wurde eine vollständig anonymisierte Auswertung der Daten zugesichert und die Freiwilligkeit an der Befragung betont. Alle Befragten hatten jederzeit die Möglichkeit, die Teilnahme an der Befragung ohne Angabe von Gründen abzulehnen.

### 6.1.1 Verwendung statistischer Programme

Die vorliegenden Daten wurden mit den Programmen SPSS 14.02 und AMOS 17 ausgewertet. Alle vorausgehenden Datenanalysen erfolgten mittels SPSS 14.02, die Berechnungen der Konfirmatorischen Faktorenanalysen (KFA) erfolgten mittels AMOS 17. Die Ersetzung fehlender Werte durch den EM-Algorithmus wird mittels des SPSS-kompatiblen Statistik-Programms NORM durchgeführt.

### 6.1.2 Bildung verschiedener Subgruppen und Ausreißeranalyse

Die ursprüngliche Excel-Datendatei zum MZP 1 enthielt Angaben zu 734 Personen, zum MZP 2 waren es 482 Personen. Gut zwei Drittel der Stichprobe zum MZP 1 waren männlich (PCa-Patienten) sowie etwa zwei Drittel waren Patientinnen und Patienten, die ihre Rehabilitation im Rahmen einer Anschlussheilbehandlung (AHB) in der Klinik durchführten. Im Rahmen des Heilverfahrens (HV) wurden überwiegend Brustkrebspatientinnen rehabilitiert. Die genauen Zahlen sind in Tabelle 16 wiedergegeben.

Tabelle 16: Allgemeine Angaben zu den Ausgangsdaten

Diagnosegruppe	Antragsverfahren	
	AHB	HV
C 50	106 (14,4)	195 (26,6)
C 61	328 (44,7)	105 (14,3)

Anmerkungen: AHB = Anschlussheilbehandlung, HV = Heilverfahren, C 50 = Brustkrebs, C 61 = Prostatakrebs, Angaben in Klammern (Prozent an der Gesamtstichprobe)

Die Diagnosegruppen verteilten sich nicht gleichmäßig über die beiden Antragsverfahren, so dass die größte homogene Subgruppe die der PCa-Patienten in der AHB waren. Aufgrund der Stichprobenhomogenität als Grundlage für die weiteren Auswertungsschritte wurde eine Eingrenzung auf diese Substichprobe für die Beantwortung der Fragestellungen vorgenommen. Somit wird auch eine Konfundierung der Variablen Diagnosegruppe und Geschlecht vermieden. Des Weiteren gibt es einen inhaltlichen Grund, die Stichprobe auf AHB-Patienten zu reduzieren: Nach response-shift-theoretischen Annahmen ist eine Veränderung

des Bewertungshintergrundes eines zentralen Konstrukts eher zu erwarten, je intensiver und je aktueller das stattgefundene Ereignis ist (Sprangers und Schwartz 2000).

Die Intensität der Auseinandersetzung mit der Diagnose Krebs und der Behandlung der Erkrankung mag für beide Diagnosegruppen vergleichbar sein. Die Dauer der Auseinandersetzung zum Befragungszeitpunkt ist jedoch bei den HV-Patienten in der Regel eine weitaus größere. Der genaue Zeitraum zwischen Diagnosestellung und anschließender Akutbehandlung bis zum Beginn der Rehabilitation kann aus den Daten für keine der Subgruppen angegeben werden. Jedoch ist der Beginn der Rehabilitation im Rahmen eines Heilverfahrens häufig erst mehrere Monate nach der Akutbehandlung. Die Anschlussheilbehandlung schließt in der Regel direkt an die Akutbehandlung an, so dass die Auseinandersetzung mit dem kritischen Ereignis Tumorbehandlung als zeitnaher eingeschätzt wird.

Von den 328 PCa-Patienten liegen zum Zeitpunkt der Katamnese 236 Prä-Post-Messungen vor. Für die verbleibenden 92 PCa-Patienten ist nicht eindeutig definierbar, ob es sich bei diesen um einen Dropout im klassischen Sinne handelt oder ob diese Personen noch nicht in die Katamnese-Erhebung aufgenommen wurden. Der Beginn der Katamnese-Erhebung wurde von der Klinik selbst festgelegt und konnte aus den zur Verfügung gestellten Daten nicht mehr eindeutig entnommen werden. Die Katamnese-Erhebung wurde zusätzlich zu einer in der Klinik standardmäßig durchgeführten Aufnahmeerhebung (Prämessung) durchgeführt. Wie bereits beschrieben, liegen zu diesen 92 Personen keine ausführlichen soziodemographischen Angaben vor, so dass diese nur marginal beschrieben werden können. Dennoch wurden diese 92 PCa-Patienten in die (heterogene) Stichprobe zur Modellableitung aufgenommen (s.u.).

Fehlende Werte wurden nach dem EM-Algorithmus ersetzt. Von den 236 vorliegenden Fällen wurden dennoch 8 Teilnehmer aufgrund unvollständiger Datensätze aus den weiteren Berechnungen ausgeschlossen. Diese Personen hatten in wenigstens einer zentralen Lebensqualitätsskala des EORTC-QLQ-C30 mehr als 50% Missings. Der Wert von 50% fehlenden Werten innerhalb einer Skala wurde festgelegt, da drei der fünf Skalen (Soziale, Kognitive und Rollenfunktionsfähigkeit) jeweils nur aus zwei Items bestehen. So dass bei einem nicht definierten Item die gesamte Skala nicht berechnet und die Person gänzlich aus der Analyse ausgeschlossen wurde. Somit wurden in die Auswertungsstichprobe 228 PCa-Patienten aufgenommen, zu denen zu beiden Messzeitpunkten ausreichend Daten vorlagen.



Die Stichprobe der 228 PCa-Patienten wurde einer Ausreißeranalyse unterzogen. Insgesamt wurden 16 Ausreißer identifiziert und aus der weiteren Berechnung ausgeschlossen. Die endgültige Analysestichprobe zur Beantwortung der ersten Fragestellung umfasst somit 212 PCa-Patienten.

In einem ersten Auswertungsschritt sollte das aus der Literatur entnommene Modell der GLQ an einer zur Beantwortung der Hauptfragestellungen unabhängigen Stichprobe bestätigt werden (Boehmer & Luszczynska 2006). Um auf eine akzeptable Stichprobengröße zur Modelltestung zurückgreifen zu können, wurde eine heterogene Stichprobe gebildet. Diese setzte sich aus 92 PCa-Patienten, die lediglich zu Beginn der AHB den EORTC-QLQ-C30 ausgefüllt hatten, und den 106 Brustkrebspatientinnen (Messzeitpunkt 1) in der AHB zusammen. Insgesamt 11 Personen (sieben Brustkrebspatientinnen, vier Prostatakrebspatienten) wurden aufgrund unvollständiger Datensätze ( $\geq 50\%$  fehlender Werte in wenigstens einer der Funktionsskalen des EORTC-QLQ-C30) von den weiteren Analysen ausgeschlossen. Nach Ausreißer-Analysen mittels Boxplots (SPSS 14.02) wurden zwei Personen (je ein PCa-Patient sowie eine Brustkrebspatientin) als Ausreißer identifiziert und ebenfalls von den weiteren Auswertungen aus dem Datensatz entfernt. Somit setzt sich die endgültige Stichprobe zur Modellableitung aus 98 Brustkrebspatientinnen und 87 Prostatakrebspatienten zusammen. Tabelle 17 bietet einen Überblick über die Zusammensetzung der Stichproben.

Tabelle 17: Stichprobenzusammensetzung

Diagnosegruppe <sup>1</sup>	Stichprobe AHB-Patienten			
	Modellableitung	Dropout	Zentrale Auswertungsstichprobe	Dropout
C 50	98	8	---	---
C 61	87	5	212	24
Insgesamt	185	13	212	24

Anmerkungen: <sup>1</sup>entspricht auch dem Geschlecht (C 50 = Brustkrebs, C61 = Prostatakrebs)

### 6.1.3 Beschreibung der Stichprobe zur Modellableitung

Eine Beschreibung der 185 Personen der Stichprobe zur Validierung des Modells der GLQ nach Boehmer und Luszczynska (2006) erfolgt lediglich anhand weniger Parameter. Da aus Gründen der Datenerhebung für die meisten Personen dieser Gruppe keine deskriptiven Angaben vorliegen, werden hier lediglich die Variable Alter sowie die Mittelwerte der Skalen, die zur Definition des latenten Konstrukts GLQ verwendet werden, vorgestellt. Die PCa-

Patienten waren im Durchschnitt etwas älter und in der Altersverteilung homogener im Vergleich zu den Brustkrebspatienten. In den Skalen Emotionale und Kognitive Funktionsfähigkeit weisen die PCa-Patienten statistisch signifikant höhere Mittelwerte auf als die Brustkrebspatientinnen. Die Angaben werden in Tabelle 18 dargestellt.

Tabelle 18: Mittelwerte und Standardabweichungen (in Klammern) von Alter und Lebensqualitätsskalen – Modellableitung

Diagnosegruppe <sup>1</sup>	n	Alter**	SF	RF	PF	EF**	CF**
C 50	98	58,6 (15,7)	57,4 (31,8)	47,3 (29,4)	67,4 (18,3)	50,3 (28,1)	67,9 (29,7)
C 61	87	64,3 (6,3)	64,6 (31,4)	40,0 (33,0)	72,9 (18,5)	66,6 (26,7)	80,8 (23,1)
Insgesamt	185	61,3 (12,5)	60,8 (31,7)	43,9 (31,3)	70,0 (18,6)	58,0 (28,6)	74,0 (27,5)

Anmerkungen: <sup>1</sup>entspricht auch dem Geschlecht (C 50 = Brustkrebs, C61 = Prostatakrebs), n = Anzahl, SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit; \*\* = p < .01

#### 6.1.4 Beschreibung der Stichprobe zur Beantwortung der Hauptfragestellung

Insgesamt liegen von 228 Personen Daten zu zwei Messzeitpunkten vor. Ausreißeranalysen mittels Boxplots erbrachten eine Selektion von 16 Studienteilnehmern, die zu einem oder zu beiden Messzeitpunkten auf wenigstens einer modellbildenden Skala des EORTC-QLQ-C30 durch Extremwerte auffällig waren. Die genauen Angaben zur Gesamtstichprobe der verbleibenden 212 PCa-Patienten in der Anschlussheilbehandlung werden in Tabelle 19 dargestellt.

Bei den Personen der Gesamtstichprobe handelt es sich ausschließlich um Patienten mit Prostatakrebs in der AHB. Das Durchschnittsalter lag bei 66,1 Jahren mit einer Streuung von 5,7 Jahren und einem Range von 40 Jahren (Min = 43, Max = 83). Die überwiegende Mehrheit war verheiratet, deutscher Nationalität und lebt gemeinsam mit ihrer Partnerin in einem Haushalt. Etwa 2/3 hatten 2 und mehr Kinder. Ebenso hoch ist der Anteil derjenigen mit Volksschulabschluss, wobei ca. die Hälfte der Stichprobe eine Lehre abgeschlossen hatte. Gut ¾ der befragten Personen war mittlerweile nicht mehr erwerbstätig und in Altersrente. Drei Personen gaben an, dass drei Monate nach Beginn der Rehabilitation ein erneuter Tumor aufgetreten sei. Obwohl anzunehmen ist, dass die Auseinandersetzung mit der Erkrankung bei

diesen Personen kritischer verlaufen kann, werden aus Gründen einer möglichst großen Teilnehmerzahl die derart Betroffenen in die weiteren Analysen mit aufgenommen.

Tabelle 19: Soziodemographische Angaben zur Gesamtstichprobe (N = 212)

Familienstand				Berufsausbildung				
	n	(%)		n	(%)		(%)	
Ledig	3	(1,4)	Lehre	112	(52,8)			
Verheiratet	187	(88,2)	Fachschule	45	(21,2)			
Geschieden	8	(3,8)	Fachhochschule	18	(8,5)			
Verwitwet	14	(6,6)	Universität	10	(4,7)			
Missing	0	(0,0)	Andere Berufsausbildung	15	(7,1)			
Nationalität				Keine Berufsausbildung				
Deutsch	208	(98,1)	Missing	5	(2,4)			
Andere	3	(1,4)	Erwerbstätigkeit					
Missing	1	(0,4)	Ja, ganztags	23	(10,8)			
Personen im Haushalt				Ja, halbtags	6	(2,8)		
Partner	160	(75,5)	Nein, Hausfrau/Hausmann	3	(1,4)			
Kinder	3	(1,4)	Nein, arbeitslos/erwerbslos	7	(3,3)			
Allein	22	(10,4)	Nein, BU, EU	5	(2,4)			
Partner und Kinder	25	(11,8)	Nein, Altersrente	160	(75,5)			
Missing	2	(0,9)	Nein, anders	7	(3,3)			
Anzahl der Kinder				Missing	1	(0,4)		
Kein Kind	14	(6,6)	Weitere Therapie nach der Rehabilitation					
1 Kind	53	(25,0)	Nein, keine	180	(84,9)			
2 Kinder	98	(46,2)	Operation	2	(0,9)			
Mehr als 2 Kinder	47	(22,2)	Chemotherapie	1	(0,4)			
Missing	0	(0,0)	Bestrahlung	17	(8,0)			
Schulabschluss				Hormontherapie	6	(2,8)		
Hauptschule/Volksschule	132	(62,3)	Anderweitige Therapie	2	(0,9)			
Realschule	35	(16,5)	Missing	4	(1,9)			
Polytechnische Oberschule	1	(0,4)	Erneutes Auftreten eine Tumors nach der AHB					
Fachhochschulreife	22	(10,4)	Nein.	198	(93,4)			
Abitur	17	(8,0)	Ja	3	(1,4)			
Anderer Schulabschluss	3	(1,4)	Verdacht auf	0	(0,0)			
Kein Schulabschluss	1	(0,4)	Weiß nicht	2	(0,9)			
Missing	1	(0,4)						

Anmerkung: n = Anzahl der Personen, (%) = prozentualer Anteil an der Gesamtstichprobe

Bei ca. 85 % der Befragten wurden nach Abschluss der AHB keine weiteren therapeutischen Maßnahmen durchgeführt. Die häufigste Form der Weiterbehandlung bestand in der Bestrahlung (8 %). Angaben zum TNM-Status der Prostatakreberkrankung waren den Daten nicht zu entnehmen und konnten im Nachhinein nicht mehr recherchiert werden.

Die Erfassung der soziodemographischen Angaben der Rehabilitanden fand erst zum MZP 2 statt. Daher wird die Dropout-Analyse nur mit den 16 Ausreißern sowie den 8 PCa-Patienten mit unvollständigen Datensätzen durchgeführt. Die Auswertungen erbrachten folgendes Ergebnis: In den Variablen *Alter*, *Personen im Haushalt*, *Schulabschluss*, *Berufsausbildung*, *Erwerbstätigkeit* und *Anzahl der Kinder* zeigten sich keine statistisch bedeutsamen Unterschiede zwischen den Personen der Auswertungsstichprobe und dem Dropout. Dies gilt ebenso bezüglich der Variable *weitere Therapie nach der Rehabilitation*. Jedoch fand sich ein statistisch signifikanter Unterschied gemessen am  $\chi^2$ -Wert in der Variable *erneutes Auftreten eines Tumors nach der Rehabilitation*, wobei in der Dropoutgruppe häufiger die Antwortkategorie „Verdacht auf“ angegeben wurde. In dieser Gruppe berichtet eine von 24 Personen von einem erneuten Auftreten eines Tumors, wohingegen zwei weitere Personen einen entsprechenden Verdacht äußern.

#### 6.1.5 Ersetzen fehlender Werte

Im nächsten Schritt wurden in der Modelleitungsstichprobe sowie in der zentralen Auswertungsstichprobe fehlende Werte ersetzt. Verschiedene Methoden stehen hierfür zu Verfügung. Einerseits besteht die Möglichkeit der paarweisen (pairwise deletion) oder listenweisen (listwise deletion) Löschung von Fällen mit fehlenden Werten oder die Ersetzung fehlender Werte durch den Mittelwert aller beobachteten Werte der jeweiligen Skala (mean imputation, MI). Brown (1983) sowie Little und Rubin (1987) konnten jedoch zeigen, dass diese Verfahren zu unangemessenen Schätzungen führen. Nach Enders (2001) existieren verschiedene Maximum Likelihood (ML-) Algorithmen, die zu robusteren Schätzern führen und den zuerst genannten Ansätzen überlegen sind (Arbuckle 1996, Enders & Bandalos 2001). Für die vorliegende Arbeit wurde der Expectation maximum (EM-) Algorithmus gewählt, bei dem es sich um einen indirekten Ansatz handelt, da vor der Berechnung von Parameterschätzern verschiedene Analysen durchgeführt werden. Der EM-Algorithmus liefert einen Schätzer der Kovarianz-Matrix und des Mittelwertvektors, der wiederum für weitere Modellierungen verwendet werden kann. Enders und Peugh (2004) kommen zu dem Schluss, dass die durch

einen EM-Algorithmus geschätzte Kovarianzmatrix zu keiner Verzerrung der Standardfehler führt.

## 6.2 Beschreibung der Messinstrumente

Im Mittelpunkt der vorliegenden Arbeit steht das Instrument *European Organization for Research and Treatment of Cancer – Quality of Life Questionnaire – Core 30* (EORTC-QLQ-C30) der gleichnamigen belgischen Institution EORTC, die diesen Fragebogen entwickelt hat (Fayers et al. 1995). Ausgangspunkt der Verfahrensentwicklung war das Ziel, ein modular aufgebautes Instrumentenpaket zur Erfassung der Lebensqualität von Krebspatienten für die Durchführung klinischer Studien und die internationale Vergleichbarkeit der Studienergebnisse zu erstellen (Aaronson et al. 1993). Die Anwendung des EORTC-QLQ-C30 soll der Erfassung der GLQ bei verschiedenen Krebserkrankungen dienen. Der EORTC-QLQ-C30 soll bei krankheitsspezifischen Studien durch entsprechend spezifische Module ergänzt werden. Die Kombination aus EORTC-QLQ-C30 und einem krankheitsspezifischen Modul erlaubt eine genaue Analyse der GLQ sowie der krankheitsspezifischen Auswirkungen auf das Erleben und Verhalten der betroffenen Personen sowie die Erfassung behandlungsspezifischer Nebenwirkungen.

Der Fragebogen besteht zum Einen aus neun Skalen, die sich aus insgesamt 24 Items zusammensetzen. Zu diesen neun Skalen zählen fünf funktionale Skalen, eine Skala zur Erfassung der Globalen Lebensqualität und drei Symptomskalen. Zum Anderen sind in dem Fragebogen sechs Einzelitems zur Erfassung verschiedener Behandlungsnebenwirkungen enthalten. Tabelle 20 gibt einen Überblick über die Inhalte der Funktions- und Symptomskalen und zeigt zudem die in verschiedenen Studien angegebenen Werte für Cronbachs Alpha einerseits für heterogene Stichproben und andererseits für PCa-Patienten. Die Einzelitems sind in dieser Darstellung nicht enthalten, da Angaben für Cronbachs Alpha nicht berechenbar sind. Inhaltlich wird jeweils die Ausprägung folgender Symptome beziehungsweise Nebenwirkungen abgefragt: Kurzatmigkeit, Schlafstörungen, Appetitmangel, Verstopfung, Durchfall, finanzielle Schwierigkeiten (Fayers et al. 1995).

Tabelle 20: Angaben zu den Skalen des EORTC-QLQ-C30 und verschiedenen Cronbachs Alpha-Werten

Skalen	Items	Range	Inhaltliche Beschreibung der geringsten und höchsten Ausprägung		Cronbachs Alpha*				
			Geringster Wert = 0	höchster Wert = 100	KS <sup>1</sup>	PCa <sup>2</sup>	OR <sup>3</sup>	AB <sup>4</sup>	
Funktionsskalen	PF	5	1-4	Sehr viele Schwierigkeiten bei körperlicher Aktivität, sehr immobil und hilfebedürftig	Überhaupt keine Schwierigkeiten bei körperlicher Aktivität, hohe Mobilität und Selbstständigkeit	.68	.64	.57	.81
	RF	2	1-4	Sehr hohe Einschränkungen in Alltagsaktivitäten in Haushalt und Beruf	In keiner Weise eingeschränkt, weder bei häuslichen, noch bei beruflichen Aktivitäten	.54	.62	.36	.89
	EF	4	1-4	Sehr hohes Ausmaß an Ängstlichkeit und Depressivität	Überhaupt nicht angespannt, reizbar, besorgt oder niedergeschlagen	.73	.85	.72	.80
	CF	2	1-4	Sehr große Schwierigkeiten, sich zu konzentrieren oder sich an Dinge zu erinnern	Keinerlei Schwierigkeiten mit der Konzentration und dem Gedächtnis	.56	.55	.85	.68
	SF	2	1-4	Aufgrund des körperlichen Zustands und der medizinischen Behandlung sind Aktivitäten mit der Familie bzw. dem sozialen Umfeld sehr stark eingeschränkt	Keine Einschränkungen der familiären und weiteren sozialen Aktivitäten aufgrund der Erkrankung und deren Behandlung	.68	.74	.78	.86
Globale Lebensqualität	2	1-7	Allgemeiner körperlicher Zustand und allgemeine Lebensqualität ist sehr schlecht	Allgemeiner körperlicher Zustand und allgemeine Lebensqualität ist ausgezeichnet	.86	.88	.90	.89	
Übelkeit und Erbrechen	2	1-4	Keine Beschwerden	Beides sehr häufig	.65	.64	.75	.65	
Schmerzen	2	1-4	Keine Schmerzen	Sehr häufig	.82	.86	.89	.84	
Fatigue	3	1-4	Kein Gefühl der Müdigkeit und Erschöpfung	Sehr starkes Gefühl von Müdigkeit und Erschöpfung	.80	.89	.86	.80	

Anmerkungen PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, KS = Krebspezifisch, PCa = Prostatakarzinom, OR = Onkologische Rehabilitation, AB = bundesdeutsche Allgemeinbevölkerung, \* = alle vorliegenden Werte für Cronbachs Alpha wurden für die Skalen PF und RF mit einem Range von 0 bis 1 berechnet, da in der vorherigen Version des EORTC-QLQ-C30 diese Items auf einer dichotomen Antwortskala beantwortet werden sollten, <sup>1</sup> = Aaronson et al. 1993, <sup>2</sup> = Borghede & Sullivan 1996, <sup>3</sup> = Dirmaier et al. 2004, <sup>4</sup> = Schwarz & Hinz (2001)

In der weiteren Darstellung wird ausschließlich auf die Funktionsskalen Bezug genommen, da nur diese in der vorliegenden Arbeit ausgewertet wurden. Diese Entscheidung wurde vor dem theoretischen Hintergrund getroffen, da nach Fayers und Hand (1997, 2002) sowie Boehmer und Luszczynska (2006) diese fünf Skalen als Indikatoren für GLQ anzusehen sind. Die Symptomskalen und Einzelitems hingegen werden als Kausalvariablen beschrieben, deren Beziehung zum Konstrukt Lebensqualität nicht ‚korrelativ‘ ist. Im Zentrum der vorliegenden Arbeit steht die Frage nach der Veränderung der Lebensqualität und der sie kennzeichnenden Dimensionen, so dass auf die Symptome und Beschwerden verzichtet wurde. Die Skala Globale Lebensqualität wird in der vorliegenden Arbeit nur für die Beantwortung der Fragestellung 2 verwendet, da ein potenzieller Response-Shift mit Hilfe latenter Variablen untersucht werden soll, so dass die manifeste Variable Lebensqualität des Fragebogens nicht mit den Lebensqualitätsdimensionen in Beziehung gesetzt wird.

Die Berechnung der Skalenwerte erfolgt durch eine Transformierung der Mittelwerte der im Fragebogen vorgegebenen vierstufigen Likertskala auf eine Skala von 0 bis 100, wobei für die Funktionsskalen und die Skala zur Globalen Lebensqualität gilt, dass höhere Werte eine positivere Ausprägung anzeigen. Für die Symptomskalen und die Einzelitems gilt das Gegenteil, je niedriger der Wert, desto weniger negativ ist die Bewertung.

Die internen Konsistenzen der Skalen wurden mittels Cronbachs Alpha berechnet und sind in der Tabelle 20 für verschiedene Populationen angegeben, da die Belastungen durch eine Krebserkrankung je nach Art und Stadium der Erkrankung sowie deren Behandlungsstrategie sehr verschieden sein kann. Anhand der Werte für Cronbachs Alpha aus der Literatur fällt auf, dass die Werte für die Skalen in einem eher kritischen Bereich liegen. Ringdal und Ringdal (1993) berichten von einer Zusammenfassung der Skalen Physische und Rollenfunktionsfähigkeit, was zu einer Erhöhung des Cronbachs Alpha für die neu entstandene Skala „Personale Funktionsfähigkeit“ auf .82 führt. Dirmaier und Kollegen (2004) postulieren aufgrund faktorenanalytischer Untersuchungen eine gemeinsame Skala der Physischen und Rollenfunktionsfähigkeit und gelangen so zu einem besseren Cronbachs Alpha Wert (.66). Kriskke (1995) bestätigt die Befunde zu den Skalen Physische - und Rollenfunktionsfähigkeit, die in der dargestellten Studie beide unter Cronbachs Alpha .70 bei einer heterogenen Stichprobe von Rehabilitanden liegen. Osoba und Kollegen (1994) erreichten eine verbesserte Interne Konsistenz der Skala Physische Funktionsfähigkeit durch eine Kombination von drei Items der Ursprungsskala sowie einem Item der Skala Rollenfunktionsfähigkeit (Cronbachs

Alpha .71). Die Kombination der übrigen Items der beiden Skalen zur Skala Rollenfunktionsfähigkeit erbrachte einen Wert für Cronbachs Alpha von .66. Jedoch schwankten diese Werte zu den verschiedenen Messzeitpunkten. Aufgrund der unklaren Ergebnisse der bisher vorliegenden Studien und der von der EORTC neu gestalteten Antwortskalierungen für die Skalen Physische und Rollenfunktionsfähigkeit (mit darin enthaltenen minimalen Änderungen der Itemformulierung) wird auf die von der EORTC (Fayers et al. 1995) vorgegebenen Skalenzusammensetzung zurückgegriffen.

Die Beantwortung und Auswertung der zweiten Fragestellung erfolgt mit Hilfe der Skala *Globale Lebensqualität* des EORTC-QLQ-C30. Diese Skala setzt sich aus zwei Items zusammen, die zum Einen eine Einschätzung des eigenen Gesundheitszustandes und zum Anderen der Globalen Lebensqualität in den vergangenen 7 Tagen erfassen sollen. Als Antwortformat wird eine 7-stufige Lickertskala mit den Polen ‚sehr schlecht‘ bis ‚ausgezeichnet‘ verwendet. Die interne Konsistenz der Skala liegt für PCa-Patienten bei Cronbachs Alpha von .88 (siehe Tabelle 20).

Zur Beantwortung der dritten Fragestellung werden verschiedene Messinstrumente zur Operationalisierung der Variablen *positive Affektivität*, *negative Affektivität*, *Selbstwirksamkeit* und *Benefit Finding* eingesetzt, die im Folgenden kurz beschrieben werden.

*Positive and Negative Affect Schedule* (PANAS, Watson et al. 1988a, deutsche Übersetzung von Krohne et al. 1996). Die PANAS ist ein Selbstbeurteilungsinstrument zur Erfassung der emotionalen Befindlichkeit, das aus insgesamt 20 Adjektiven besteht. Das Instrument setzt sich aus zwei Skalen bestehend aus jeweils 10 Adjektiven zusammen, wobei eine Skala positive (beispielsweise aktiv, freudig erregt, stolz) und die zweite Skala negative Adjektive (beispielsweise verärgert, bekümmert, nervös) umfasst. Grundlage der Skalenbildung ist die Annahme, dass sich Affektivität auf zwei unabhängigen Dimensionen zeigt, die als positiver und negativer Affekt bezeichnet werden. Eine geringe Ausprägung der positiven Affektivität geht somit nicht einher mit einer gleichzeitigen hohen Ausprägung der negativen Affektivität. Die Ausprägung zu den einzelnen Adjektiven werden auf einer fünfstufigen Skala von *gar nicht – ein bißchen – einigermaßen – erheblich – äußerst* angegeben. In der vorliegenden Arbeit wurde die habituelle Ausprägung der positiven und negativen Affektivität erfasst. Es wurde nach der Ausprägung der einzelnen Variablen *im Allgemeinen* gefragt. Die Skala Positive Affektivität weist einen Wert für Cronbachs Alpha von .84 auf, die Skala Negative Af-



fektivität von .86 (Krohne et al. 1996). Die Werte der internen Konsistenz liegen somit in einem guten Bereich. Grundlage des Fragebogens ist die Einteilung von emotionalen Zuständen in positive und negative Erfahrungsqualitäten. Auf eine weitere Differenzierung (bspw. nach Basisemotionen) wird bewusst verzichtet. Krohne und Kollegen (1996) wiesen anhand der deutschen Version der PANAS nach, dass die dispositionelle Beurteilung der positiven und negativen Affektivität (‚im Allgemeinen‘) nicht unabhängig ist vom Erleben und von Befindlichkeiten in den vergangenen Wochen.

*Allgemeine Selbstwirksamkeitserwartung* (Schwarzer & Jerusalem 1999). Das Konzept erfasst die persönliche Einschätzung der eigenen Kompetenz, allgemein mit Schwierigkeiten und Barrieren im täglichen Leben zurechtzukommen. Solche Kompetenzerwartungen stellen eine wichtige personale Ressource dar, um schwierige Dinge im Alltag zu bewältigen. Die Skala setzt sich aus 10 Items zusammen, die auf einer vierstufigen Antwortskala beurteilt werden. Das Antwortformat ist unterteilt in *stimmt nicht* – *stimmt kaum* – *stimmt eher* – *stimmt genau*. Beispielitems lauten: ‚Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen‘, ‚Die Lösung schwieriger Probleme gelingt mir immer, wenn ich mich darum bemühe‘. Cronbachs Alpha liegt bei .90 für die Allgemeinbevölkerung. Luszczynska und Kollegen (2005) geben für eine Stichprobe von 238 Krebspatienten einen Wert für Cronbachs Alpha von .89 an.

*Benefit Finding Skala* (Mohamed & Böhmer 2004). Das Konzept bezeichnet einen Prozess, bei dem Individuen nach den positiven Folgen eines negativen Ereignisses suchen oder nach guten Seiten von schwierigen Ereignissen (Schwarzer & Knoll 2003). Die Benefit Finding Skala existiert als Langform mit 17 Items und als Kurzform mit 10 Items. In der vorliegenden Arbeit kam die Kurzversion der Skala zur Anwendung. Das Antwortformat ist fünfstufig von *gar nicht* – *kaum* – *mittelmäßig* – *ziemlich* – *sehr*. Beispielitems der Skala lauten ‚Meine Erkrankung hat mich gelehrt, mich an Umstände anzupassen, die ich nicht ändern kann‘, ‚Meine Erkrankung hat mir geholfen, die Dinge so zu nehmen, wie sie sind‘. Die Werte für Cronbachs Alpha in der amerikanischen Originalversion von Antoni und Kollegen (2001) liegen bei .95. Mohamed und Böhmer (2004) geben für eine deutsche Stichprobe von Krebspatienten ein Cronbachs Alpha von .90 an.

*Soziodemographische Daten* (*Soz-Dat*, Deck & Röckelein 1999). Der Fragebogen erfasst verschiedene zentrale soziodemographische Daten wie beispielsweise Geschlecht, Staatsan-

gehörigkeit, Familienstand, Schulabschluss und Berufsausbildung. Das Instrument wurde von der „AG Routinedaten“ des Förderschwerpunkts Rehabilitationswissenschaften des Verbands Deutscher Rentenversicherungsträger (VDR) als Vorschlag zur Homogenisierung deskriptiver Informationen zur Soziodemographie entworfen. Die Frage zum monatlichen Nettoeinkommen des Haushalts wurde nicht in die Erfassung aufgenommen. Ergänzend wurden Fragen zu nachfolgenden Therapien und Behandlungen sowie dem Wiederauftreten eines Tumors integriert.

### 6.2.1 Prüfung auf Normalverteilung

Eine Voraussetzung für die Analyse von Strukturgleichungsmodellen ist die multivariate Normalverteilung der Daten. Bei der Verwendung der Maximum-Likelihood-Schätzmethode als Schätzverfahren können nicht normalverteilte Daten zu Überschätzungen des  $\chi^2$ -Wertes führen. Nach West und Kollegen (1995) ist davon auszugehen, dass bei Stichprobenumfängen zwischen 200 bis 500 Personen und einer Schiefe  $\leq 2$  und einer Kurtosis  $\leq 7$  die Maximum-Likelihood-Schätzung zu keinen abweichenden Ergebnissen führt. Kline (2005) setzt die Kriterien noch etwas lockerer an und bezeichnet eine Schiefe  $> 3$  als nicht akzeptabel sowie eine Kurtosis  $> 10$  als problematisch. Zudem erklären Hoogland und Boosma (1998) die Maximum-Likelihood-Methode als robust gegenüber kleineren Verletzungen der Normalverteilungsvoraussetzungen. In Tabelle 21 sind die Angaben zu Schiefe und Kurtosis für beide Stichproben enthalten.

Tabelle 21: Schiefe und Kurtosis in den zentralen Modellvariablen für die beiden Stichproben

	Variablen	Modellableitung (N = 185)		Auswertungsstichprobe (N = 212)	
		Schiefe	Kurtosis	Schiefe	Kurtosis
MZIP 1	PF	-0.28	-0.57	-0.35	-0.71
	RF	0.25	-0.87	0.30	-1.22
	EF	-0.27	-0.86	-0.47	-0.42
	CF	-0.92	0.12	-1.04	0.20
	SF	-0.26	-1.10	-0.35	-0.74
MZIP 2	PF	---	---	-0.79	-0.05
	RF	---	---	-0.69	0.12
	EF	---	---	-0.78	0.13
	CF	---	---	-0.84	-0.26
	SF	---	---	-0.78	0.11

Anmerkungen: MZIP = Messzeitpunkt, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit

Entsprechend den oben genannten Kriterien ist von annähernd normalverteilten Daten auszugehen. Zur weiteren Anpassung der Variablen an die Voraussetzungen einer Normalverteilung wurden sämtliche Berechnungen zusätzlich mit Hilfe des Bootstrapping-Verfahrens aufbereitet. Dabei wird eine definierte Anzahl an Zufallsstichproben ‚mit Zurücklegen‘ aus der vorliegenden Stichprobe gezogen (beispielsweise 400 Bootstraps, bedeutet 400 Zufallsstichproben aus den vorliegenden Daten). Für jede dieser künstlich generierten Stichproben werden die einzelnen Kennwerte berechnet. Nach Durchlaufen der vorgegeben Anzahl an Bootstraps werden (im obigen Beispiel aus den 400 zufällig gezogenen Stichproben) die Populationsschätzer bestimmt (Efron & Tibshiran 1993). Die derart erhaltenen Populationsschätzer entsprechen den Grundwerten einer Normalverteilung noch deutlicher. Die Stichprobengröße für Bootstrapping soll bei  $N \geq 200$  liegen (Byrne 2001).

### 6.2.2 Skalenüberprüfung

Die Überprüfung der verwendeten Skalen erfolgte in Form der Berechnung der internen Konsistenz (Cronbachs Alpha) für alle Skalen, die für die Fragestellungen relevant sind. Im Folgenden wird ein Überblick über die Werte von Cronbachs Alpha für die verwendeten Skalen des EORTC-QLQ-C30 in der vorliegenden Arbeit dargestellt. Die Berechnungen wurden für beide Teilstichproben, der zur Modellableitung und der zentralen Auswertungsstichprobe (zu beiden Messzeitpunkten) durchgeführt. Die Anzahl der Items pro Skala variiert zwischen fünf (Skala Physische Funktionsfähigkeit) und zwei (beispielsweise Kognitive Funktionsfähigkeit). Tabelle 22 zeigt alle Werte.

Tabelle 22: Cronbachs Alpha für die Indikatoren der GLQ des EORTC-QLQ-C30

Variablen	Modellableitung (N = 185)	Auswertungsstichprobe (N = 212)	
		MZP 1	MZP 2
PF	.75	.72	.65
RF	.85	.89	.83
EF	.87	.82	.82
CF	.76	.59	.69
SF	.79	.71	.74

Anmerkungen: MZP = Messzeitpunkt, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit

Die Werte der Skalen unterscheiden sich zum Teil recht deutlich. Die Skalen Rollenfunktionsfähigkeit und Emotionale Funktionsfähigkeit zeigen über alle Teilstichproben gute

Kennwerte. Die Werte der Skala Soziale Funktionsfähigkeit sind insgesamt in einem zufrieden stellenden Bereich. Die Skalen Physische Funktionsfähigkeit und Kognitive Funktionsfähigkeit weisen in jeweils einer Teilstichprobe einen kritischen Wert auf. Bei der Skala Kognitive Funktionsfähigkeit handelt es sich um eine Skala, die sich aus zwei Items zusammensetzt, so dass der Wert von Cronbachs Alpha vom Messzeitpunkt 1 der Auswertungsstichprobe von .59 weniger kritisch zu sehen ist als der Wert .65 der Skala Physische Funktionsfähigkeit (fünf Items). Nach Bortz und Döring (1995) wird ein Cronbachs Alpha für stabile Merkmale von  $>.80$  als mittelmäßig bis hoch eingeschätzt. Insgesamt wird die interne Konsistenz der zentralen Skalen des EORTC-QLQ-C30 als ausreichend bis gut bewertet.

In einer weiteren Auswertung wurden die Werte für Cronbachs Alpha für die Instrumente zur Beantwortung der Fragestellungen 2 und 3 berechnet. Mit Ausnahme der Skala *Allgemeine Lebensqualität* wurden die Fragebögen nur zum zweiten Messzeitpunkt eingesetzt, so dass sich die Angaben nur auf diesen Zeitpunkt beziehen. Die Skala Allgemeine Lebensqualität setzt sich aus zwei Items zusammen. Die Skalen Benefit Finding, Selbstwirksamkeit, Positive und Negative Affektivität bestehen aus jeweils 10 Items. Insgesamt zeigt sich für alle Skalen ein guter Kennwert für Cronbachs Alpha. Die Angaben dazu sind in Tabelle 23 zu finden.

Tabelle 23: Cronbachs Alpha für weitere Skalen

Variablen	MZP 2
EORTC-QLQ-Globale LQ	.89
BF	.89
SW	.90
PANAS – POS	.87
PANAS – NEG	.84

Anmerkungen: MZP = Messzeitpunkt, EORTC-QLQ-Globale LQ = Skala der Globalen Lebensqualität des EORTC-QLQ-C30; BF = Skala ‚Benefit Finding‘, SW = Skala ‚Selbstwirksamkeit‘, PANAS – POS = Skala ‚Positive Affektivität‘ der PANAS, PANAS – NEG = Skala ‚Negative Affektivität‘ der PANAS

### 6.3 Vorgehensweise der Modellprüfung

In der Terminologie der Strukturgleichungsmodelle (Structural Equation Model, SEM) stellt das in der vorliegenden Arbeit berücksichtigte Modell ein so genanntes Messmodell dar (Backhaus et al. 2003). Das Ziel von SEM und Konfirmatorischer Faktorenanalyse (KFA) ist die Überprüfung, ob eine empirisch ermittelte Kovarianzmatrix mit einer theoretisch postu-

lierten Kovarianzmatrix übereinstimmt. Die theoretisch postulierte Kovarianzmatrix resultiert aus der Modellbildung und Modellableitung, die nach sachlichen und inhaltlichen Aspekten im ersten Schritt *vor* der eigentlichen Modellüberprüfung vorgenommen wird. Die Beziehungen, die aufgrund der theoretischen Annahmen zwischen den Modellvariablen bestehen (sollten), werden in einem zweiten Schritt anhand einer empirischen Datenmatrix, die diese Variablen enthält, auf ihr tatsächliches Auftreten und ihr Ausmaß überprüft. Die Passung zwischen theoretisch postulierter und empirisch ermittelter Modellstruktur wird anhand verschiedener Fit-Indizes überprüft. Bevor die in dieser Arbeit verwendeten Indizes zur Überprüfung der Modellpassung beschrieben werden, sollen vorab einige Grundlagen zur Vorgehensweise der Modellableitung dargestellt werden.

### 6.3.1 Allgemeine Voraussetzungen zur Ableitung eines Messmodells

Mit Hilfe der KFA lässt sich die Güte eines Messmodells überprüfen. Ein Messmodell muss nach Bollen (1989) vier Voraussetzungen genügen:

1. Definition eines theoretischen Konzepts: Bedeutsam sind inhaltliche Aspekte, um das Konzept gut und präzise zu beschreiben und zu erfassen, sowie eine mögliche Ordnung der relevanten inhaltlichen Aspekte, um ein Konzept auch in seiner hierarchischen Gliederung darzustellen;
2. Formulierung der latenten Variable(n): Jeder inhaltlich relevante Aspekt wird in einem Messmodell anhand einer latenten Variable definiert, die durch mehrere Indikatoren repräsentiert wird;
3. Formulierung der manifesten Variablen: Diese stellen die tatsächlich erhobenen Indikatoren der latenten Konstrukte dar, die empirisch erfasst werden;
4. Formulierung der Beziehungen zwischen latenten und manifesten Variablen: Ein Messmodell wird erst durch Messgleichungen formal spezifiziert, die Aussagen über die Bedeutung der latenten und manifesten Variablen und von Messfehlern zulassen.

Die Testung der Parameter dieser Messgleichungen bietet die Möglichkeit, die inhaltlich postulierten Beziehungen zwischen den Variablen zu überprüfen. Um jedoch die einzelnen Modellparameter zu identifizieren, müssen zwei Voraussetzungen gegeben sein (Reinecke 2005): (1) Das Modell muss identifizierbar sein, worunter zu verstehen ist, dass die Anzahl der Varianzen und Kovarianzen hinsichtlich der manifesten Variablen ausreichen muss, um die Anzahl der freien Parameter im spezifizierten Modell zu ermitteln. Diese Grundannahme

betrifft die Anzahl der Freiheitsgrade in einem Modell. Die Anzahl der Freiheitsgrade muss  $> 0$  sein, um ein Modell überhaupt identifizieren zu können; (2) die Parameter des Modells müssen skaliert werden, d.h. entweder die Varianz eines latenten Faktors oder der Regressionskoeffizient einer manifesten Variablen pro latenten Faktor muss fixiert werden. In der Regel wird einer der genannten Parameter auf 1.0 gesetzt. Die Wahl der Fixierung ist abhängig von der Fragestellung der Untersuchung. Da in der vorliegenden Arbeit die Schätzung der Faktorladungen aller manifesten Variablen auf die latente Variable bedeutsam ist, wird die Varianz des latenten Faktors im Messmodell auf 1.0 gesetzt. Somit resultiert eine Skalierung aller Faktorladungen am Ausmaß der Varianz der latenten Variablen.

In der vorliegenden Arbeit ist die Überprüfung von Restriktionen im Messmodell von zentraler Bedeutung. Von Restriktionen wird in einem Messmodell (oder in einem SEM) dann gesprochen, wenn bestimmte messtheoretische Hypothesen oder bestimmte Festlegungen von Parametern innerhalb des Modells überprüft werden (Arbuckle 2005). In der vorliegenden Arbeit werden die Intercepts (Achsenschnittpunkte auf Kovarianzebene), die Faktorladung und die Fehlervarianzen der Indikatoren zu beiden Messzeitpunkten als gleich gesetzt, da von einer Invarianz der Parameter zu beiden Messzeitpunkten ausgegangen wird. Die in der vorliegenden Arbeit vorgenommenen Restriktionen werden in Tabelle 24 dargestellt.

Tabelle 24: Restriktionen zur Testung von Parameterinvarianz zu zwei Messzeitpunkten

Faktorladungen	Intercepts *	Messfehler	Latentes Konstrukt
$\lambda_{11} = \lambda_{12}$	$\tau_{11} = \tau_{12}$	$\delta_{11} = \delta_{12}$	GLQ_var = 1
$\lambda_{21} = \lambda_{22}$	$\tau_{21} = \tau_{22}$	$\delta_{21} = \delta_{22}$	GLQ_M = 0
$\lambda_{31} = \lambda_{32}$	$\tau_{31} = \tau_{32}$	$\delta_{31} = \delta_{32}$	
$\lambda_{41} = \lambda_{42}$	$\tau_{41} = \tau_{42}$	$\delta_{41} = \delta_{42}$	
$\lambda_{51} = \lambda_{52}$	$\tau_{51} = \tau_{52}$	$\delta_{51} = \delta_{52}$	

Anmerkungen: \* = Achsenschnittpunkte, werden nur in der Kovarianzberechnung berücksichtigt,  $x_1$  = Parameter X zum Messzeitpunkt 1,  $x_2$  = Parameter X zum Messzeitpunkt 2, GLQ\_var = Varianz des latenten Faktors Gesundheitsbezogene Lebensqualität, GLQ\_M = Mittelwert des latenten Faktors Gesundheitsbezogene Lebensqualität

Zur Beantwortung der ersten Fragestellung wird überprüft, welche der genannten Parameter restringiert bleiben können oder frei gesetzt werden müssen. Das Vorgehen, welche Restriktionen aufgehoben werden, orientiert sich an einer Kombination aus statistischen Auswertungen mittels so genannter Modifikationsindizes (AMOS 17.0) und inhaltlicher Kriterien. Modifikationsindizes wurden bereits von Jöreskog und Sörbom (1986) für LISREL entwickelt und geben an, inwieweit sich der  $\chi^2$ -Wert des gesamten Modells (wenigstens) ver-

ändert, wenn zwischen zwei Parametern, die zuvor unkorreliert waren, im theoretisch postulierten Modell eine Beziehung angenommen wird. Somit verliert das Gesamtmodell einen Freiheitsgrad pro Modifikation. Jedoch führt das Vorgehen dann zu einer besseren Modellpassung, wenn der  $\chi^2$ -Wert für das Gesamtmodell deutlicher abnimmt als die Anzahl der Freiheitsgrade. Die wichtigste Voraussetzung für die Anwendung der Modifikationsindizes ist die inhaltliche Begründung und Plausibilität der Restriktionen, das heißt, es sollten nur zwischen solchen Variablen Beziehungen hergestellt werden, die theoretisch und inhaltlich begründbar sind. Streng genommen stellt die Anwendung der Modifikationsindizes einen Schritt von der konfirmatorischen Modellprüfung zur explorativen Modellentwicklung dar, da möglicherweise Veränderungen in ein bestehendes Modell aufgenommen werden, die vorher nicht postuliert wurden.

In der vorliegenden Arbeit wurden ebenfalls inhaltliche und statistische Kriterien berücksichtigt, um die definierten Restriktionen zu prüfen. Die Modifikationsindizes wurden eingesetzt, um den Einfluss der Restriktionen der Parameter auf den  $\chi^2$ -Wert zu schätzen. Überschreitet die Veränderung des  $\chi^2$ -Werts durch eine Restriktion ein bestimmtes Kriterium, werden diese Restriktionen im Parameter-Output angezeigt. Die Programm-Voreinstellungen in AMOS wurden so angelegt, dass ein  $\chi^2$ -Wert um wenigstens eine Einheit verändert wird, wenn eine (einzelne) Restriktion vorgenommen wird. Restriktionen mit einem geringeren Einfluss auf den  $\chi^2$ -Wert werden vom Programm nicht angezeigt. Die Schwelle für eine Einheit im  $\chi^2$ -Wert wird deshalb so niedrig angesetzt, um alle inhaltlich relevanten Parameterfreisetzungen statistisch zu prüfen. AMOS 17.0 gibt bei der Betrachtung der Modifikationsindizes ebenfalls (bereits) restringierte Beziehungen als  $\chi^2$ -Wert-veränderungsrelevant an und bietet somit einen Hinweis auf eine Freisetzung dieser Parameterrestriktion. Die inhaltliche Begründung der Freisetzung liegt in dem theoriegeleiteten Ansatz, dass lediglich response-shift-relevante Parameter entsprechend Tabelle 24 freigesetzt werden. Die Reihenfolge der Freisetzung richtet sich wiederum nach statistischen Kriterien.

### 6.3.2 Prüfung eines Messmodells

Die Überprüfung der Restriktionen beginnt mit der Bestimmung des Modell-Fit des vollständig restringierten Modells. Anschließend wird die Parameter-Freisetzung vorgenommen, die zu der größten kalkulierbaren Verbesserung des Modell-Fit führt. Danach wird die Modell-Passung neu berechnet und überprüft, ob weitere Parameter freigesetzt werden sollten. Ausschlaggebendes Kriterium für die Akzeptanz einer Parameterfreisetzung ist eine statis-

tisch bedeutsam verbesserte Modellpassung im Vergleich zum stärker ‚geschachtelten‘ (genesteten) Modell. Das Modell, in dem eine mögliche Parameterfreisetzung zu keiner statistisch bedeutsamen Verbesserung der Modellpassung führt, wird als Zielmodell bezeichnet. Dieses Vorgehen orientiert sich am Ablauf von Modellvergleichen zwischen geschachtelten Modellen (Schermelleh-Engel et al. 2003). Ein vollständig restringiertes Modell ist demnach eine mögliche, aber sehr spezifische Variante eines vollkommen unrestringierten Modells. In der vorliegenden Arbeit sind Modelle dann geschachtelt, wenn die Anzahl der freien Parameter eines stärker restringierten Modells eine Teilmenge der Anzahl der freien Parameter eines weniger restringierten Modells ist. Je mehr Restriktionen innerhalb eines Modells vorgenommen werden, desto spezifischer ist dieses Modell definiert. Der Unterschied zwischen der Passung zweier genesteter Modelle kann mittels  $\chi^2$ -Differenztest überprüft werden (Steiger et al. 1985). Ein stärker genestetes beziehungsweise restringiertes Modell hat weniger freie Parameter und mehr Freiheitsgrade als ein weniger restringiertes Modell (Arbuckle 2005). Für die vorliegende Arbeit bedeutet dies, dass Restriktionen so lange freigesetzt werden, wie die Freisetzung eines Parameters beim Vergleich zweier geschachtelter Modelle zu einer statistisch signifikanten Verbesserung der Modellpassung führt. Resultiert aus dem Vergleich eines Modells (Modell A) mit einem weniger restringierten Modell (Modell B), in dem ein (im Modell A noch restringierter) Modellparameter *mehr* freigesetzt ist, kein statistisch signifikanter Unterschied im  $\chi^2$ -Differenztest, werden keine weitere Parameterfreisetzungen mehr vorgenommen. Das Modell mit mehr Restriktionen (Modell A) enthält mehr spezifische Beziehungen zwischen einzelnen Modellvariablen als ein Modell mit weniger Restriktionen (Modell B). Deshalb wird bei einem nicht signifikanten  $\chi^2$ -Differenztest das stärker restringierte Modell (Modell A) favorisiert. Dieses Modell stellt das Ergebnis der Überprüfung der Response-Shift-Parameter dar.

### 6.3.3 Testung von Diskrepanzfunktion und Goodness-of-Fit-Indizes

Grundlage der Schätzung von Diskrepanzfunktion und Gütemaßen beziehungsweise Fit-Indizes ist die Entscheidung für eine bestimmte Schätzfunktion. Unter der Annahme multivariat normalverteilter Variablen und einer genügend großen Stichprobe ist die Maximum-Likelihood-Schätzfunktion (ML-Schätzung) zu wählen. In der vorliegenden Arbeit wird die ML-Schätzung verwendet. Boosma und Hoogland (2001) konnten zeigen, dass die ML-Schätzung selbst bei nicht-normalverteilten Variablen zu angemessenen Ergebnissen führte. Bei kleinen Stichproben wird das Bootstrapping-Verfahren als Möglichkeit empfohlen, um bessere Voraussetzungen für die ML-Schätzfunktion zu bieten (Efron & Tibshirani 1993). Da die vorlie-



gende Stichprobe aus einem  $N < 400$  besteht, wird das Bootstrapping-Verfahren bei jeder Berechnung verwendet.

Unter der Annahme, dass die Anwendungsvoraussetzungen für die ML-Schätzfunktion erfüllt sind, liefert diese Schätzung einen Funktionswert von Null, wenn die empirische und die modelltheoretische Kovarianzmatrix genau übereinstimmen ( $\Sigma = \Sigma(\Theta)$ ). Die exakte Übereinstimmung ist eher unwahrscheinlich, so dass die Werte der ML-Schätzung ungleich Null sind. Wird der geschätzte Wert mit der um eins verringerten Stichprobengröße multipliziert, folgt der so erhaltene Wert einer  $\chi^2$ -Verteilung. Je größer der  $\chi^2$ -Wert, desto eher liegt eine Diskrepanz zwischen empirischen Daten und modelltheoretischen Annahmen zugrunde. Die Diskrepanzfunktion ( $\chi^2$ -Wert) geht unter anderem von der oben genannten Voraussetzung aus, dass  $\Sigma = \Sigma(\Theta)$  ist, was bedeutet, dass eine eindeutige Beschreibung der Realität möglich ist. Diese Annahme ist jedoch unrealistisch. Deshalb ist zu prüfen, ob die Differenz zwischen der modelltheoretischen und empirischen Kovarianzmatrix zu vernachlässigen ist. Die Verteilung der  $\chi^2$ -Werte ist allerdings stichprobenabhängig, so dass bei zunehmender Stichprobengröße die Nullhypothese ( $\Sigma = \Sigma(\Theta)$ ) immer abgelehnt wird und somit die theoretisch postulierten Modelle immer widerlegt werden (Reinecke 2005). Aus diesem Grund wurden verschiedene deskriptive Fit-Indizes konzipiert, für deren Anwendung jeweils bestimmte Voraussetzungen erfüllt sein müssen und die für unterschiedliche Modellüberprüfungen geeignet sind. Hu und Bentler (1999) unterscheiden grob zwischen zwei Fit-Indizes, den *Absolut-Fit-Indizes* und den *Incremental-Fit-Indizes*. Erstgenannte werden dann verwendet, wenn die Überprüfung eines theoretisch postulierten Modells mit einer empirischen Kovarianzmatrix verglichen werden soll. Die Indizes zur Testung auf ‚incremental fit‘ werden beim Vergleich geschachtelter Modelle eingesetzt, wobei ein Modell mit einem stärker restringierten Modell verglichen wird. Aufgrund der unterschiedlichen Beschaffenheiten der einzelnen Fit-Indizes kommen Hu und Bentler (1999) zu dem Ergebnis, dass für die Modellprüfung immer wenigstens zwei Fit-Indizes verwendet werden sollen, um die jeweiligen Unsicherheiten eines Index zu berücksichtigen.

In der vorliegenden Arbeit werden folgende Maße zur Beurteilung der Modelle eingesetzt: der Quotient aus der Diskrepanzfunktion relativiert an der Anzahl der Freiheitsgrade ( $\chi^2/df$ ). Als Absolut-Fit-Index zur Erfassung der Modellgüte wird der *Root Mean Square Error of Approximation* (RMSEA) von Steiger (1990) eingesetzt. Der RMSEA berücksichtigt die Annahme der approximativ ermittelten Passung zwischen tatsächlich geltendem und the-

oretisch postuliertem Modell, indem die Stichprobengröße sowie die Anzahl der Freiheitsgrade stärker berücksichtigt werden. Die Anzahl der Freiheitsgrade ist für die vorliegende Fragestellung relevant, da das Ausmaß der Restriktionen diesen Wert beeinflusst. Als Incremental-Fit-Indizes werden der *Comparative Fit Index* (CFI) von Bentler (1990) sowie der *Tucker-Lewis-Index* (TLI) beziehungsweise der *Non-Normed Fit Index* (NNFI) verwendet (zum TLI bzw. NNFI siehe Bollen 1989). Den zuletzt genannten Fit-Indizes liegt die Annahme zugrunde, die Passung eines aktuellen Modells mit einem weniger/stärker restringierten Modell zu vergleichen, wobei dieses Vergleichsmodell entsprechend mehr/weniger Freiheitsgrade aufweist (Marsh et al. 1996). Der TLI ist unabhängig von der Stichprobengröße und berücksichtigt besonders die Komplexität eines Modells. Als zentrales Kriterium erfasst er die Unterschiede zwischen Modellen hinsichtlich ihrer Fehlspezifikationen. Bei einem Wert über 1.0 weist der TLI darauf hin, dass im getesteten Modell mehr Parameter spezifiziert sind als notwendig. In einem solchen Fall wird von Overfitting gesprochen (Bollen 1989). Der CFI als inkrementelles Maß berücksichtigt die Stichprobengröße als wichtiges Einflusskriterium, wobei insbesondere Fehlspezifikationen bei kleinen Stichproben vermieden werden (Schermelleh-Engel et al. 2003). In Tabelle 25 werden die verwendeten Indizes und Gütemaße mit ihren Werten für akzeptables und gutes Fit angegeben.

Tabelle 25: Angaben zur Modellbewertung mit Wertebereichen für Modellpassungen

Gütemaß	gute Passung	akzeptable Passung
$\chi^2/df$	$0 \leq \chi^2/df \leq 2$	$2 \leq \chi^2/df \leq 3$
RMSEA	$0 \leq RMSEA \leq .05$	$0.5 < RMSEA \leq .08$
CFI <sup>1</sup>	$.97 \leq CFI \leq 1.00$	$.95 \leq CFI < .97$
TLI	$.95 \leq TLI \leq 1.00$	$.90 \leq TLI < .95$
AIC	Je kleiner der Wert, um so besser	
ECVI	Je kleiner der Wert, um so besser	

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden; RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index, AIC = Akaike Information Criterion, ECVI = Expected Cross Validation Index, <sup>1</sup> Angaben zum CFI aus Schermelleh-Engel et al. (2003), die diese Grenzen eher empfehlen als die herkömmliche Einteilung nach .90 bis .95 für akzeptablen Fit und .95 bis 1.0 für guten Fit

Zur Ergänzung werden noch zwei weitere Indizes aufgenommen, die sich ebenfalls auf den Vergleich verschiedener (genesteter) Modell beziehen. Diese zählen zu den so genannten *Sparsamkeitsmaßen*, die insbesondere die Anzahl zu schätzender Parameter mit dem  $\chi^2$ -Wert in Beziehung setzen. Dabei handelt es sich um das *Akaike Information Criterion* (AIC) von Akaike (1974) und den *Expected Cross Validation Index* (ECVI) von Browne und Cudeck

(1993), der neben der Anzahl der zu schätzenden Parameter auch noch die Stichprobengröße mit einbezieht.

### 6.3.4 Ableitung eines Messmodells

Ausgangspunkt der vorliegenden Arbeit ist ein Modell der GLQ nach Fayers und Hand (1997) sowie Boehmer und Luszczynska (2006), das GLQ als latenten Faktor beschreibt, der durch die Komponenten Rollenfunktionsfähigkeit, Emotionale, Soziale, Kognitive und Physische Funktionsfähigkeit repräsentiert wird. Die Funktionsskalen des EORTC-QLQ-C30 von Fayers und Kollegen (1995) erfassen diese Inhaltsbereiche mit jeweils einer Skala. Dieses theoretisch und empirisch begründete Modell wurde in einem ersten Schritt der Modellableitung auf seine Passung bei einer heterogenen Stichprobe von Brustkrebspatientinnen und PCa-Patienten in der Anschlussheilbehandlung überprüft. In Abbildung 6 wird das Messmodell graphisch dargestellt, das mittels KFA getestet wurde.

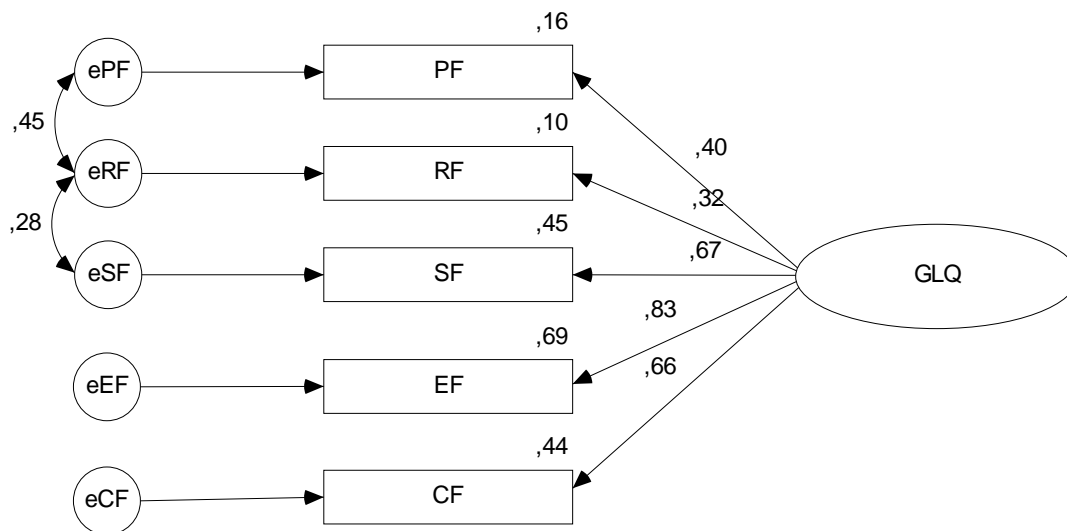


Abbildung 6: Messmodell zur GLQ bei 185 Krebspatientinnen und -patienten

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, e = Messfehleranteile der jeweiligen Indikatoren, GLQ = Gesundheitsbezogene Lebensqualität

Bei der Modellprüfung wurde auf die Modifikationsindizes zurückgegriffen, um die Passung des Modells zu verbessern. Es zeigte sich, dass der Messfehler eRF (Messfehleranteil der Indikatorvariable Rollenfunktionsfähigkeit) sowohl mit dem Messfehler ePF (Messfehler der Indikatorvariable Physische Funktionsfähigkeit) als auch mit dem Messfehler eEF (Mess-

fehler der Indikatorvariable Emotionale Funktionsfähigkeit) korreliert. Nach Zulassung dieser beiden Korrelationen wurden keine weiteren Modifikationen mehr vorgenommen. Das Modell weist allgemein eine gute Passung auf, was durch die Höhe der Faktorladungen bestätigt wird. Kritisch anzumerken ist eine Faktorladung von  $\lambda = .32$  für den Indikator Rollenfunktionsfähigkeit auf der GLQ.

Aus studientechnischen Gründen liegen für diese Stichprobe keine ausreichenden soziodemographischen Daten vor, so dass keine differenzierenden Aussagen über die Modellpassung abgeleitet werden können. Ziel dieses ersten Arbeitsschrittes ist die Überprüfung des aus der Literatur bekannten Modells der GLQ anhand einer heterogenen Stichprobe von Krebspatientinnen und Krebspatienten. Die Ergebnisse der ersten Modellprüfung werden in Tabelle 26 dargestellt. Alle Personen nahmen an der onkologischen Rehabilitation im Rahmen einer Anschlussheilbehandlung teil.

Tabelle 26: Bewertung des Modells zur gesundheitsbezogenen Lebensqualität bei 185 Brustkrebspatientinnen und Prostatakrebspatienten

Gütemaß	Index-Wert	Passung
$\chi^2/df$	2.14	akzeptabel
RMSEA	0.08	akzeptabel
CFI	0.99	gut
TLI	0.95	gut

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

In einem weiteren Schritt wurde das aus der Literatur bestätigte Modell gegen ein zweifaktorielles Modell getestet, in dem die Indikatorvariablen Physische, Soziale und Rollenfunktion zu einer latenten Variable *Physische Skala der GLQ* und die beiden Skalen emotionale und kognitive Funktionsfähigkeit zur latenten Variable *Psychische Skala der GLQ* zusammengefasst wurden. Dieses Alternativmodell wies in den allgemeinen Gütemaßen etwas bessere Werte auf. Jedoch sprechen zwei Aspekte gegen das Zweifaktoren-Modell: Zum Einen wird der latente Faktor Psychische Skala der GLQ lediglich durch zwei manifeste Variablen berechnet. Zum Anderen weist das zweifaktorielle Modell eine überschätzte Korrelation zwischen den beiden latenten Variablen auf, die mit einem Wert  $> 1.0$  die Ablehnung des Alternativmodells bestätigt. Aus diesen Gründen wird in der vorliegenden Arbeit das einfaktorielle Modell verwendet. Die Angaben zu den Fit-Indizes sowie eine graphische Darstellung des zweifaktoriellen Modells sind im Anhang enthalten (Anhang A1).

Aufgrund der zum Teil kritischen Werte für Cronbachs Alpha der einzelnen Skalen des EORTC-QLQ-C30 wurde eine zusätzliche KFA auf Itemebene durchgeführt. In diesem Modell wird davon ausgegangen, dass die jeweiligen Items einer Skala einen latenten Faktor 1. Ordnung repräsentieren. Auf der latenten Ebene zweiter Ordnung wurde ein latenter Faktor GLQ postuliert, der die latenten Faktoren erster Ordnung bestimmt. Somit wurde entsprechend der modelltheoretischen Vorgaben von einem Generalfaktormodell ausgegangen. Ebenso wie im Modell auf Skalenebene wurde zwischen den Messfehlern der latenten Faktoren Rollen- und Physische Funktionsfähigkeit sowie zwischen den Messfehler der latenten Faktoren Rollen- und Soziale Funktionsfähigkeit eine Interkorrelation angenommen. Die Überprüfung der Modellpassung zeigt insgesamt einen akzeptablen Fit des Gesamtmodells (siehe Tabelle 27).

Tabelle 27: Bewertung des Modells zur gesundheitsbezogenen Lebensqualität bei 185 Brustkrebspatientinnen und Prostatakrebspatienten auf Itemebene

Gütemaß	Index-Wert	Passung
$\chi^2/df$	1.72	gut
RMSEA	0.06	akzeptabel
CFI	0.95	akzeptabel
TLI	0.92	akzeptabel

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Bei der Betrachtung der Ladungen auf Ebene der latenten Faktoren fällt auf, dass die Emotionale Funktionsfähigkeit einen ausgesprochen hohen Wert besitzt (siehe Abbildung 7). Insgesamt liegen auf Ebene der latenten Faktoren Ladungen in einem guten bis zufriedenstellenden Wertebereich. Auf der Ebene der Ladungen der Einzelitems auf den entsprechenden latenten Faktoren erster Ordnung liegen die Werte in einem Bereich von .21 bis .94. Insgesamt zeigen sich gute bis angemessene Faktorladungen. Die Parameterschätzungen der Prüfung des abgeleiteten Modells auf Ebene der Einzelitems sowie auf erster und zweiter Ordnung latenter Faktoren bestätigen die Verwendung des vorliegenden Fragebogens sowie die Definition des abgeleiteten Modells. In den weiteren Analysen werden die Auswertungen mittels KFA auf Ebene der Skalenwerte des EORTC-QLQ-C30 vorgenommen.

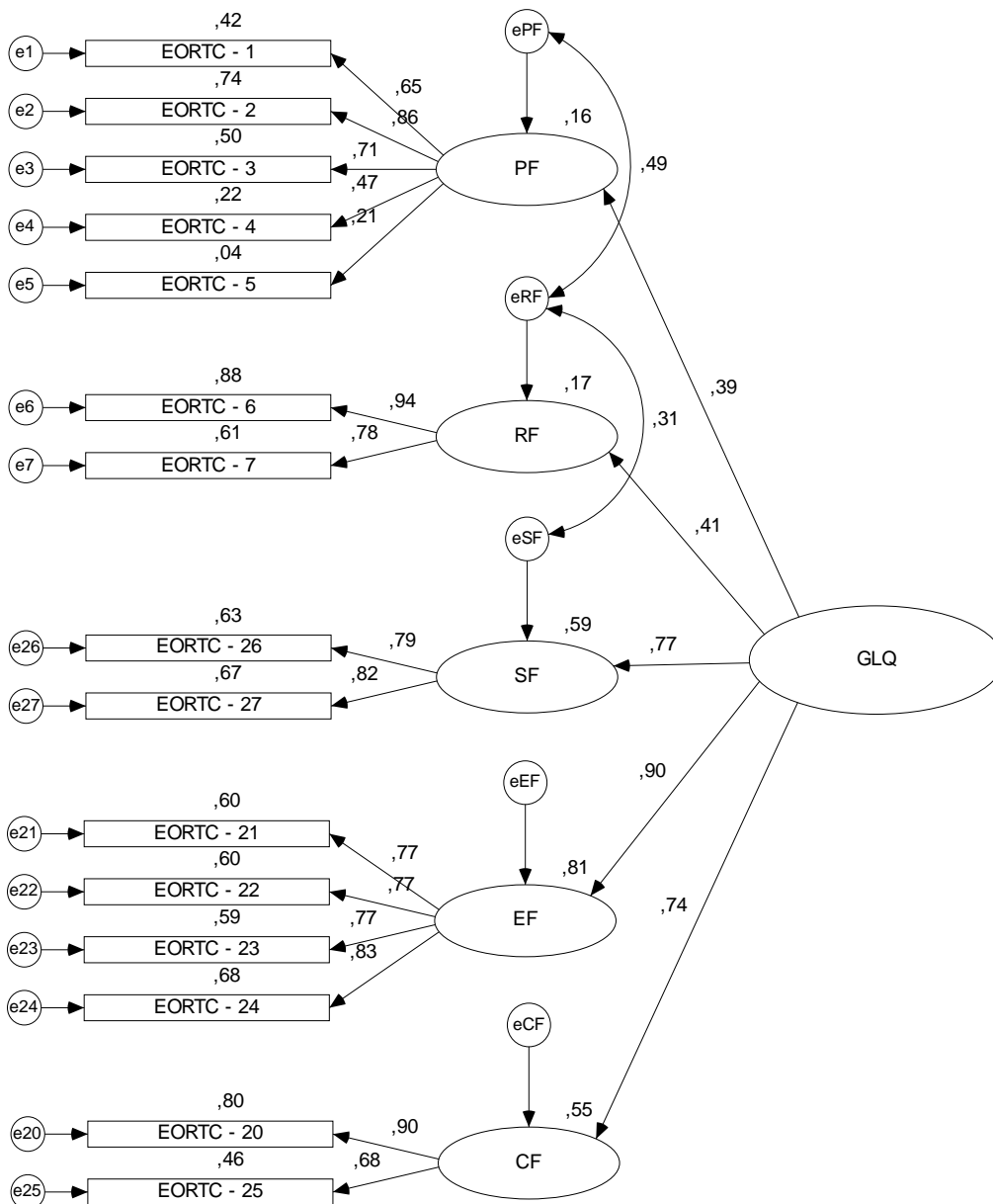


Abbildung 7: Messmodell auf Itemebene zur GLQ bei 185 Brustkrebspatientinnen und Prostatakrebspatienten auf Ebene der Einzelitems und latenten Faktoren 1. und 2. Ordnung

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, e = Messfehleranteile der jeweiligen Indikatoren oder latenten Variablen, GLQ = Gesundheitsbezogene Lebensqualität, EORTC – 1 bis 27 = Items des Fragebogens EORTC-QLQ-C30

## 7. Ergebnisse

Im voran gegangenen Kapitel wurden die methodischen Grundlagen für die Auswertungen der vorliegenden Fragestellungen dargestellt. Im Folgenden werden die Ergebnisse der Berechnungen mittels Konfirmatorischer Faktoren Analyse (KFA) präsentiert. Zuerst soll überprüft werden, ob in der zugrundeliegenden Hauptstichprobe der 212 PCa-Patienten mit AHB von Beginn der Maßnahme bis drei Monate nach deren Beginn ein Response-Shift auftritt (Fragestellung 1). Anschließend wird der Einfluss der Veränderung der Globalen Lebensqualität auf Response-Shift-Prozesse untersucht (Fragestellung 2). Abschließend werden potenzielle Moderatorvariablen aus dem Modell von Sprangers und Schwartz (1999) auf ihren Einfluss auf einen Response-Shift geprüft. Dazu zählen Benefit Finding, Selbstwirksamkeit, positive und negative Affektivität (Fragestellung 3).

### 7.1 Ergebnisse zur Fragestellung 1

Im Folgenden werden die Ergebnisse der Prüfung auf Response-Shift mittels KFA an der Gesamtstichprobe von 212 PCa-Patienten in der Anschlussheilbehandlung dargestellt. Vorab wird das im vorhergehenden Kapitel an einer heterogenen Stichprobe aus Brustkrebspatientinnen und PCa-Patienten bestätigte Modell der GLQ auf Passung in der Gesamtstichprobe untersucht. Anschließend werden die verschiedenen Parameterrestriktionen eingeführt und sukzessive auf ihre Haltbarkeit geprüft, bis ein Endmodell, das Response-Shift-Modell, erreicht wird.

#### 7.1.1 Prüfung des Basismodells

In einem ersten Schritt wird das im vorangegangenen Kapitel bestätigte Messmodell im Sinne eines Basismodells als Kombinationsmodell aus beiden Messzeitpunkten berechnet. Die Berechnungsgrundlage stellt die Stichprobe der 212 PCa-Patienten in der AHB dar. Zu diesem Zweck wird das bisherige Messmodell, bestehend aus den fünf Skalen zu den Indikatoren und dem latenten Faktor GLQ, in zweifacher Form in das Basismodell aufgenommen. Jeweils ein Messmodell bildet die Faktorenstruktur zu einem Messzeitpunkt ab, so dass im Basismodell beide Messzeitpunkte simultan getestet werden. Um einerseits eine Modellprüfung überhaupt zu ermöglichen und andererseits inhaltlich relevante Beziehungen zwischen den Variablen und Faktoren der beiden Messzeitpunkte zuzulassen, werden verschiedene Korrelationen im Basismodell vorab definiert. Diese umfassen die Korrelation von Messfehlern

einer manifesten Variablen zu beiden MZPen sowie die Korrelation der latenten Variable GLQ zu MZP 1 und MZP 2. Das Modell wird in Abbildung 8 dargestellt. Die Angaben über die Modellpassung sind Tabelle 28 zu entnehmen.

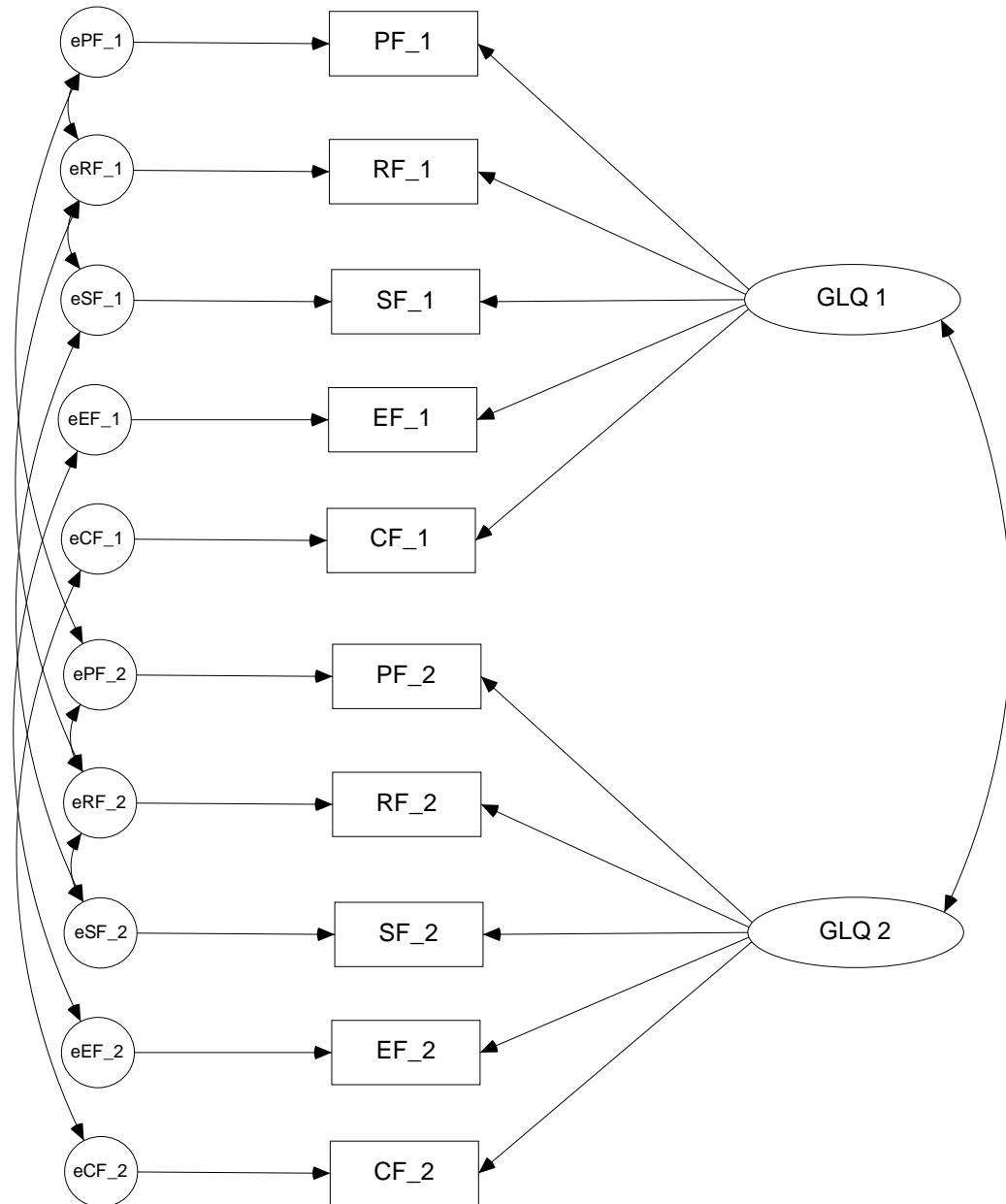


Abbildung 8: Modellstruktur des Basismodells

Anmerkungen:

e = Messfehler, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ 1 = Gesundheitsbezogene Lebensqualität zum MZP 1, GLQ 2 = Gesundheitsbezogene Lebensqualität zum MZP 2, \_1 = Parameter zum Messzeitpunkt 1, \_2 = Parameter zum Messzeitpunkt 2.



Tabelle 28: Bewertung des Basismodells zur GLQ bei 212 PCa-Patienten zu beiden MZPen

Gütemaß	Index-Wert	Passung
$\chi^2/df$	1.50	gut
RMSEA	0.05	gut
CFI	0.98	gut
TLI	0.97	gut

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Die Überprüfung der Modellpassung liefert gute Fit-Werte für das Basismodell. Alle vier Güte-Maße liegen im Bereich einer guten Passung zwischen theoretisch postulierter und empirisch vorliegender Kovarianzmatrix.

Um das Vorgehen im weiteren Verlauf zu veranschaulichen und einen besseren Vergleich der überprüften Modelle zu bieten, werden in Tabelle 29 die anhand des Modells in Abbildung 8 berechneten Werte sämtlicher Response-Shift-Parameter vorgestellt.

Tabelle 29: Schätzung der Response-Shift-Parameter im Basismodell (ohne Restriktionen)

Faktorladungen	Korrelationen		Kovarianzen	
	Prätest GLQ	Posttest GLQ	Prätest GLQ	Posttest GLQ
	PF	.46	.56	8.27
RF	.42	.69	14.43	16.96
SF	.67	.78	20.22	19.84
EF	.64	.75	16.06	16.16
CF	.51	.50	10.74	9.35

Intercepts	PF	RF	SF	EF	CF
	Prätest	74.06	41.90	61.16	65.92
Posttest	83.81	70.60	72.72	74.92	83.10

Residual-Varianzen	PF	RF	SF	EF	CF
	Prätest	258.19	983.33	490.62	364.16
Posttest	129.15	323.23	250.95	199.30	257.83

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität

Auf der Ebene der Faktorladungen zeigen sich für alle Skalen zufriedenstellende bis gute Kennwerte. Die Werte der Intercepts stellen die Mittelwerte der überprüften Skalen dar, was dadurch bedingt ist, dass die Mittelwerte der latenten Variablen zu beiden Messzeitpunkten auf Null gesetzt wurden. In allen Skalen zeigen sich Werteveränderungen im Prä-Post-Vergleich, die auf eine ‚Verbesserung‘ im jeweiligen Bereich hinweisen. Je höher die Werte, desto positiver die Ausprägung im entsprechenden Funktionsbereich. Der Vergleich der Intercepts kann jedoch für keine der weiteren Response-Shift-Modelle übernommen werden, da die Restriktion der latenten Variable GLQ zum MZP 2 bei den weiteren Analysen aufgehoben wird. Der Messfehlereinfluss nimmt zum MZP 2 auf allen Skalen ab.

### 7.1.2 Prüfung des ‚Null-Modells‘

In einem nächsten Auswertungsschritt werden alle einzelnen Response-Shift-Parameter (Faktorladungen, Intercepts und Fehlervarianzen) für beide Messzeitpunkte als invariant definiert und restringiert. Inhaltlich bedeuten diese Restriktionen, dass beispielsweise die Faktorladungen zwischen der manifesten Variable Physische Funktionsfähigkeit und dem latenten Konstrukt GLQ zu beiden MZPen identisch sind. Die Bezeichnung ‚Null-Modell‘ bezieht sich im Kontext der Arbeit auf die Annahmen der Nullhypothesen, die davon ausgehen, dass sich der Wert eines Response-Shift-Parameters zwischen den MZPen nicht verändert. Die Restriktionen im vorliegenden ‚Null-Modell‘ werden in Tabelle 30 dargestellt. Die Varianz des latenten Faktors GLQ zum MZP 1 wird weiterhin auf 1 gesetzt, um die Parameter des Messmodells zum MZP 1 an der Varianz der latenten Variable zu standardisieren. Über die Parameterrestriktionen ist eine Freisetzung der Varianz des latenten Faktors zum MZP 2 möglich. Ebenso wird der Mittelwert des latenten Faktors zum MZP 2 nicht mehr restringiert.

Tabelle 30: Restriktionen zur Testung von Parameterinvarianz zwischen zwei Messzeitpunkten

Faktorladungen	Intercepts	Messfehler	Latenter Faktor
FL_PF_1 = FL_PF_2	IN_PF_1 = IN_PF_2	e_PF_1 = e_PF_2	GLQ_Var_1 = 1
FL_RF_1 = FL_RF_2	IN_RF_1 = IN_RF_2	e_RF_1 = e_RF_2	GLQ_M_1 = 0
FL_SF_1 = FL_SF_2	IN_SF_1 = IN_SF_2	e_SF_1 = e_SF_2	
FL_EF_1 = FL_EF_2	IN_EF_1 = IN_EF_2	e_EF_1 = e_EF_2	
FL_CF_1 = FL_CF_2	IN_CF_1 = IN_CF_2	e_CF_1 = e_CF_2	

Anmerkungen: FL = Faktorladung, IN = Intercepts, e = Messfehlervarianz, GLQ\_Var = Varianz der GLQ, GLQ\_M = Mittelwert des latenten Konstrukts, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, \_1 = Parameter zum MZP 1, \_2 = Parameter zum MZP 2.

Die Überprüfung des ‚Null-Modells‘ auf Passung zwischen theoretisch erwarteter und empirisch vorhandener Kovarianzmatrix erbringt Ergebnisse, die auf eine inakzeptable Passung hinweisen. Insgesamt lässt sich aus den Werten der Fit-Maße ableiten, dass die vorgenommenen Restriktionen in dieser Form nicht haltbar sind. Tabelle 31 enthält die Ergebnisse.

Tabelle 31: Passung des ‚Null-Modells‘

Gütemaß	Index-Wert	Passung
$\chi^2/df$	5.30	inakzeptabel
RMSEA	0.14	inakzeptabel
CFI	0.76	inakzeptabel
TLI	0.71	inakzeptabel

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

In Tabelle 32 werden zur besseren Veranschaulichung der Parameterprüfung die Werte aller Response-Shift-Parameter dargestellt.

Tabelle 32: Schätzung der Response-Shift-Parameter im vollständig restringierten Modell (‚Null-Modell‘)

Faktorladungen	Korrelationen		Kovarianzen	
	Prätest GLQ	Posttest GLQ	Prätest GLQ	Posttest GLQ
	PF	.57	.51	9.83
RF	.68	.62	23.10	23.10
SF	.74	.68	21.01	21.01
EF	.65	.58	15.22	15.22
CF	.40	.34	7.73	7.73

Intercepts	PF	RF	SF	EF	CF	
	Prätest	74.63	48.29	59.10	65.11	79.93
	Posttest	74.63	48.29	59.10	65.11	79.93

Residual-Varianzen	PF	RF	SF	EF	CF	
	Prätest	197.36	630.79	362.22	322.94	324.23
	Posttest	197.36	630.79	362.22	322.94	324.23

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität

Der direkte Modellvergleich zwischen dem Basismodell und dem ‘Null-Modell’ anhand einer  $\chi^2$ -Statistik ist aufgrund der unterschiedlichen Parameterrestriktionen der latenten Faktoren nicht möglich. Auf eine umfassende Gegenüberstellung aller Modelle wird weiter unten eingegangen. Beim Vergleich der Werte der Tabellen 29 und 32 werden durchaus beträchtliche Unterschiede ersichtlich. Das gilt insbesondere für die Intercepts und Messfehlervarianzen.

In den Korrelationsberechnungen des standardisierten Modells ergeben sich Abweichungen in der Höhe der Faktorladungen zwischen den Messzeitpunkten. Das ist darauf zurückzuführen, dass Mittelwerte und Varianzen der latenten Faktoren variieren dürfen und somit zu beiden Messzeitpunkten (geringfügige) Abweichungen in den Faktorladungen resultieren. Für die Kovarianzen im unstandardisierten Modell trifft dies nicht zu.

### 7.1.3 Ableitung eines Response-Shift-Modells

In einem weiteren Schritt wurden die in Tabelle 30 aufgeführten Parameterrestriktionen auf ihren Einfluss auf die Modellpassung überprüft. Dabei wurde ein Vorgehen gewählt, bei dem einerseits auf die beschriebenen Modifikationsindizes zurückgegriffen wurde und andererseits inhaltliche Überlegungen ausschlaggebend waren, welche Restriktionen beibehalten und welche ‚befreit‘ werden konnten. Die inhaltlichen Kriterien bezogen sich darauf, dass lediglich response-shift-relevante Parameter (Faktorladungen, Intercepts, Fehlervarianzen) freigesetzt wurden. Weiterhin wurden verschiedene Reihenfolgen der Parameterfreisetzungen überprüft, um ein abschließendes Response-Shift-Modell zu erhalten. Die Variation der Reihenfolge der Parameterfreisetzungen wurde durchgeführt, da mit jeder einzelnen Parameterfreisetzung die Kovarianzen zwischen den im Modell definierten Variablen insgesamt verändert werden, so dass die Möglichkeit geprüft werden sollte, ob alle im ersten Durchlauf identifizierten response-shift-beeinflussenden Parameter tatsächlich auch freigesetzt werden sollten.

Die Parameterfreisetzungen wurden mittels  $\chi^2$ -Statistik auf ihre statistische Bedeutsamkeit überprüft. In einem ersten Schritt wurden alle (response-shift-relevanten) Freisetzungen vorgenommen, um die Veränderung der Modellpassung zu beschreiben, unabhängig davon, ob diese statistisch bedeutsam war oder nicht. Die Entscheidung für das Response-Shift-Modell wurde aufgrund der Fitmaße zum Modellvergleich geschachtelter Modelle getroffen.

Aufgrund des Vergleichs verschiedener Variationen der Parameterfreisetzungen resultierte ein ‚Response-Shift-Modell‘ mit den in Tabelle 33 dargestellten Fitmaßen. Aus der Tabelle ist erkennbar, dass ein Modell abgeleitet wurde, das über alle Fit-Maße eine gute Passung aufweist.

Tabelle 33: Passung des ‚Response-Shift-Modells‘

Gütemaß	Index-Wert	Passung
$\chi^2/df$	1.31	gut
RMSEA	0.04	gut
CFI	0.99	gut
TLI	0.98	gut

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Im Verlauf der Überprüfung des Response-Shifts stellte sich heraus, dass zwei Modelle ableitbar sind, die sich durch eine nahezu identische Güte auszeichnen. Neben dem Response-Shift-Modell wird das zweite Modell mit guter Passung als ‚Alternativ-Modell‘ bezeichnet. Diese zwei Modelle unterscheiden sich bezüglich ihrer Parameterrestriktionen dadurch, dass im Response-Shift-Modell die Faktorladung der Skala Rollenfunktionsfähigkeit restringiert und diese im Alternativ-Modell freigesetzt wird. Die Fitmaße der beiden Modelle werden in Tabelle 34 dargestellt.

Tabelle 34: Vergleich zweier verschiedener Response-Shift-Modelle

Gütemaß	$\chi^2/df$	RMSEA	CFI	TLI	AIC	ECVI
Response-Shift-Modell	1.31	.04	.99	.98	109.35	.52
Alternativ-Modell	1.31	.04	.99	.98	110.11	.52

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index, AIC = Akaike Information Criterion, ECVI = Expected Cross Validation

Wie in Kapitel 6.3.2 (Prüfung eines Messmodells) beschrieben, wurde aufgrund der (beinahe) identischen Werte der Fit-Maße das Kriterium des spezifischer formulierten Modells als Entscheidungskriterium herangezogen. Im vorliegenden Fall unterscheidet sich das Response-Shift-Modell vom Alternativ-Modell dadurch, dass eine Restriktion und somit ein Freiheitsgrad mehr im Modell enthalten ist. Somit ist das stärker restringierte Modell, das Response-Shift-Modell, zu bevorzugen, da es inhaltlich eine spezifischere Aussage mehr zulässt.

Die Reihenfolge der Freisetzung der restringierten Response-Shift-Parameter wird in Abbildung 9 dargestellt. Als Beschreibungskriterium wurde der Root Mean Square Error of

Approximation (RMSEA, Steiger 1990) gewählt. Die Freisetzungen der ersten acht Response-Shift-Parameter führten zu einer stetigen Verbesserung der Modellpassung. Ausgehend vom als RS-Modell gekennzeichneten Response-Shift-Modell zeigten sich keine weiteren signifikanten Verbesserungen der Modellpassung.

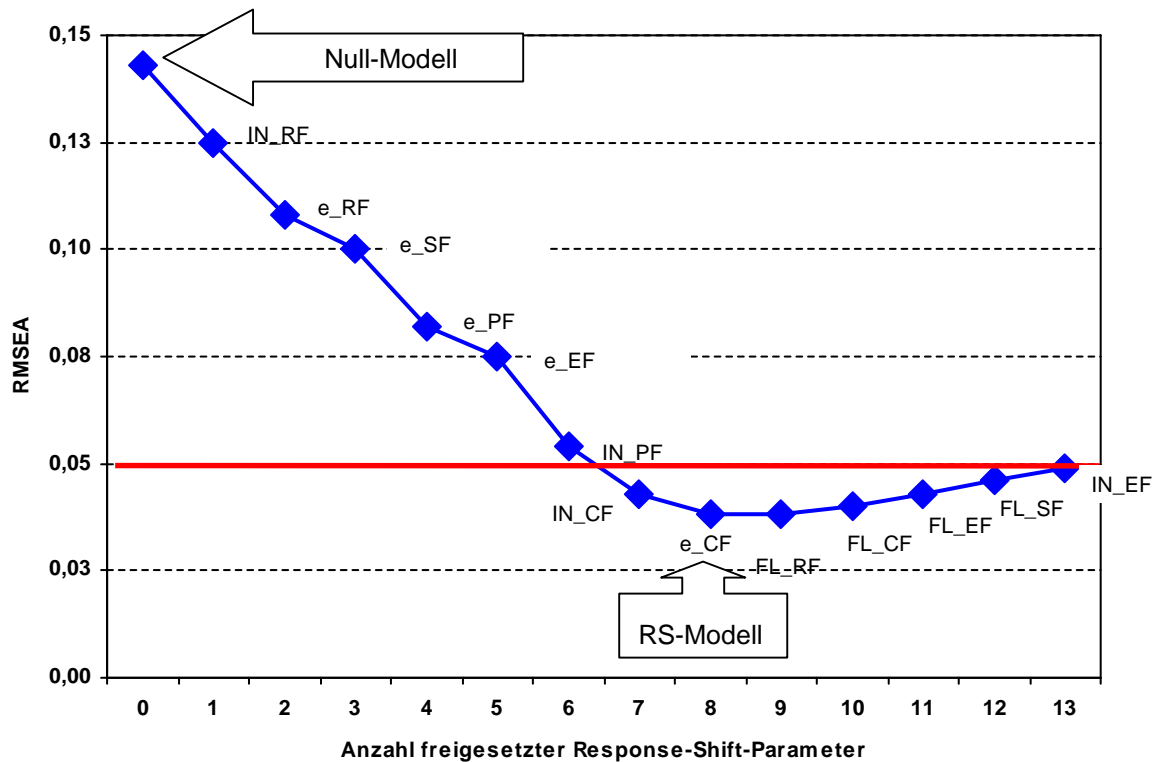


Abbildung 9: Beziehung zwischen RMSEA-Veränderungen und Parameterfreisetzungen

Anmerkungen: IN = Intercept, e = Messfehler, FL = Faktorladung, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, RS-Modell = Response-Shift-Modell

In Tabelle 35 werden Angaben über den statistischen Vergleich anhand der  $\chi^2$ -Diskrepanzfunktion zwischen dem Response-Shift-Modell und den geschachtelten (Vorläufer-)Modellen dargestellt. Das Ausgangsmodell ist das vollständig restringierte ‚Null-Modell‘, da ausgehend von diesem die Freisetzung der Parameter vorgenommen wurde. Die Angaben zur statistischen Signifikanz beziehen sich auf den Unterschied zwischen dem Response-Shift-Modell und den jeweils stärker restringierten Modellen. Die Einzelvergleiche zeigen, dass sich die Modelle vom Ausgangsmodell bis zum Modell mit sechs freigesetzten Parametern (Modell 6 – IN\_PF) im direkten Vergleich mit dem RS-Modell jeweils statistisch hochsignifikant unterscheiden ( $p < .01$ ). Der Unterschied in der Modellpassung zwischen dem Modell mit sieben Parameterfreisetzungen (Modell 7 – IN\_CF) und dem RS-Modell verfehlt mit ei-

nem  $p = 0.059$  knapp die Grenze zur statistischen Signifikanz. Aufgrund der Nähe zum Signifikanzniveau von 5% wird das Modell 8 – e\_CF dennoch als Modellverbesserung gewertet.

Wie oben bereits beschrieben, resultierte aus der nächsten Parameterfreisetzung (Modell 9 – FL\_RF) ein Modell mit nahezu identischer Passung, wobei die Parameterfreisetzung zu keiner beachtenswerten Verbesserung der Modellpassung beiträgt. Somit wird das Modell 8 als Response-Shift-Modell definiert.

Tabelle 35: Vergleich der geschachtelten Modelle mit dem Response-Shift-Modell

Geschachtelte Modelle	Vergleich: Response-Shift-Modell (Modell 8 – e_CF)		
	df	CMIN	p-Wert
„Null-Modell“	8	161,42	< .001
Modell 1 – IN_RF	7	118,25	< .001
Modell 2 – e_RF	6	84,93	< .001
Modell 3 – e_SF	5	68,98	< .001
Modell 4 – e_PF	4	43,41	< .001
Modell 5 – e_EF	3	32,59	< .001
Modell 6 – IN_PF	2	12,57	< .01
Modell 7 – IN_CF	1	3,56	.059
Alternativ-Modell – Modell 9 – FL_RF	1	1,25	.264

Anmerkungen: IN = Intercept, e = Messfehler, FL = Faktorladung; df = Freiheitsgrade, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit

In Tabelle 36 wird ein zusammenfassender Überblick gegeben, welche Parameter im ‚Response-Shift-Modell‘ freigesetzt wurden und welche weiterhin restringiert blieben.

Tabelle 36: Restriktionen zur Testung auf Parameterinvarianz zwischen zwei Messzeitpunkten

Faktorladungen	Intercepts	Messfehler	Latenter Faktor
FL_PF_1 = FL_PF_2	<b>IN_PF_1 ≠ IN_PF_2</b>	<b>e_PF_1 ≠ e_PF_2</b>	Var_GLQ_1 = 1
FL_RF_1 = FL_RF_2	<b>IN_RF_1 ≠ IN_RF_2</b>	<b>e_RF_1 ≠ e_RF_2</b>	Mean_GLQ_1 = 0
FL_SF_1 = FL_SF_2	IN_SF_1 = IN_SF_2	<b>e_SF_1 ≠ e_SF_2</b>	
FL_EF_1 = FL_EF_2	IN_EF_1 = IN_EF_2	<b>e_EF_1 ≠ e_EF_2</b>	
FL_CF_1 = FL_CF_2	<b>IN_CF_1 ≠ IN_CF_2</b>	<b>e_CF_1 ≠ e_CF_2</b>	

Anmerkungen: FL = Faktorladung, IN = Intercepts, e = Messfehlervarianz, Var\_GLQ = Varianz der GLQ, Mean\_GLQ = Mittelwert des latenten Konstrukts, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, \_1 = Parameter zum MZP 1, \_2 = Parameter zum MZP 2.

Es wurde keine Faktorladung freigesetzt. Die Intercepts der Variablen Soziale und Emotionale Funktionsfähigkeit blieben restringiert. Die Restriktionen aller anderen Intercepts sowie der gesamten Messfehlervarianzen wurden aufgehoben.

Die Werte der einzelnen Response-Shift-Parameter werden in Tabelle 37 angegeben. Die Response-Shift-Überprüfung erfolgt auf der Ebene der Kovarianzen der Modellparameter. Die Faktorladungen beziehungsweise Kovarianzen zwischen der GLQ und den Indikatoren sind auch im Response-Shift-Modell konstant zu halten. Die standardisierten Faktorladungen variieren zwischen 0.44 bis 0.68 zum Messzeitpunkt 1 und 0.52 bis 0.78 zum Messzeitpunkt 2. Somit liegen die Werte der standardisierten Lösung im Bereich zufriedenstellender bis guter Faktorladungen. Die Höhe der Korrelationen variiert trotz Restriktion. Es wurden die Intercepts der Skalen Physische, Kognitive und Rollenfunktionsfähigkeit freigesetzt. Die beiden weiteren Skalen blieben restringiert. Auf den Skalen Physische und Rollenfunktionsfähigkeit wurde eine Parameterveränderung in einen positiven Bereich festgestellt, auf der Skala Kognitive Funktionsfähigkeit in einen negativen Wertebereich.

Tabelle 37: Schätzung der Response-Shift-Parameter

<b>Faktorladungen</b>					
	Korrelationen		Kovarianzen		
	Prätest GLQ	Posttest GLQ	Prätest GLQ	Posttest GLQ	
PF	0.44	0.56	7.97	7.97	
RF	0.48	0.67	16.78	16.78	
SF	0.68	0.78	20.45	20.45	
EF	0.66	0.75	16.44	16.44	
CF	0.48	0.52	10.03	10.03	
<b>Intercepts</b>					
	PF	RF	SF	EF	CF
Prätest	<b>74.06</b>	<b>41.95</b>	61.29	65.82	<b>82.15</b>
Posttest	<b>79.37</b>	<b>61.24</b>	61.29	65.82	<b>77.52</b>
<b>Residual-Varianzen</b>					
	PF	RF	SF	EF	CF
Prätest	<b>260.76</b>	<b>957.28</b>	<b>484.34</b>	<b>359.52</b>	<b>333.04</b>
Posttest	<b>129.15</b>	<b>333.11</b>	<b>251.31</b>	<b>199.05</b>	<b>255.80</b>
Anmerkungen:	PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität, fett = Response-Shift-Prozess				



Die Residualvarianzen aller fünf manifesten Variablen wurden im Response-Shift-Modell freigesetzt. In allen Parametervergleichen einer Skala ergab sich eine Reduktion der Varianz des entsprechenden Messfehlers zum MZP 2.

Bezüglich der formulierten Hypothesen lassen sich die bisher vorgestellten Ergebnisse folgendermaßen zusammenfassen: Aufgrund der Restriktion aller Faktorladungen zu beiden MZPen ist für die vorliegende Studie davon auszugehen, dass in der Beziehung zwischen den einzelnen Komponenten und der GLQ in der Gesamtstichprobe keine Repriorisierung stattgefunden hat. Daraus resultiert ebenso, dass es nicht zu einer Neukonzeptualisierung der GLQ gekommen ist. Die Parameterfreisetzungen der entsprechenden Intercepts und aller Messfehler weist auf eine Rekalibrierung hin.

Als uniforme Rekalibrierung lassen sich die Veränderungen der Intercepts der Skalen Physische, Kognitive und Rollenfunktionsfähigkeit bezeichnen, da hier Veränderungen auf der Ebene der Achsenschnittpunkte stattgefunden haben. Im Falle einer ausschließlichen vorliegenden uniformen Rekalibrierung wären die Veränderungen der Achsenschnittpunkte auf einheitliche Veränderung der gesamten Skalierungen der Indikatoren zwischen den MZPen zurückzuführen. Die Varianzveränderungen der Messfehler aller Skalen belegen jedoch die non-uniforme Rekalibrierung auf allen Skalen. Die Änderungen in den Messfehlervarianzen der manifesten Variablen sind ein Anzeichen dafür, dass sich die Wertebereiche der Skalen zwischen den MZPen in unterschiedlicher Weise (non-uniform) verändert haben. Nach Oort (2005a) ist dann von einer uniformen Rekalibrierung auszugehen, wenn eine Parameterfreisetzung den Intercept, nicht aber zusätzlich die Messfehlervarianz einer Skala betrifft. Dies ist auf den Skalen Physische, Kognitive und Rollenfunktionsfähigkeit jedoch nicht der Fall. *Somit gilt, dass ausgehend von den Parameterfreisetzungen innerhalb der vorliegenden Gesamtstichprobe ein Response-Shift in Form einer non-uniformen Rekalibrierung stattgefunden hat.*

In Tabelle 38 werden weitere Ergebnisse der Parameterschätzungen dargestellt. Die Korrelationen der Messfehler der manifesten Variablen zu beiden Messzeitpunkten liegen in einem mäßigen Bereich. Unter der Annahme der KTT sollten die Residuen nicht miteinander korreliert sein. Dass dies trotzdem der Fall ist, ist inhaltlich erklärbar: Beispielsweise können zu beiden MZPen die gleichen Fehlereinflüsse vorliegen, was in der vorliegenden Studie nicht kontrolliert wurde. Wegen der Identifizierbarkeit des Modells mit abhängigen Stichproben

sind die Korrelationen jedoch notwendig. Die Varianz der GLQ zum MZP 2 wurde an der Varianz der latenten Variable zum MZP 1 standardisiert und fällt etwas geringer aus als zum MZP 1. Zwischen der GLQ zum MZP 1 und der zum MZP 2 besteht eine hohe, aber akzeptable Korrelation. Die Mittelwertsdifferenz zwischen den beiden latenten Variablen beträgt 0.56 Einheiten und ist statistisch hochsignifikant. Eine Zunahme in den Mittelwerten der latenten Variablen zeigt eine positive Veränderung der GLQ an. *Dementsprechend ist beim vorliegenden Mittelwertvergleich von einer statistisch signifikanten Verbesserung der GLQ zum MZP 2 auszugehen.*

Tabelle 38: Weitere Parameterschätzungen des ‚Response-Shift-Modells‘

Korrelationen der Residuen					
	PF	RF	SF	EF	CF
Prätest x Posttest	0.39	0.17	0.25	0.32	0.37
Varianz latenter Faktoren					
	Prätest GLQ		Posttest GLQ		
	1.00		0.95		
Korrelation latenter Faktor					
	Prätest GLQ				
Posttest GLQ	0.67				
Mittelwerte latenter Faktor GLQ					
	Prätest	Posttest			
	0.00	0.56	p < .001		
Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität					

Tabelle 39 zeigt die drei zu vergleichenden Modelle mit den angegebenen Fit-Werten. Die Gegenüberstellung des Basismodells mit dem ‚Null-Modell‘ zeigt eine deutlich bessere Passung des Basismodells. Der Unterschied zwischen den geschachtelten Modellen (Null-Modell und Response-Shift-Modell) wurde auf statistische Signifikanz anhand der  $\chi^2$ -Statistik geprüft. Der Vergleich zwischen dem ‚Null-Modell‘ und dem ‚Response-Shift-Modell‘ zeigt eine deutlich bessere Passung zugunsten des Response-Shift-Modells. Unter der Annahme, das ‚Response-Shift-Modell‘ sei korrekt, unterscheidet sich das ‚Null-Modell‘ von diesem statistisch bedeutsam ( $\chi^2 = 161.42$ ,  $p < .001$ ). Die bessere Passung des ‚Response-Shift-Modells‘ im Vergleich zum Basismodell resultiert aus einer höheren Anzahl an Freiheitsgraden im zuletzt genannten Modell durch die beibehaltenen Restriktionen. Das Basismodell enthält fünf Restriktionen weniger als das ‚Response-Shift-Modell‘, was sich für das Response-Shift-

Modell auf allen Fit-Maßen günstig auswirkt. Die Werte der zusätzlich angegebenen Fit-Maße AIC und ECVI bestätigen das Gesamtergebnis. Nach diesen beiden Gütemaßen ist das Modell zu bevorzugen, das geringere Werte auf den genannten Indices ausweist.

Tabelle 39: Vergleich der geschätzten Modelle

Gütemaß	$\chi^2/df$	RMSEA	CFI	TLI	AIC	ECVI
Basismodell	1.50	.05	.98	.97	117.42	.56
Null-Modell	5.30	.14	.76	.71	255.21	1.21
Response-Shift-Modell	1.31	.04	.99	.98	109.30	.52

Anmerkungen:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index, AIC = Akaike Information Criterion, ECVI = Expected Cross Validation

Abschließend wurde ausgehend vom Response-Shift-Modell zur Überprüfung der Veränderung der Varianz der latenten Variable ein weiteres Modell berechnet, in dem die Varianzen der GLQ restringiert wurden. Nach Oort (2005a) kann mit Hilfe der Restriktion der Varianz der latenten Variablen getestet werden, ob deren Veränderung statistisch signifikant ist. Die Passung des so getesteten Modells unterscheidet sich vom Response-Shift-Modell formal durch die zusätzliche Festlegung der Varianz der GLQ vom MZP 2 auf den Faktor 1 und weist somit einen Freiheitsgrad mehr auf. Das so definierte Modell weist eine geringfügig bessere Passung auf als das Response-Shift-Modell, wobei der Unterschied zwischen den beiden Modellen statistisch nicht signifikant ist ( $\chi^2 = 0.123$ ,  $p = .726$ ). Aus dem Ergebnis läßt sich ableiten, dass die Veränderung der Varianz der GLQ zwischen beiden MZPen statistisch nicht signifikant ist. Eine ausführliche Darstellung der Modellparameter ist im Anhang A2 enthalten.

Auf eine Restriktion des Mittelwertes der latenten Variablen zu beiden MZPen wird verzichtet, da AMOS 17.0 die Veränderung der Mittelwerte der GLQ bereits auf statistische Bedeutsamkeit prüft. Es zeigte sich ein statistisch hoch signifikanter Unterschied zwischen beiden Mittelwerten (siehe Tabelle 38), der das Ausmaß der „wahren Veränderung“ der GLQ repräsentiert.

#### 7.1.4 Response-Shift und Effektstärkeschätzungen in der Gesamtstichprobe

Anhand der von Oort (2005a) publizierten Formel wurde der Einfluss des Response-Shifts auf die Größe der Effekte der beobachteten Mittelwertsunterschiede der relevanten Skalen bestimmt. Hinsichtlich der Größe der Effekte wird auf die von Cohen (1988) beschrie-

bene Unterteilung in kleine ( $d \sim 0.2$ ), mittlere ( $d \sim 0.5$ ) und große Effekte ( $d \sim 0.8$ ) zurückgegriffen. Außerdem soll die von Norman und Kollegen (2003) beschriebene Kategorisierung von Effektstärken in klinisch nicht relevante ( $d < 0.5$ ) und klinisch relevante ( $d \geq 0.5$ ) Veränderungen als weitere Beurteilungsgrundlage mit einbezogen werden.

Wie von Oort (2005a) vorgestellt, lässt sich der Einfluss einzelner Response-Shift-Prozesse auf die beobachtete Effektstärke bestimmen. In der vorliegenden Arbeit ist ein Response-Shift in Form einer non-uniformen Rekalibrierung festgestellt worden. Die Stabilität der Faktorladungen über beide MZPe ist ein Beleg für die Abwesenheit einer Repriorisierung der untersuchten Skalen sowie der Neukonzeptualisierung der GLQ innerhalb der untersuchten Gesamtstichprobe.

Wie aus der zugrundeliegenden Formel ebenfalls ersichtlich ist, zeigt sich der Einfluss der non-uniformen Rekalibrierung (Freisetzung der Messfehlervarianzen der Indikatoren) in unterschiedlichen Varianzen und Kovarianzen, die sich im Nenner der Gesamtformel ausdrücken. Die non-uniforme Rekalibrierung hat dann einen Einfluss auf die Differenz der Skalenmittelwerte der Indikatoren, wenn diese so ausgeprägt ist, dass sie neben der Varianzveränderung auch eine Interceptveränderung bedingt. Eine reine Varianzveränderung im Sinne der non-uniformen Rekalibrierung hat keine Auswirkungen auf die Mittelwertvergleiche im Prä-Post-Test. Für die Interceptveränderungen der Skalen Physische, Rollen- und Kognitive Funktionsfähigkeit wird davon ausgegangen, dass die non-uniforme Rekalibrierung zur Interceptfreisetzung beiträgt. Für die non-uniforme Rekalibrierung auf den Skalen Soziale und Emotionale Funktionsfähigkeit trifft das nicht zu. Aus der Formel nach Oort ist ebenfalls ersichtlich, dass bei Aufrechterhaltung der Restriktion der Intercepts im Falle der reinen non-uniformen Rekalibrierung sich diese im Zähler zu Null summieren und der Umfang des Effektes nur auf die Veränderung durch die GLQ zurückzuführen ist. In Tabelle 40 werden die Ergebnisse zur Effektstärkeschätzung in der Gesamtstichprobe dargestellt.

Tabelle 40: Effektkalkulation und Response-Shift

Skala	Response-Shift	Effektstärke Gesamtstichprobe		
		beobachtet	Response-Shift	'wahre Veränderung'
SF	non-uniforme Rek.	0.40	-----	0.40
<b>RF</b>	non-uniforme Rek. <sup>#</sup>	<b>0.81</b>	<b>0.55</b>	<b>0.26</b>
<b>PF</b>	non-uniforme Rek. <sup>#</sup>	<b>0.60</b>	<b>0.33</b>	<b>0.27</b>
EF	non-uniforme Rek.	0.37	-----	0.37
<b>CF</b>	non-uniforme Rek. <sup>#</sup>	<b>0.05</b>	<b>-0.22</b>	<b>0.27</b>

Anmerkung: Rek. = Rekalibrierung; <sup>#</sup> = mit Interceptbeeinflussung; SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit; fett = Skalen mit Response-Shift Einfluss

Aus der non-uniformen Rekalibrierung resultiert eine Differenzierung der Effektstärken auf den Skalen Rollen-, Physische und Kognitive Funktionsfähigkeit. Auf der Skala Rollenfunktionsfähigkeit ist der beobachtete Effekt nach Cohen (1988), berechnet nach der herkömmlichen Effektkalkulation, als groß zu bezeichnen. Bemerkenswert daran ist, dass dieser Effekt lediglich zu einem geringen Anteil aufgrund der Veränderungen der GLQ zustande kommt. Der Einfluss der non-uniformen Rekalibrierung entspricht in etwa dem Ausmaß eines mittleren Effektes. Der beobachtete Effekt ist demnach zu einem großen Anteil auf die Veränderung der Skalierung der Rollenfunktionsfähigkeit innerhalb der Stichprobe zurückzuführen. Der beobachtete Effekt auf der Skala Rollenfunktionsfähigkeit ist nach Norman und Kollegen (2003) von klinischer Relevanz ( $ES > 0.5$ ). Aus der Tabelle ist jedoch ebenfalls zu entnehmen, dass dies auf einen klinisch relevanten Response-Shift zurückzuführen ist. Das Ausmaß der Veränderung der GLQ ist im Vergleich dazu definitionsgemäß klinisch nicht relevant.

Die Differenzierung des beobachteten Effekts auf der Skala Physische Funktionsfähigkeit zeigt ein ähnliches Bild. Jedoch liegt der beobachtete Effekt eher im Bereich einer mittleren Ausprägung, der in etwa zu gleichen Teilen auf einen Response-Shift sowie auf die Veränderung der GLQ zurückzuführen ist. Insgesamt betrachtet ist der beobachtete Effekt klinisch relevant, beide Einzeleffekte jedoch sind es für sich genommen nicht.

Der beobachtete Effekt auf der Skala Kognitive Funktionsfähigkeit erweist sich als nahezu unbedeutend. Auffällig ist in diesem Zusammenhang jedoch die Wirkung des Response-Shift, der einen kleinen Effekt auf der Ebene der wahren Veränderung wiederum *neutralisiert*.

Inhaltlich lassen sich die Ergebnisse zur non-uniformen Rekalibrierung folgendermaßen zusammenfassen: In der Gesamtstichprobe ergibt sich eine uneinheitliche Veränderung in der Bewertung der Skalierung aller Indikatoren. Die befragten PCa-Patienten zeigen auf den Skalen Physische und Rollenfunktionsfähigkeit eine Tendenz dahingehend, dass sie ihren Zustand zum MZP 2 *positiver* beurteilten als zum MZP 1, wobei diese Veränderung nicht auf die Veränderung der GLQ zurückzuführen ist. Im Sinne der non-uniformen Rekalibrierung bedeutet dies, dass für denselben Wert der GLQ zu MZP 1 und MZP 2 unterschiedliche Werte in den Skalen Physische und Rollenfunktionsfähigkeit resultieren, wobei es sich dabei um eine uneinheitliche Veränderung einzelner Skalenabstufungen handelt. Diese Veränderung ist auf der Skala Rollenfunktionsfähigkeit stärker ausgeprägt als im Bereich der Physischen Funktionsfähigkeit. Auf die Teilnehmer der Gesamtstichprobe bezogen lassen sich die Ergebnisse auch so formulieren, dass die Befragten eine Veränderung in der Bewertung einzelner Antwortoptionen der Skalen der genannten Funktionsfähigkeiten zeigen, die sich durch eine positivere Bewertung einzelner Skalenstufen zum MZP 2 im Vergleich zum MZP 1 äußert, ohne dass dies auf die Veränderung der GLQ zurückzuführen ist.

Für die Skala Kognitive Funktionsfähigkeit ergibt sich eine Veränderung dahingehend, dass die Personen zum MZP 2 diese Fähigkeit eher schlechter bewerten als zum MZP 1, ebenfalls wiederum losgelöst von der Veränderung der GLQ. Bei gleichbleibender GLQ schätzen die befragten Männer ihre kognitive Funktionsfähigkeit zum MZP 2 teilweise *schlechter* ein als zum MZP 1.

Für die anderen Ergebnisse der non-uniformen Rekalibrierung auf den Skalen Soziale und Emotionale Funktionsfähigkeit lassen sich folgende inhaltliche Aussagen zusammenfassen: Durch die Abnahme der Varianz der Messfehler der Skalen und der gleichzeitig konstant gebliebenen Varianz der GLQ zu beiden MZP'en ist davon auszugehen, dass sich einige Skalenstufen der verwendeten Indikatoren (quantitativ) verändert haben, andere hingegen konstant geblieben sind. Die Skalenabstufungen haben sich demnach lediglich innerhalb bestimmter Wertebereiche der Skalen verändert. Jedoch sind die Varianzveränderungen nicht ausgeprägt genug, um einen Einfluss auf die Mittelwerte der genannten Skalen zu erreichen.

#### 7.1.5 Zusammenfassung der Ergebnisse der Fragestellung 1

Im voran gegangenen Kapitel wurden die Ergebnisse zur ersten Fragestellung präsentiert. Mittels Konfirmatorischer Faktoren Analyse (KFA) wurden verschiedene theoretisch abge-

leitete Modelle auf ihre Passung mit den vorliegenden Daten der Stichprobe der 212 PCa-Patienten mit AHB getestet.

In einem ersten Schritt wurde die Passung eines Basismodells, das sich aus den Messmodellen der GLQ zu beiden MZPen zusammensetzt, getestet. In diesem Modell wurden lediglich zur Berechnung des Gesamtmodells notwendige, an den üblichen Konventionen zur Durchführung konfirmatorischer Faktorenanalysen vorgegebene Parameterdefinitionen vorgenommen. Das theoretisch abgeleitete Modell zeigte insgesamt eine gute Passung mit der empirischen Kovarianzmatrix.

In einem zweiten Schritt wurden alle für die Überprüfung eines Response-Shifts notwendigen Parameterrestriktionen vorgenommen: Die Ausprägungen der Faktorladungen, Intercepts und Messfehlervarianzen der manifesten Variablen wurden festgelegt, indem angenommen wurde, dass sich die Werte der genannten Parameter einer manifesten Variable zu beiden MZPen nicht unterscheiden. Das so berechnete ‚Nullmodell‘ zeigte eine inakzeptable Passung mit den empirischen Daten der Stichprobe. Aufgrund dieses Ergebnisses ist davon auszugehen, dass in der vorliegenden Stichprobe zwischen den beiden MZPen ein Response-Shift stattgefunden hat.

In einem dritten Schritt wurde mit Hilfe von Modifikationsindizes ein Modell ausgehend vom ‚Nullmodell‘ abgeleitet, in dem alle Parameter freigesetzt wurden, die von einem Response-Shift betroffen waren. In diesem ‚Response-Shift-Modell‘ wurden alle Messfehlervarianzen und die Intercepts der Skalen Physische, Kognitive und Rollenfunktionsfähigkeit freigesetzt. Die Skalen Soziale und Emotionale Funktionsfähigkeit sowie alle Faktorladungen blieben restringiert. Das gleichzeitige Freisetzen des Intercepts und der Messfehlervarianz eines Indikators wurde auf eine ausgeprägte non-uniforme Rekalibrierung zurückgeführt. Daraus folgt, dass in der vorliegenden Stichprobe ein Response-Shift in Form einer non-uniformen Rekalibrierung aufgetreten ist. Eine Repriorisierung und eine Neukonzeptualisierung konnten in der Gesamtstichprobe nicht nachgewiesen werden.

Weitere Analysen erbrachten das Ergebnis, dass sich die GLQ als latentes Konstrukt zwischen den beiden MZPen statistisch signifikant verbesserte. Die Veränderung der Varianz der GLQ zwischen beiden MZPen war jedoch nicht statistisch bedeutsam.

Ein Einfluss der non-uniformen Rekalibrierung mit Interceptfreisetzung auf die beobachteten Effekte auf den entsprechenden Skalen wurde anhand einer Formel nach Oort (2005a) berechnet. Auf der Skala Rollenfunktionsfähigkeit zeigte sich ein Response-Shift-Effekt in der Höhe eines mittleren Effektes nach Cohen (1988). Auf der Skala Physische Funktionsfähigkeit war der nachgewiesene Response-Shift-Effekt klein. Dies gilt ebenso für die Skala Kognitive Funktionsfähigkeit. Auf allen drei Skalen zeigten sich kleine Effekte durch die Veränderung der GLQ. Der Response-Shift auf den Skalen der Physischen und Rollenfunktionsfähigkeit führte zu einer ‚Überschätzung‘ der wahren Veränderung. Auf der Skala Kognitive Funktionsfähigkeit hingegen ‚überdeckte‘ der Response-Shift eine wahre Veränderung.

## **7.2 Ergebnisse zur Fragestellung 2**

Nachdem im Rahmen der Beantwortung der Fragestellung 1 ein Response-Shift bei der Erfassung der GLQ bei PCa-Patienten in der Anschlussheilbehandlung in der vorliegenden Stichprobe nachgewiesen werden konnte, soll in einem weiteren Auswertungsschritt geprüft werden, ob das Ausmaß der Veränderung der *Globalen Lebensqualität* der Patienten zwischen den beiden MZPen einen Einfluss auf den Response-Shift hat. Zu diesem Zweck wurde die vorliegende Gesamtstichprobe der 212 PCa-Patienten anhand des Medians der Veränderung auf der Skala ‚Globale Lebensqualität‘ des EORTC-QLQ-C30 dichotomisiert und die Response-Shift-Ableitung und Modelltestung für beide Subgruppen getrennt durchgeführt. Aufgrund des geringen Stichprobenumfangs der gebildeten Subgruppen von 106 Personen pro Gruppe werden die Ergebnisse der Auswertungen lediglich als Hinweise für weitere Hypothesengenerierungen betrachtet. Bevor die Ergebnisse der Untersuchung der Variable ‚Veränderung der Globalen Lebensqualität‘ auf den Response-Shift dargestellt werden, wird eine Beschreibung der Substichproben vorgenommen. Eine Dropout-Analyse wird nicht durchgeführt, da alle Teilnehmer der Gesamtstichprobe in die Auswertungen der Fragestellung 2 eingehen.

AMOS 17.0 bietet gute Möglichkeiten für einfache Gruppenvergleiche, allerdings ist aufgrund der Restriktionen verschiedener Modellparameter die simultane Auswertung verschiedener Gruppen methodisch ungenau. Bei einem Simultanvergleich verschiedener Gruppen würden die Restriktionen dazu führen, dass die restringierten Parameter in allen Gruppen den gleichen Wert bekommen (Arbuckle 2005). Diese allgemeine Definition über die Gruppen hinaus beeinflusst jedoch die Ableitung der entsprechenden Varianz-Kovarianz-Matrizen.



Neben verzerrten Modellschätzungen resultieren daraus fehlerhafte Berechnungen zur Kalkulation der Effektgrößen. Aus diesem Grund werden die KFAen für beide Subgruppen getrennt voneinander berechnet.

### 7.2.1 Dichotomisierung der Gesamtstichprobe

Die Dichotomisierung der 212 PCa-Patienten erfolgte anhand des Medians der Skala Globale Lebensqualität. Diese Skala besteht aus den beiden Items: „Wie würden Sie insgesamt Ihren Gesundheitszustand während der letzten Woche einschätzen?“ sowie „Wie würden Sie insgesamt Ihre gesamte Lebensqualität während der letzten Woche einschätzen?“. Die Fragen sind jeweils auf einer siebenstufigen Lickertskala von ‚1‘ (sehr schlecht) bis ‚7‘ (ausgezeichnet) zu beantworten. Aus beiden Werten wird ein Skalenwert der Globalen Lebensqualität berechnet. Die Skala weist zu beiden MZPen eine gute interne Konsistenz auf. Im Folgenden werden die deskriptive Statistik und die Ableitung der Dichotomisierungsvariable beschrieben.

Die Werte zur Globalen Lebensqualität zu beiden MZPen werden in Tabelle 41 vorgestellt. Wie bereits anhand der Auswertungen zum latenten Konstrukt GLQ zu beiden MZPen anhand der KFA in Fragestellung 1 zu erwarten war, unterscheiden sich die Mittelwerte der Skala Globale Lebensqualität zu beiden MZPen statistisch hochsignifikant ( $t = 13,64$ ,  $p < .001$ ). Je höher der Wert, desto positiver ist die Ausprägung der Globalen Lebensqualität, so dass im vorliegenden Fall von einer Verbesserung der Globalen Lebensqualität zum MZP 2 auszugehen ist. Die Werte zu den beiden Messzeitpunkten entsprechen den in der Literatur bekannten Ausprägungen (Aaronson et al. 1993, Borghede & Sullivan 1996, Dirmaier et al. 2004).

Tabelle 41: Angaben zur Globalen Lebensqualität (N = 212)

Variable	M	SD	Min	Max	Md
Globale Lebensqualität MZP 1	49,9	18,5	0,0	100,0	50,0
Globale Lebensqualität MZP 2	68,7	17,0	16,7	100,0	66,7
Globale LQ-Differenz	18,8	20,1	-33,3	83,3	16,7

Anmerkung: M = Mittelwert, SD = Standardabweichung, Min = Minimum, Max = Maximum, Md = Median, LQ = Lebensqualität

Da die Dichotomisierung der Gesamtstichprobe anhand der *Veränderung* der Globalen Lebensqualität zwischen den MZPen erfolgte, wurde für jede Person ein Differenzwert aus

den beiden Einzelwerten berechnet. Aus den so gewonnenen 212 Differenzwerten wurde der Median bestimmt. Dieser lag in der vorliegenden Stichprobe bei 16,7 Skaleneinheiten. Als Vergleichswerte lassen sich die Daten von Dirmaier und Kollegen (2004) heranziehen, die eine heterogene Stichprobe onkologischer Patienten vor und nach der Rehabilitation befragten und aus deren Angaben sich ein Differenzwert von ca. 16 Skaleneinheiten zwischen Prä- und Postmessung ableiten lässt.

Wie beschrieben, lag der Median der Veränderungswerte in der vorliegenden Arbeit bei 16,7 Skaleneinheiten, wobei dieser Wert bei insgesamt 53 Personen berechnet wurde. Da eine annähernd gleiche Verteilung in zwei Subgruppen anhand des Wertes 16,7 nicht möglich war, das heißt nicht alle Personen mit diesem Veränderungswert lediglich einer der beiden Subgruppen zuzuordnen waren, wurden die Personen mit diesem Veränderungswert per Zufall auf beide Gruppen verteilt, bis eine exakte Stichprobenhalbierung erreicht wurde. Insgesamt wurden 31 PCa-Patienten in die Gruppe *Globale LQ-Differenz  $\leq$  Md* gelöst.

Die Werte der Globalen Lebensqualität der beiden Gruppen (Globale Lebensqualität  $\leq$  Median  $\geq$  Globale Lebensqualität) werden in Tabelle 42 vorgestellt. In der Tabelle werden die Werte der Globalen Lebensqualität zu beiden MZP für die beiden Gruppen sowie die mittlere Differenz der Einzelwerte in beiden Gruppen dargestellt. Zu den Differenzwerten sind noch weitere Verteilungsmaße angegeben. Es zeigt sich eine geringe Veränderung der Globalen Lebensqualität in der Gruppe *Globale LQ-Differenz  $\leq$  Md* sowie jeweils Mittelwerte der Globalen Lebensqualität dieser Gruppe zu beiden MZP, die in einem durchschnittlichen Bereich lagen. Für die Gruppe *Globale LQ-Differenz  $\geq$  Md* zeigt sich eine deutliche Veränderung, die um das 10-fache größer ist als in der erstgenannten Gruppe. Die Einzelwerte zur Globalen Lebensqualität waren zum MZP 1 eher unterdurchschnittlich und zum MZP 2 eher überdurchschnittlich. Der Unterschied zwischen den beiden Veränderungswerten ist statistisch hochsignifikant ( $t = 17,1$ ;  $p < .001$ ).

Tabelle 42: Subgruppenbildung und Globale LQ-Differenz

Gruppe	n	M <sub>t1</sub>	M <sub>t2</sub>	M <sub>Diff</sub>	SD	Min	Max
Globale LQ-Differenz $\leq$ Md	106	58,6	61,1	3,5	12,3	-33,3	16,7
Globale LQ-Differenz $\geq$ Md	106	41,2	75,2	34,0	13,7	16,7	83,3

Anmerkung: LQ = Lebensqualität, Md = Median, n = Stichprobenumfang, M<sub>t1</sub> = Mittelwert zum MZP 1, M<sub>t2</sub> = Mittelwert zum MZP 2, M<sub>Diff</sub> = Differenz der Mittelwerte zwischen MZP 1 und MZP 2, SD = Standardabweichung, Min = Minimum, Max = Maximum.

## 7.2.2 Stichprobenbeschreibung

Die beiden Substichproben sollen im Folgenden kurz anhand der wesentlichen soziodemographischen Variablen beschrieben und auf Unterschiede zwischen den Gruppen getestet werden. In beiden Gruppen ist die überwiegende Mehrheit der Patienten verheiratet, deutscher Nationalität, lebt mit einem Partner im Haushalt und hat mehr als ein Kind. Der häufigste Schulabschluss ist in beiden Gruppen die abgeschlossene Hauptschule beziehungsweise Volksschule, die häufigste Berufsausbildung die Lehre. Ebenfalls in beiden Gruppen beziehen mehr als  $\frac{3}{4}$  der Personen Altersrente. Über  $\frac{4}{5}$  der PCa-Patienten in beiden Gruppen waren nach der Reha nicht mehr in einer Therapie. Ebenfalls gilt für beide Gruppen, dass nur bei wenigen Personen erneut ein Tumor nach der AHB entdeckt wurde. Die Überprüfung der Verteilung mittels  $\chi^2$ -Test zeigt lediglich in der Variable *Schulabschluss* einen signifikanten Unterschied zwischen den Gruppen, der aufgrund der Differenzialanalyse vor allem darauf zurückzuführen ist, dass sich in der Gruppe Globale LQ-Differenz  $\geq$  Md mehr Personen mit Abitur befinden. Es lassen sich jedoch keine statistisch signifikanten Unterschiede bezüglich der Verteilung zur Berufsausbildung zwischen den Gruppen nachweisen. Ein Überblick befindet sich in Tabelle 43.

Tabelle 43: Soziodemographische Angaben zur Subgruppenanalyse Globale LQ-Differenz

Variable	Globale LQ-Differenz $\leq$ Md (n = 106)		Globale LQ-Differenz $\geq$ Md (n = 106)	
	n	(%)	n	(%)
<b>Familienstand</b>				
Verheiratet	92	(86,8)	95	(89,6)
Anderer	14	(13,2)	11	(10,4)
Missing	0	(0,0)	0	(0,0)
<b>Nationalität</b>				
Deutsch	105	(99,1)	103	(97,2)
Andere	1	(0,9)	2	(1,9)
Missing	0	(0,0)	1	(0,9)
<b>Personen im Haushalt</b>				
Partner	78	(73,6)	82	(77,4)
Andere	28	(26,4)	22	(20,7)
Missing	0	(0,0)	2	(1,9)
<b>Anzahl der Kinder</b>				
Kein Kind	7	(6,6)	7	(6,6)
1 Kind	33	(31,1)	20	(18,9)
Mehr als 1 Kinder	66	(62,3)	79	(74,5)
Missing	0	(0,0)	0	(0,0)

Fortsetzung nächste Seite

Schulabschluss				
Hauptschule/Volksschule	76	(73,2)	65	(61,3)
Anderer	30	(26,8)	40	(37,8)
Missing	0	(0,0)	1	(0,9)
Berufsausbildung				
Lehre	61	(57,5)	51	(48,1)
Andere	44	(41,6)	51	(48,1)
Missing	1	(0,9)	4	(3,8)
Erwerbstätigkeit				
Altersrente	76	(71,7)	84	(79,2)
Anderes	29	(27,4)	22	(20,7)
Missing	1	(0,9)	0	(0,0)
Weitere Therapie nach der Reha				
Nein	87	(82,2)	93	(87,7)
Ja	17	(15,9)	13	(12,3)
Missing	2	(1,9)	0	(0,0)
Erneutes Auftreten eines Tumors nach der AHB				
Nein.	98	(92,5)	100	(94,3)
Ja	2	(1,9)	1	(0,9)
Weiß nicht	1	(0,9)	1	(0,9)
Missing	5	(4,7)	4	(3,8)

Anmerkungen: LQ = Lebensqualität, Md = Median

Insgesamt ist aufgrund der vorliegenden Ergebnisse von einer Vergleichbarkeit der Gruppen auszugehen. Das Durchschnittsalter liegt für die Gruppe Globale LQ-Differenz  $\leq$  Md bei 66,4 Jahren (SD = 5,2, Min = 53, Max = 83) und für die Gruppe Globale LQ-Differenz  $\geq$  Md bei 65,7 (SD = 6,2, Min = 43, Max = 81). Der Altersunterschied zwischen beiden Gruppen ist statistisch nicht signifikant.

Zusätzlich werden für beide Subgruppen die Werte der Indikatoren der GLQ zum MZP 1 und MZP 2 dargestellt und die Unterschiede auf statistische Signifikanz getestet. Die Angaben sind in Tabelle 44 dargestellt. Es zeigt sich, dass sich die beiden Subgruppen hinsichtlich der Mittelwerte der Indikatoren (mit Ausnahme der Physischen Funktionsfähigkeit) der GLQ zum MZP 1 statistisch signifikant unterscheiden. Die Werte weisen darauf hin, dass die Gruppe mit deutlichen Veränderungen der Globalen Lebensqualität zum MZP 1 die Funktionsfähigkeiten insgesamt deutlich schlechter beurteilt. Da zu der Gesamtstichprobe keine Daten über den objektiven Gesundheitszustand oder Angaben über die Art der Akutbehand-

lung vorliegen, kann über den Hintergrund dieser Unterschiede keine weiteren Aussagen getroffen werden. Zum MZP 2 unterscheiden sich die Werte der einzelnen Indikatoren nicht mehr statistisch signifikant. Dies ist ein Hinweis darauf, dass sich die Personen beider Subgruppen bezüglich der Funktionsfähigkeiten deutlich annähern. Jedoch soll darauf hingewiesen werden, dass die p-Werte zum MZP 2 zur Überprüfung der Unterschiede zwischen den Indikatorwerten der Skala Rollenfunktionsfähigkeit bei  $p = .054$  und der Skala Physische Funktionsfähigkeit bei  $p = .052$  liegen.

Tabelle 44: Subgruppen Globale LQ-Differenz und Indikatoren

Gruppe	Indikator									
	SF		RF		PF		EF		CF	
	M <sub>t1</sub>	M <sub>t2</sub>	M <sub>t1</sub>	M <sub>t2</sub>	M <sub>t1</sub>	M <sub>t2</sub>	M <sub>t1</sub>	M <sub>t2</sub>	M <sub>t1</sub>	M <sub>t2</sub>
Globale LQ-Differenz $\leq$ Md	65,6	69,7	50,3	67,3	75,5	81,9	70,1	73,1	86,9	83,0
Globale LQ-Differenz $\geq$ Md	56,8	75,8	33,5	73,9	72,6	85,7	61,7	76,7	77,4	83,2
p-Wert	<.05	n.s.	<.01	n.s.	n.s.	n.s.	<.05	n.s.	<.01	n.s.

Anmerkung: LQ = Lebensqualität, Md = Median, SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, M<sub>t1</sub> = Mittelwert zum MZP 1, M<sub>t2</sub> = Mittelwert zum MZP 2, n.s. = nicht signifikant.

### 7.2.3 Überprüfung der Modellpassung in beiden Gruppen

Die anhand des Medians der Differenzwerte der Skala Globale Lebensqualität gebildeten Substichproben weisen jeweils einen Umfang von 106 Personen auf. Ein solch geringer Stichprobenumfang ist für die Durchführung der KFA als kritisch anzusehen. Nach Kline (2005) ist davon auszugehen, dass gute Voraussetzungen zur Berechnung einer KFA bei einem Mindeststichprobenumfang von  $n > 200$  liegt. Der Autor fügt jedoch hinzu, dass ein Stichprobenumfang von  $100 > n < 200$  als ausreichend für die Durchführung einer KFA angesehen werden kann, wobei es einerseits auf die Anzahl im Modell zu bestimmender Parameter ankommt sowie die Güte der Schätzer, unabhängig von der Passung aufgrund der Fitmaße. Für Stichprobenumfänge kleiner 200 schlägt Garson (2008) den RMSEA und den CFI als weniger anfällige Fitmaße vor.

Ein anderer Zugang zur Bestimmung der Stichprobengröße orientiert sich an der Anzahl zu berechnender Parameter im Modell (freier Parameter). Nach Kline (2005) sollte das Verhältnis zwischen Versuchspersonen und zu bestimmenden Parametern 5:1 nicht unterschrei-

ten. Für die Ableitung des Basismodells an der gesamten Stichprobe ist diese Forderung erfüllt. Dies gilt jedoch nicht für die Auswertungen der Substichproben bezüglich der Nebenfragestellungen. Da es jedoch bei der Überprüfung des Response-Shifts in den Subgruppen nicht im klassischen Sinne um die Ableitung eines neuen Modells geht, sondern um den Vergleich der Modellpassung zwischen zwei Gruppen, werden die Berechnungen dennoch durchgeführt. Insbesondere kann das Basismodell als gut passend bezeichnet werden, da es einerseits in der Ableitungsstichprobe gute Fitmaße aufweist und diese durch die Gesamtstichprobe der 212 PCa-Patienten bestätigt wurde. Ebenfalls für die Akzeptanz der vorliegenden Stichprobenumfänge spricht das Vorgehen, theoriegeleitet geschachtelte Modelle zu überprüfen, d.h. entsprechend der Überprüfung der Response-Shift-Parameter verschiedene Modellalternativen miteinander und die Veränderung der Passungen anhand der Fitmaße zu vergleichen.

Marsh und Hau (1999) berichten über die Beziehungen zwischen der Stichprobengröße und verschiedenen anderen Parametern bei Strukturgleichungsmodellen und Konfirmatorischen Faktorenanalysen. Die Autoren gehen unter anderem davon aus, dass der Grundsatz „je mehr, desto besser“ bezüglich der Stichprobengröße zwar wünschenswert, jedoch nicht immer notwendig ist. So ist nachweisbar, dass die Güte von Modellen sowie die Präzision von Parameterschätzungen einerseits mit zunehmender Stichprobengröße steigen, andererseits jedoch auch mit der Anzahl der Freiheitsgrade innerhalb eines Modells. Daraus leiten die Autoren ab, dass bei kleinen Stichproben eine möglichst hohe Anzahl an Freiheitsgraden kompensierend wirken kann. In der vorliegenden Untersuchung wurde die Zahl der Freiheitsgrade durch die Einführung der Parameterrestriktionen erhöht, was sich günstig auf die Berechnung der Fitmaße auswirkt. Marsh und Kollegen (1996) konnten darüber hinaus nachweisen, dass die Wahrscheinlichkeit der Passung von theoretischen Modellen und empirischen Daten neben der Stichprobengröße auch von dem Verhältnis zwischen der Anzahl der Indikatoren pro Faktor abhängt. Grundsätzlich wird ein Verhältnis von 3 (Indikatoren) zu 1 (Faktor) als Mindestmaß postuliert. Anhand von Simulationsstudien konnten Marsh und Kollegen (1996) nachweisen, dass selbst ein Stichprobenumfang von  $n = 50$  mit steigendem Verhältnis Indikatoren/Faktor ausreichend sein kann, um gute Schätzungen zu liefern, die sich je nach Anzahl der Indikatoren nur geringfügig von größeren Stichproben unterscheiden. Marsh und Hau (1999) bestätigen diesen Zusammenhang an einer weiteren Simulationsstudie. Bei einem Verhältnis von fünf Indikatoren pro Faktor ist der vorliegenden Stichprobenumfang von 106 Patienten pro Gruppe und einer vorher überprüften Modellpassung an zwei Stichproben mit

212 und 185 Personen akzeptabel. Aus denselben Gründen wird auch für die Fragestellung 3.1 bis 3.4 von einer kritischen, jedoch verwertbaren Stichprobengröße ausgegangen.

Im Folgenden werden die Ausprägungen der Güte-Indizes für das Basismodell und das Response-Shift-Modell für beide Gruppen angegeben. Anschließend wird die Modellableitung für beide Gruppen auf Unterschiede geprüft. Auf eine Ausreißeranalyse sowie auf die Überprüfung von Schiefe und Kurtosis wird aufgrund des explorativen Charakters der Fragestellungen verzichtet.

Die Ausprägungen der Güteindizes in Tabelle 45 weisen insgesamt auf eine akzeptable bis gute Passung zwischen modelltheoretischen Annahmen und empirischen Daten hin. Kritisch ist die Passung des Basismodells in der Gruppe Globale LQ Differenz  $\leq$  Md.

Tabelle 45: Modellpassung nach Globale LQ-Differenz

Variable	Gruppe	Modell	Fit-Maße			
			$\chi^2/df$	RMSEA	CFI	TLI
Globale LQ	Differenz $\leq$ Md	Basismodell	1.90	.09 [.05;.13]	.95	.90
		RS-Modell	1.70	.08 [.04;.12]	.95	.93
	Differenz $\geq$ Md	Basismodell	1.04	.02 [.00;.08]	.98	.99
		RS-Modell	1.21	.04 [.00;.09]	.98	.97

Anmerkung: LQ = Lebensqualität, Md = Median, RS = Response-Shift,  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Die berechneten Fitmaße zeigen für beide Subgruppen eine unterschiedliche Passung an. Die Ausprägung der Indizes für die Gruppe Globale LQ-Differenz  $\leq$  Md weist auf eine akzeptable Passung des Basismodells in der zugrunde liegenden Subgruppe hin, wobei die Ausprägungen der einzelnen Maße unterschiedliche Interpretationen zulassen. Der  $\chi^2/df$  zeigt eine gute Passung zwischen Modell und empirischer Kovarianzmatrix an. Der RMSEA hingegen weist einen grenzwertig inakzeptablen Fit auf (RMSEA  $>$  .08 = inakzeptabel). Die beiden Indizes CFI und TLI weisen wiederum auf eine grenzwertige Akzeptanz des Modells hin. In der Gruppe Globale LQ-Differenz  $\geq$  Md sind die Indexwerte durchweg als gut zu bezeichnen. Das theoretisch definierte Basismodell scheint die empirisch gefundene Kovarianzmatrix gut zu beschreiben. Trotz einer grenzwertigen Passung des Basismodells in der erstgenannten Gruppe wird separat für beide Gruppen die Response-Shift Testung wiederholt. Bezüglich des Response-Shift-Modells, das im Rahmen der Beantwortung der Fragestellung 1 abgeleitet

wurde, zeigt sich in der Gruppe Globale LQ-Differenz  $\leq$  Md eine gute Passung. In der Vergleichsgruppe weisen die Fitmaße ebenfalls auf eine gute Passung hin. Auffällig ist jedoch, dass das Basismodell in der zuletzt genannten Gruppe numerisch bessere Indexwerte aufweist als das Response-Shift-Modell. Eine mögliche Ursache kann darin liegen, dass die Response-Shift-Prozesse innerhalb der Gesamtstichprobe nicht auf beide Subgruppen gleichermaßen übertragen werden können. Somit würden bei der Verwendung des Response-Shift-Modells aus Fragestellung 1 Fehlkalkulationen vorgenommen. Um dies zu überprüfen, werden gruppenspezifische Response-Shift-Ableitungen vorgenommen.

#### 7.2.4 Beziehung zwischen der Veränderung der Globalen Lebensqualität und Response-Shift

Im Folgenden werden die Freisetzungen der Parameterrestriktionen vom Basismodell bis zum Response-Shift-Modell entsprechend dem Ablauf in der Gesamtstichprobe für beide Subgruppen getrennt überprüft. Es soll getestet werden, ob die in der Gesamtstichprobe identifizierten Parameterfreisetzungen und Restriktionen des Response-Shift-Modells auch auf beide Subgruppen übertragbar sind. Auch wenn die Fitmaße des Response-Shift-Modells der beiden Subgruppen akzeptabel bis gut sind, wird jede einzelne Restriktion nochmals genauer betrachtet. In der Tabelle 46 sind die Ergebnisse der Überprüfung dargestellt.

Tabelle 46: Vergleich der geschachtelten Modelle mit dem Response-Shift-Modell nach Gruppe Globale LQ-Differenz  $\leq$  Md vs.  $>$ Md

Geschachtelte Modelle	df	Response-Shift-Modell (Modell 8 – e_CF)			
		Gruppe Globale LQ-Differenz $\leq$ Md		Gruppe Globale LQ-Differenz $\geq$ Md	
		CMIN	p	CMIN	p
‚Null-Modell‘	8	57,92	< .001	91,68	< .001
Modell 1 – IN-RF	7	48,76	< .001	77,08	< .001
Modell 2 – e_RF	6	37,34	< .001	58,11	< .001
Modell 3 – e_SF	5	36,88	< .001	40,74	< .001
Modell 4 – e_PF	4	24,10	< .001	27,52	< .001
Modell 5 – e_EF	3	23,74	< .001	12,57	< .01
Modell 6 – IN_PF	2	11,58	< .01	4,53	.10
Modell 7 – IN_CF	1	1,36	.244	1,68	.20

Anmerkungen: Md = Median, IN = Intercept, e = Messfehler, LQ = Lebensqualität, Md = Median, df = Freiheitsgrade, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit



Die Ergebnisse lassen erkennen, dass die in der Gesamtstichprobe identifizierten Parameterfreisetzungen auf beide Gruppen nicht vollständig übertragbar sind. Das in Kapitel 7.1 abgeleitete Response-Shift-Modell unterscheidet sich in der Gruppe Globale LQ-Differenz  $\leq$  Md nicht statistisch signifikant vom Modell 7 – IN\_CF (Intercept Kognitive Funktionsfähigkeit). Das heißt, der Parameter ‚Messfehler Kognitive Funktionsfähigkeit‘ (ursprünglich Modell 8 – e\_CF) kann in dieser Subgruppe restringiert bleiben. In der Gruppe Globale LQ-Differenz  $\geq$  Md ergibt die Response-Shift-Überprüfung eine noch schlechtere Übereinstimmung. Bereits die Restriktion der Parameter ‚Intercept Physische Funktionsfähigkeit‘ (Modell 6 – IN\_PF) ist aufrecht zu halten. Nach diesem Vorgehen ist in dieser Subgruppe sogar auf weniger Parametern ein Response-Shift identifizierbar.

Da jedoch jede einzelne Veränderung innerhalb der Kovarianzmatrix – z. B. durch die Aufhebung einer Parameterrestriktion – die Berechnung aller übrigen Parameter beeinflusst, wird für beide Gruppen nochmals eine gruppenspezifische Response-Shift-Ableitung vorgenommen. Die Ergebnisse dazu werden in Tabelle 47 dargestellt.

Tabelle 47: Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Globale LQ-Differenz

Gruppe	Ausprägung	Fit-Maße				
		$\chi^2/df$	RMSEA	CFI	TLI	
Globale Lebensqualität	Differenz $\leq$ Md	Nullmodell	2.87	.13 [.10;.16]	.83	.80
		Modell 1 – IN_CF	2.60	.12 [.09;.15]	.86	.83
		Modell 2 – IN_RF	2.42	.12 [.09;.15]	.88	.85
		Modell 3 – IN_PF	2.22	.11 [.08;.14]	.90	.87
		Modell 4 – e_RF	1.94	.10 [.06;.13]	.92	.90
		Modell 5 – e_PF	1.61	.08 [.04;.11]	.95	.94
	Differenz $\geq$ Md	Nullmodell	3.36	.15 [.12;.18]	.67	.60
		Modell 1 – e_EF	3.05	.14 [.11;.17]	.72	.66
		Modell 2 – e_PF	2.86	.13 [.10;.16]	.75	.69
		Modell 3 – e_RF	2.47	.12 [.09;.15]	.81	.75
		Modell 4 – e_SF	2.11	.10 [.07;.14]	.86	.82
		Modell 5 – FL_RF	1.70	.08 [.04;.12]	.92	.88
		Modell 6 – IN_RF	1.29	.05 [.00;.08]	.97	.95
		Modell 7 – IN_PF	1.11	.03 [.00;.08]	.99	.98

Anmerkung: Md = Median, IN = Intercept, e = Messfehler, FL = Faktorladung, LQ = Lebensqualität, Md = Median, RS = Response-Shift,  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Die gruppenspezifische Response-Shift-Ableitung weist auf deutliche Unterschiede im Response-Shift zwischen den beiden untersuchten Gruppen hin. In der Gruppe Globale LQ-Differenz  $\leq$  Md findet auf insgesamt fünf Parametern ein Response-Shift statt. Das Response-Shift-Modell (Modell 5 – e\_PF) dieser Subgruppe weist folgende Parameterfreisetzungen auf: Die Intercepts der Skalen Physische, Kognitive und Rollenfunktionsfähigkeit sowie die beiden Messfehler der Skalen Physische und Rollenfunktionsfähigkeit. In der Gruppe Globale LQ-Differenz  $\geq$  Md zeigen sich insgesamt sieben Parameter durch einen Response-Shift beeinflusst. In dieser Gruppe sind im Response-Shift-Modell (Modell 7 – IN\_PF) folgende Parameter freizusetzen: die Messfehlervarianzen der Skalen Emotionale, Physische, Soziale und Rollenfunktionsfähigkeit, die Faktorladung der Skala Rollenfunktionsfähigkeit sowie die beiden Intercepts der Skalen Physische und Rollenfunktionsfähigkeit. Besonders hervorzuheben ist, dass die Freisetzung der Faktorladung der Skala Rollenfunktionsfähigkeit auf eine Repriorisierung oder Neukonzeptualisierung hinweist (Modell 5 – FL\_RF).

Der Vergleich der Modelle untereinander wird für jede der beiden Gruppen getrennt berechnet und die Unterschiede zwischen den gruppenspezifischen Response-Shift-Modellen und den jeweiligen ‚Vorläufermodellen‘ auf statistische Bedeutsamkeit geprüft. Die Ergebnisse sind in Tabelle 48 enthalten.

Tabelle 48: Vergleich der geschachtelten Modelle nach Gruppe Globale LQ-Differenz  $\leq$  Md vs.  $>$ Md

Gruppe Globale LQ-Differenz $\leq$ Md			Gruppe Globale LQ-Differenz $\geq$ Md		
Geschachtelte Modelle	CMIN	p	Geschachtelte Modelle	CMIN	p
‚Null-Modell‘	55,64	< .001	‚Null-Modell‘	98,52	< .001
Modell 1 – IN-CF	42,96	< .001	Modell 1 – e_EF	78,38	< .001
Modell 2 – IN_RF	33,86	< .001	Modell 2 – e_PF	68,46	< .001
Modell 3 – IN_PF	24,76	< .001	Modell 3 – e_RF	52,02	< .001
Modell 4 – e_RF	12,69	< .001	Modell 4 – e_SF	37,29	< .001
Modell 5 – e_PF			Modell 5 – FL_RF	21,79	< .001
			Modell 6 – IN_RF	7,00	< .01
			Modell 7 – IN_PF		

Anmerkungen: Md = Median, IN = Intercept, e = Messfehler, FL = Faktorladung; RS = Response-Shift, df = Freiheitsgrade, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, LQ = Lebensqualität

Die Ergebnisse zeigen, dass zwischen den beiden untersuchten Gruppen quantitative und qualitative Unterschiede in der Form des Response-Shifts bestehen. Einerseits weist die Gruppe Globale LQ-Differenz  $\geq$  Md mehr Parameterfreisetzungen auf. Andererseits unterscheidet sich durch die aufgedeckte Repriorisierung beziehungsweise Neukonzeptualisierung auch die Art des Response-Shifts zwischen den Gruppen. Diese Ergebnisse weisen darauf hin, dass ein Response-Shift in der Gruppe Globale LQ-Differenz  $\geq$  Md in ausgeprägterer Form stattfindet.

Im Weiteren werden alle Modell-Parameter gruppenbezogen betrachtet. Tabelle 49 gibt eine Gegenüberstellung der Werte in beiden Subgruppen wieder.

Tabelle 49: Schätzung der Response-Shift-Parameter nach Globale LQ-Differenz

	Globale LQ-Differenz $\leq$ Md					Globale LQ-Differenz $\geq$ Md				
	Faktorladungen									
	Korrelationen mit GLQ		Kovarianzen mit GLQ		Korrelationen mit GLQ		Kovarianzen mit GLQ			
	Prä	Post	Prä	Post	Prä	Post	Prä	Post	Prä	Post
PF	.46	.58	8.34	8.34	.41	.54	7.37	7.37		
RF	.51	.66	17.70	17.70	<b>.23</b>	<b>.73</b>	<b>7.23</b>	<b>17.41</b>		
SF	.76	.76	21.94	21.94	.63	.81	19.21	19.21		
EF	.77	.77	17.84	17.84	.57	.70	15.01	15.01		
CF	.52	.52	9.22	9.22	.42	.40	8.84	8.84		
-----										
	Intercepts									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>75.47</b>	<b>50.36</b>	65.66	70.03	<b>86.95</b>	<b>72.64</b>	<b>33.70</b>	57.35	62.31	76.06
Posttest	<b>80.47</b>	<b>64.12</b>	65.66	70.03	<b>81.38</b>	<b>78.65</b>	<b>57.35</b>	57.35	62.31	76.06
-----										
	Residual-Varianzen									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>261.22</b>	<b>894.20</b>	346.67	214.35	226.72	<b>263.48</b>	<b>976.99</b>	<b>567.11</b>	<b>469.39</b>	358.98
Posttest	<b>134.24</b>	<b>384.24</b>	346.67	214.35	226.72	<b>118.87</b>	<b>244.81</b>	<b>174.02</b>	<b>206.99</b>	358.98

Anmerkungen: LQ = Lebensqualität, Md = Median, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität, fett = Parameterfreisetzung

Während in der Gruppe Globale LQ-Differenz  $\leq$  Md Faktorladungen beziehungsweise Kovarianzen zwischen dem latenten Konstrukt GLQ und den einzelnen Skalen restringiert bleiben, ist in der Gruppe Globale LQ-Differenz  $\geq$  Md die Faktorladung der Skala Rollenfunktionsfähigkeit freizusetzen. Während die Faktorladung mit einem Wert von .23 zum MZP

1 auf eine geringe Bedeutung der Skala Rollenfunktionsfähigkeit für das latente Konstrukt GLQ hinweist, besteht zum MZP 2 eine enge Beziehung zwischen dem latenten Konstrukt und der Skala. Die Veränderung der Faktorladung vom MZP 1 zum MZP 2 ist als grenzwertig zwischen einer Repriorisierung und einer Neukonzeptualisierung zu bewerten. Aufgrund theoretischer Annahmen wird im vorliegenden Fall von einer Repriorisierung gesprochen, da die Skala RF auch zu MZP 1 bereits von der GLQ beeinflusst wurde. Hinsichtlich der Intercepts der Prätestwerte (die gleichzeitig auch die beobachteten Mittelwerte der jeweiligen Skalen darstellen) fällt auf, dass diese in der Gruppe Globale LQ-Differenz  $\geq$  Md zum Teil deutlich unter den Werten der Vergleichsgruppe liegen. Dieses Ergebnis weist darauf hin, dass es sich bei den PCa-Patienten in dieser Gruppe um Personen handelt, die sich zu Beginn der AHB deutlich eingeschränkter beurteilten als die Personen in der Gruppe Globale LQ-Differenz  $\leq$  Md.

Der Darstellung ist ebenfalls zu entnehmen, dass in der Gruppe Globale LQ-Differenz  $\geq$  Md häufiger eine non-uniforme Rekalibrierung stattfindet (auf vier von fünf Skalen), dafür jedoch die Interceptfreisetzung in der Gruppe Globale LQ-Differenz  $\leq$  Md einmal häufiger auftritt und im Falle der Skala Kognitive Funktionsfähigkeit auch unabhängig ist von einer non-uniformen Rekalibrierung. In beiden Gruppen ist davon auszugehen, dass die Interceptveränderungen auf den Skalen Physische und Rollenfunktionsfähigkeit zu einem großen Teil durch die Varianzveränderung des spezifischen Faktors der Skalen mit verursacht sind. In der Gruppe Globale LQ-Differenz  $\leq$  Md treten somit insgesamt drei Response-Shift-Prozesse auf: Zwei non-uniforme Rekalibrierungen (mit Interceptbeeinflussung) auf den Skalen Physische und Rollenfunktionsfähigkeit sowie eine uniforme Rekalibrierung auf der Skala Kognitive Funktionsfähigkeit. In der Gruppe Globale LQ-Differenz  $\geq$  Md finden fünf Response-Shift-Prozesse statt: Vier non-uniforme Rekalibrierungen (davon zwei mit Interceptbeeinflussung) und eine Repriorisierung. *Die unterschiedliche Anzahl der Response-Shift-Prozesse in den beiden Gruppen spricht für die formulierte Hypothese.*

Abschließend soll vergleichend für beide Gruppen berechnet werden, in welchem Umfang sich die Response-Shift-Effekte auf das Ausmaß der beobachteten Effekte auswirken. Tabelle 50 zeigt die Aufteilung der beobachteten Effekte in Response-Shift-Effekte und Effekte aufgrund der 'wahren Veränderung' der GLQ auf den entsprechenden Skalen. In der Gruppe Globale LQ-Differenz  $\leq$  Md werden mittlere Effekte auf den Skalen Rollen- und Physische Funktionsfähigkeit beobachtet, wobei diese beobachteten Effekte in überwiegendem

Maße durch einen Response-Shift bedingt sind. Das Ausmaß der Veränderung durch die GLQ ist eher zu vernachlässigen. Auf der Skala Kognitive Funktionsfähigkeit zeigt sich ebenfalls ein eher kleiner Response-Shift-Effekt. Der Effekt durch die Veränderung der GLQ auf dieser Skala ist ebenfalls eher unbedeutend.

Tabelle 50: Globale LQ-Differenz als Einflussfaktor auf Response-Shift-Effekte

Skala	Effektstärke					
	LQ-Differenz $\leq$ Md			LQ-Differenz $\geq$ Md		
	beobachtet	Response-Shift	‘wahre Veränderung’	beobachtet	Response-Shift	‘wahre Veränderung’
SF	0.16	-----	0.16	0.65	-----	0.65
RF	<b>0.49</b>	<b>0.40</b>	<b>0.09</b>	<b>1.24</b>	<b>1.03<sup>+</sup></b>	<b>0.21</b>
PF	<b>0.44</b>	<b>0.34</b>	<b>0.10</b>	<b>0.77</b>	<b>0.36</b>	<b>0.41</b>
EF	0.13	-----	0.13	0.64	-----	0.64
CF	<b>-0.23</b>	<b>-0.33</b>	<b>0.10</b>	0.26	-----	0.26

Anmerkung: Rek. = Rekalibrierung; SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit; fett = Skalen mit Response-Shift Einfluss; <sup>+</sup> = Response-Shift setzt sich zusammen aus Rekalibrierung und Repriorisierung

In der Gruppe Globale LQ-Differenz  $\geq$  Md ist die Ausprägung der Response-Shift-Effekte eher heterogen. Auf beiden Skalen (Rollen- und Physische Funktionsfähigkeit) zeigen sich große Effekte auf der Ebene der beobachteten Effekte, wobei der beobachtete Effekt auf der Skala Rollenfunktionsfähigkeit nochmals deutlich größer ist als der auf der Skala Physische Funktionsfähigkeit. Ist der große Effekt auf der Skala Physische Funktionsfähigkeit in etwa zu gleichen Teilen auf einen Response-Shift sowie einen Anteil der Veränderung durch die GLQ zurückzuführen, stellt sich das Verhältnis auf der Skala Rollenfunktionsfähigkeit deutlich anders dar. Der beobachtete große Effekt kommt vor allem durch einen großen Response-Shift-Effekt zustande, der sich zusammensetzt aus einem großen Effekt bedingt durch eine non-uniforme Rekalibrierung (ES = 0.73) sowie einem kleinen Effekt durch eine Repriorisierung (ES = 0.30). Das Ausmaß der Veränderung durch die GLQ liegt für diese Skala im Bereich eines kleinen Effektes.

Aufgrund der angegebenen Größe der Effekte ist lediglich der Effekt der (non-)uniformen Rekalibrierung auf der Skala Rollenfunktionsfähigkeit als klinisch bedeutsam (nach Norman et al. 2003) anzusehen. Alle anderen Response-Shift-Effekte erreichen dieses Kriterium nicht, unabhängig davon, in welcher Subgruppe sie aufgedeckt werden.

Abschließend ist anzumerken, dass der kalkulierte Response-Shift-Effekt auf der Skala Rollenfunktionsfähigkeit als Hinweis gesehen werden kann, dass ein Response-Shift durch das Ausmaß der Veränderungen der Globalen Lebensqualität beeinflusst wird. Die Response-Shift-Effekte auf der Skala Physische Funktionsfähigkeit sind vom Ausmaß in beiden Subgruppen jedoch vergleichbar. Ebenfalls gegen die Hypothese spricht die uniforme Rekalibrierung auf der Skala Kognitive Funktionsfähigkeit, die lediglich in der Gruppe Globale LQ-Differenz  $\leq$  Md auftritt. *Die berichteten Angaben zu den Effekten ergeben keine klaren Hinweise auf die Bestätigung der Hypothese, dass mit zunehmender Differenz zwischen den Werten der Globalen Lebensqualität zum MZP 1 und MZP 2 ein Response-Shift umso deutlicher auftritt.*

Im Folgenden werden die statistischen Ergebnisse des Gruppenvergleichs inhaltlich kurz beschrieben. In der Gruppe Globale LQ-Differenz  $\leq$  Md bedeutet die non-uniforme Rekalibrierung auf den Skalen Physische und Rollenfunktionsfähigkeit eine Veränderung der Bewertung der entsprechenden Skalen hin zu einer positiveren Bewertung der Funktionsfähigkeiten, ohne dass diese durch die GLQ bedingt ist. Die uniforme Rekalibrierung der Skala Kognitive Funktionsfähigkeit weist den gegenläufigen Trend auf: Die befragten Männer beurteilen ihre kognitive Leistungsfähigkeit zum MZP 2 schlechter, ohne dass dieser Prozess in einem Zusammenhang mit der Veränderung der GLQ der Personen steht. Die non-uniforme Rekalibrierung auf den Skalen Physische und Rollenfunktionsfähigkeit weist darauf hin, dass einzelne Skalenooptionen von den Befragten umbewertet werden.

In der Gruppe Globale LQ-Differenz  $\geq$  Md bedeutet die Veränderung der Faktorladung der Skala Rollenfunktionsfähigkeit, dass diese zum MZP 2 durch die GLQ stärker beeinflusst wird als zum MZP 1. Anders formuliert heißt das: Zum MZP 1 wird die Rollenfunktionsfähigkeit lediglich zu einem marginalen Anteil durch die GLQ bedingt, der jedoch zum MZP 2 deutlich zunimmt. Die (non-)uniforme Rekalibrierung der Skalen Physische und Rollenfunktionsfähigkeit besagt eine positivere Bewertung dieser Funktionsfähigkeiten zum MZP 2 im Vergleich zum MZP 1, ohne dass diese auf die GLQ zurückzuführen ist. Die non-uniforme Rekalibrierung auf den Skalen Physische, Emotionale, Soziale und Rollenfunktionsfähigkeit kennzeichnet eine uneinheitliche Veränderung der Skalierung dieser Indikatoren in der vorliegenden Substichprobe.

### 7.2.5 Zusammenfassung der Ergebnisse zur Fragestellung 2

Mit den Auswertungen zur Fragestellung 2 der vorliegenden Arbeit wurde geprüft, ob das Ausmaß an Veränderung der globalen Lebensqualität einen Einfluss auf das Ausmaß eines Response-Shifts hat.

In einem ersten Schritt wurde die vorliegende Stichprobe anhand des Medians der individuellen Differenzwerte zwischen der Globalen Lebensqualität zum MZP 1 und MZP 2, gemessen mit Hilfe der gleichnamigen Skala des EORTC-QLQ-C30, in zwei Gruppen mit je 106 Personen eingeteilt. Die Gruppen unterschieden sich im Umfang der Differenzwerte statistisch hochsignifikant.

Aufgrund der mangelnden Übertragbarkeit des unter Fragestellung 1 abgeleiteten Response-Shift-Modells auf beide Subgruppen wurden für diese jeweils spezifische Response-Shift-Ableitungen vorgenommen. Es zeigten sich Unterschiede zwischen den beiden Gruppen in dem Sinne, dass in der Gruppe Globale LQ Differenz  $\leq$  Md auf weniger Parametern mittels KFA ein Response-Shift identifiziert werden konnte, sowie dem Ergebnis, dass in der Gruppe der Personen mit großen Veränderungen der Globalen Lebensqualität neben Rekalibrierungsprozessen eine Repriorisierung aufgedeckt wurde.

In der Gruppe Globale LQ Differenz  $\leq$  Md zeigten sich Response-Shift-Prozesse auf folgenden Parametern: den Intercepts der Skalen Rollen-, Physische und Kognitive Funktionsfähigkeit sowie in den Messfehlervarianzen der Skalen Physische und Rollenfunktionsfähigkeit. In der Gruppe Globale LQ Differenz  $\geq$  Md wurde ein Response-Shift auf den Intercepts der Skalen Rollen- und Physische Funktionsfähigkeit identifiziert sowie auf allen Messfehlervarianzen mit Ausnahme der Skala Kognitive Funktionsfähigkeit. Auffällig ist die Repriorisierung, die zusätzlich auf der Skala Rollenfunktionsfähigkeit festgestellt wurde.

Ein eindeutiger Einfluss der Veränderung der globalen Lebensqualität auf die Größe der Response-Shift-Effekte konnte nicht festgestellt werden. In beiden Subgruppen war die Größe des Response-Shift-Effekts auf der Skala Physische Funktionsfähigkeit in etwa vergleichbar und lag im mittleren Bereich. Auf der Skala Rollenfunktionsfähigkeit zeigte sich in der Gruppe Globale LQ Differenz  $\geq$  Md jedoch ein größerer Response-Shift-Effekt als in der Vergleichsgruppe.

Zusätzlich sprechen für die Annahme, dass das Ausmaß der Veränderung der globalen Lebensqualität einen Einfluss auf das Ausmaß eines Response-Shift hat, die Ergebnisse, dass mittels KFA in der Gruppe Globale LQ Differenz  $\geq$  Md *mehr Response-Shift-Parameter* freizusetzen waren und *mehr Response-Shift-Prozesse* nachgewiesen werden konnten. Das Ergebnis der Differenzierungen der beobachteten Effekte lässt keinen eindeutigen Schluss zu. Hier weisen die Ergebnisse nur teilweise darauf hin, dass Unterschiede im Ausmaß des Response-Shift in Abhängigkeit von der Veränderung der globalen Lebensqualität bestehen.

### 7.3 Ergebnisse zur Fragestellung 3

Ausgehend vom Response-Shift-Modell nach Sprangers und Schwartz (1999) und den im Kontext dieser Arbeit aufgedeckten Response-Shift-Prozessen, wird in weiteren Analysen überprüft, ob es moderierende Variablen gibt, die das Ausmaß des Response-Shift beeinflussen. Dazu wurde ausgehend vom Response-Shift-Modell eine Auswahl an Merkmalen und Mechanismen getroffen, die im Zusammenhang mit Response-Shift-Phänomenen als relevant angesehen werden. Zu diesen Variablen zählen Benefit Finding, Selbstwirksamkeit sowie positive und negative Affektivität. Das Vorgehen zur Überprüfung der Moderatorvariablen entspricht in etwa dem Vorgehen zur Auswertung der Fragestellung 2. Ein Unterschied liegt in der Gruppenbildung, da der Median der jeweils untersuchten Variable eine Verteilung der Gesamtgruppe in zwei Substichproben ermöglichte, ohne dass – wie zur Fragestellung 2 – eine zufallsbedingte Zuordnung einer größeren Anzahl Studienteilnehmer in eine der beiden Subgruppen erfolgen musste. Daraus resultieren jedoch jeweils zwei ungleich große Vergleichsstichproben pro Moderatoranalyse.

Um den Einfluss der vier Variablen auf den Response-Shift zu überprüfen, wurde die zugrunde liegende Stichprobe anhand des Medians der entsprechenden Variablen in jeweils zwei Subgruppen unterteilt. Für jede Subgruppe wurde anschließend eine Einzelberechnung durchgeführt, so dass insgesamt acht gruppenspezifische Response-Shift-Ableitungen vorgenommen wurden, um Aussagen über die Wirkung der Zielvariablen zu treffen.

Die Analysen und Auswertungen der Nebenfragestellungen sind aufgrund der methodischen Kritizierbarkeit der Ausgangsbedingungen (geringe Stichprobengröße) lediglich als deskriptive Ergebnisse zu verstehen.



### 7.3.1 Beschreibung der Substichproben zur Auswertung der Fragestellung 3

Aufgrund der unterschiedlichen Anzahl fehlender Werte in den einzelnen Moderatorvariablen schwankt die Größe der zugrundeliegenden Stichproben zwischen  $n = 166$  und  $n = 208$ . Im Folgenden werden die einzelnen Substichproben (je Moderatorvariable) in ihren wesentlichen Charakteristika beschrieben. Anschließend wird eine Dropoutanalyse für jede Moderatorvariable vorgenommen.

#### *Beschreibung der Substichproben*

Insgesamt 166 Personen machten ausreichend Angaben auf der Skala Benefit Finding. Die überwiegende Mehrheit dieser Patienten ist verheiratet, deutscher Nationalität, lebt gemeinsam mit einer Partnerin im Haushalt und hat mehr als ein Kind. Etwas weniger als  $2/3$  der Personen beendete die Schule nach der Volksschule, wobei etwa die Hälfte der Substichprobe eine Lehre absolvierte. Etwa  $3/4$  der Patienten beziehen Altersrente. Ca. 14% der Personen unterzogen sich nach der Rehabilitationsmaßnahme einer weiteren Therapie, wobei 94% angaben, dass bei ihnen kein neuer Tumor diagnostiziert wurde. Die angegebenen Häufigkeiten treffen in vergleichbarem Maße auch auf die drei anderen Moderatorvariablen zu, so dass auf deren genauere Darstellung verzichtet werden kann. Die gesamten Angaben sind den Tabellen 51 und 52 zu entnehmen.

Tabelle 51: Soziodemographische Angaben zu den jeweiligen Substichproben - Moderatoren

Variable	BF (n = 166)		SW (n = 166)		PA (n = 207)		NA (n = 208)	
	n	(%)	n	(%)	n	(%)	N	(%)
<b>Familienstand</b>								
Verheiratet	146	(88,0)	146	(88,0)	182	(87,9)	183	(88,0)
Anderer	20	(12,0)	20	(12,0)	25	(12,1)	25	(12,0)
Missing	0	(0,0)	0	(0,0)	0	(0,0)	0	(0,0)
<b>Nationalität</b>								
Deutsch	162	(97,6)	162	(97,6)	203	(98,1)	204	(98,1)
Andere	3	(1,8)	3	(1,8)	3	(1,4)	3	(1,4)
Missing	1	(0,6)	1	(0,6)	1	(0,5)	1	(0,5)
<b>Personen im Haushalt</b>								
Partner	127	(76,5)	127	(76,5)	155	(74,9)	156	(75,0)
Andere	38	(22,9)	38	(22,9)	50	(24,1)	50	(24,0)
Missing	1	(0,6)	1	(0,6)	2	(1,0)	2	(1,0)

Fortsetzung nächste Seite

Anzahl der Kinder									
Kein Kind	10	(6,0)	11	(6,6)	14	(6,8)	14	(6,7)	
1 Kind	40	(24,1)	40	(24,1)	51	(24,6)	52	(25,0)	
Mehr als 1 Kinder	116	(69,9)	115	(69,3)	142	(68,6)	142	(68,3)	
Missing	0	(0,0)	0	(0,0)	0	(0,0)	0	(0,0)	
Schulabschluss									
Hauptschule/Volksschule	99	(59,6)	99	(59,6)	127	(61,4)	128	(61,5)	
Anderer	66	(39,8)	66	(39,8)	79	(38,1)	79	(38,0)	
Missing	1	(0,6)	1	(0,6)	1	(0,5)	1	(0,5)	
Berufsausbildung									
Lehre	86	(51,8)	86	(51,8)	107	(51,7)	108	(51,9)	
Andere	76	(45,8)	76	(45,8)	95	(45,9)	95	(45,7)	
Missing	4	(2,4)	4	(2,4)	5	(2,4)	5	(2,4)	
Erwerbstätigkeit									
Altersrente	124	(74,7)	124	(74,7)	157	(75,8)	157	(75,5)	
Anderes	42	(25,3)	42	(25,3)	49	(23,7)	50	(24,0)	
Missing	0	(0,0)	0	(0,0)	1	(0,5)	1	(0,5)	
Weitere Therapie nach der Reha									
Nein	143	(86,1)	144	(86,7)	178	(86,0)	178	(85,6)	
Ja	21	(12,7)	20	(12,1)	25	(12,1)	26	(12,5)	
Missing	2	(1,2)	2	(1,2)	4	(1,9)	4	(1,9)	
Erneutes Auftreten eines Tumors nach der AHB									
Nein.	156	(94,0)	156	(94,0)	195	(94,2)	195	(93,8)	
Ja	3	(1,8)	3	(1,8)	3	(1,4)	3	(1,4)	
Weiß nicht	1	(0,6)	1	(0,6)	2	(1,0)	2	(1,0)	
Missing	6	(3,6)	6	(3,6)	7	(3,4)	8	(3,8)	

Anmerkungen: BF = Benefit Finding, SW = Selbstwirksamkeit, PA = Positive Affektivität, NA = Negative Affektivität, n = Stichprobenumfang

Tabelle 52: Angaben zum Alter der jeweiligen Substichproben - Moderatoren

Variable	M	SD	Min	Max
Benefit Finding (n = 166)	65,9	5,7	43	79
Selbstwirksamkeit (n = 166)	65,9	5,7	43	79
Positive Affektivität (n = 207)	66,1	5,7	43	83
Negative Affektivität (n = 208)	66,0	5,7	43	83

Anmerkung: n = Stichprobenumfang, M = Mittelwert, SD = Standardabweichung, Min = Minimum, Max = Maximum

Auffällig ist, dass zu den Variablen Benefit Finding und Selbstwirksamkeit die gleiche Anzahl von Personen verwertbare Angaben machten. Der Umfang der Substichproben zu den Moderatorvariablen der Affektivität ist nahezu gleich, jedoch sind die Umfänge der beiden Gruppen deutlich höher als die zu den erstgenannten Moderatorvariablen. Die Subgruppen zu den Moderatorvariablen Benefit Finding und Selbstwirksamkeit haben 45 gemeinsame Dropouts. Die Gruppen der Variablen positive und negative Affektivität haben vier gemeinsame Dropouts. Eine mögliche Erklärung für diese Verteilungen kann darin gesehen werden, dass die PANAS als Instrument beide Skalen (Positive und Negative Affektivität) enthält und die beiden anderen Skalen (Benefit Finding und Selbstwirksamkeit) als zwei separate Instrumente am Schluss des Fragebogenpakets aufgeführt wurden. Eine Erklärung für die unterschiedliche Anzahl der Dropouts pro Skala ist möglicherweise die abnehmende Motivation der befragten Patienten, alle Fragen beziehungsweise Fragebögen komplett auszufüllen.

### *Dropout-Analyse*

Im Folgenden wird für jede Moderatorvariable eine Dropoutanalyse dargestellt. Als Dropout wird ein Fall bezeichnet, der eine zu hohe Anzahl fehlender Werte auf der entsprechenden Skala aufweist, so dass der Skalenwert nicht berechnet wurde. Als kritischer Wert für die Anzahl fehlender Werte pro Skala wurde eine 50%-Missing-Quote angesetzt. Der Ausschluss der Personen erfolgte skalenbezogen, so dass für jede Skalenauswertung die maximale Anzahl an verfügbaren Daten (Patienten) verwendet wurde.

Für die Variablen Benefit Finding und Selbstwirksamkeit ergeben sich jeweils 46 Ausfälle. Für die Variable positive Affektivität werden fünf Dropouts identifiziert, für die Variable negative Affektivität sind es vier. Die Ergebnisse der Dropoutanalyse bezüglich der soziodemographischen Variablen werden in Tabelle 53 vorgestellt. Als statistisches Verfahren zur Prüfung der Unterschiede zwischen den Gruppen wurde der  $\chi^2$ -Test durchgeführt. Es zeigten sich lediglich hinsichtlich der Moderatorvariablen positive Affektivität statistisch signifikante Unterschiede zwischen den Teilnehmern der Substichprobe und den Dropouts. Die Dropouts erhielten nach der Reha häufiger eine Strahlentherapie (14 von 207 Personen in der Substichprobe vs. 2 von 5 Personen bei den Dropouts). Weiterhin hatten die Personen der Dropouts relativ gesehen häufiger ‚4 Kinder‘ als die Teilnehmer der Substichprobe (8 von 207 Personen in der Substichprobe vs. 2 von 5 Personen bei den Dropouts). Der zuletzt genannte Unterschied gilt auch für die Dropoutanalyse zur Moderatorvariable negative Affek-

tivität. Insgesamt zeigen sich in den Dropoutanalysen jedoch keine gravierenden Unterschiede, wobei vor allem bezüglich der Moderatorvariablen positive und negative Affektivität darauf hinzuweisen ist, dass nur sehr wenige Personen als Dropouts bezeichnet wurden und somit die Vergleichbarkeit stark eingeschränkt ist.

Tabelle 53: Dropoutanalyse zu den jeweiligen Substichproben - Moderatoren

Variable	BF		SW		PA		NA	
	n (166)	DO (46)	n (166)	DO (46)	n (207)	DO (5)	n (208)	DO (4)
Familienstand	n.s.		n.s.		n.s.		n.s.	
Nationalität	n.s.		n.s.		n.s.		n.s.	
Personen im Haushalt	n.s.		n.s.		n.s.		n.s.	
Anzahl der Kinder	n.s.		n.s.		p < .05		p < .05	
Schulabschluss	n.s.		n.s.		n.s.		n.s.	
Berufsausbildung	n.s.		n.s.		n.s.		n.s.	
Erwerbstätigkeit	n.s.		n.s.		n.s.		n.s.	
Therapie n. der Reha	n.s.		n.s.		p < .01		n.s.	
Erneutes Auftreten eines Tumors nach der AHB	n.s.		n.s.		n.s.		n.s.	
Indikatoren der GLQ	n.s.		n.s.		n.s.		n.s.	
SF	n.s.		n.s.		n.s.		n.s.	
PF	n.s.		n.s.		n.s.		n.s.	
RF	n.s.		n.s.		n.s.		n.s.	
EF	n.s.		n.s.		n.s.		n.s.	
CF	n.s.		n.s.		n.s.		n.s.	

Anmerkungen: BF = Benefit Finding, SW = Selbstwirksamkeit, PA = Positive Affektivität, NA = Negative Affektivität, DO = Dropout, n = Anzahl der Personen in der Stichprobe, (Anzahl der Dropouts), SF = Soziale Funktionsfähigkeit, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, n.s. = nicht statistisch signifikant

Eine weitere Dropoutanalyse wurde anhand der Ausprägung der Skalenwerte der Indikatorvariablen der GLQ durchgeführt. Es wurden die Werte der Skalen Soziale, Physische, Emotionale, Kognitive und Rollenfunktionsfähigkeit zwischen den Teilnehmern und den Dropouts verglichen. Die Ergebnisse sind ebenfalls in Tabelle 53 dargestellt. Es zeigten sich in keiner Moderatorvariablen statistisch signifikante Unterschiede zwischen den Gruppen. Somit ist auch hinsichtlich der Ausprägung der Indikatoren von keiner Selektion beziehungsweise Verzerrung aufgrund von Dropouts auszugehen.

Auf eine tabellarische Darstellung der Dropoutanalyse zur Variable Alter wird verzichtet. Es zeigten sich in allen vier Moderatorvariablen keine signifikanten Unterschiede zwischen den Personen der Substichproben und den Dropouts.

### *Beschreibung der Subgruppen anhand der Indikatorvariablen*

Nachdem die verschiedenen Subgruppen zur Überprüfung der Response-Shift-Prozesse anhand der soziodemographischen Merkmale vorgestellt wurden, sollen diese im Folgenden nochmals anhand der Moderatoren und der Indikatorvariablen des EORTC-QLQ-C30 beschrieben werden.

Wie im vorangegangenen Kapitel dargestellt, reduzierten sich die Substichproben mit verwertbaren Angaben in den Moderatorvariablen Benefit Finding und Selbstwirksamkeit beträchtlich. Daraus ergab sich im weiteren Verlauf der Gruppenbildung durch Dichotomisierung am Median der jeweiligen Variable zum Teil sehr kleine Stichprobenumfänge. Diese sind für alle Subgruppen in Tabelle 54 dargestellt.

Tabelle 54: Ausprägung der Moderatorvariablen in den Subgruppen

Variable	Subgruppe	n	M	SD	Min.	Max.
Benefit Finding	≤ Median	81	3,07	0,53	1,3	3,7
	> Median	85	4,38	0,38	3,8	5,0
Selbstwirksamkeit	≤ Median	88	2,88	0,29	1,7	3,2
	> Median	78	3,66	0,26	3,3	4,0
PANAS – Positive Affektivität	≤ Median	105	2,69	0,38	1,6	3,2
	> Median	102	3,79	0,41	3,3	5,0
PANAS – Negative Affektivität	≤ Median	109	1,18	0,14	1,0	1,4
	> Median	99	1,94	0,45	1,5	3,7

Anmerkung: n = Stichprobenumfang, M = Mittelwert, SD = Standardabweichung, Min = Minimum, Max = Maximum

Ebenfalls angegeben in der Tabelle sind die durchschnittlichen Werte und Standardabweichungen der jeweiligen Moderatorvariablen. Aus der Tabelle wird ersichtlich, dass es sich bei den Mittelwerten der Subgruppen nicht um Extremwerte handelt, sondern dass in beiden Gruppen die Werte der jeweiligen Moderatorvariablen zwar unterschiedlich, jedoch nicht hinsichtlich einer größtmöglichen Differenz ausgeprägt sind. Bei einer genügend großen Stichprobe wäre der Vergleich von Extremgruppen (bspw. 1. und 4. Quartil oder andere Perzentile

im Bereich der positiven und negativen Polung der Skala) wünschenswert gewesen. Aufgrund der Stichprobengröße konnte auf diesen Zugang zur Moderatoranalyse jedoch nicht zurückgegriffen werden. Den Daten der Tabelle 54 ist zu entnehmen, dass hinsichtlich der Variablen Benefit Finding (BF) die inhaltliche Ausprägung der Werte zwischen der Einschätzung ‚mittelmäßig‘ (3,07) und ‚ziemlich‘ (4,38) variiert. In beiden Gruppen findet somit ein Benefit Finding in nicht geringem Maße statt. Ähnliches gilt für die Subgruppenauswertungen der anderen Moderatorvariablen.

Selbstwirksamkeitserwartungen sind in beiden Gruppen ausgeprägt vorhanden (‚stimmt eher‘ (2,88) vs. ‚stimmt genau‘ (3,66)). In beiden Gruppen erleben die Patienten positive Affekte (‚einigermaßen‘ (2,69) vs. ‚erheblich‘ (3,79)) sowie lediglich ein geringes Maß an negativen Affekten (‚gar nicht‘ (1,18) vs. ‚ein bisschen‘ (1,94)). Zwar unterscheiden sich alle Subgruppen hinsichtlich der Ausprägungen der Moderatorvariablen statistisch hochsignifikant voneinander, jedoch besitzen die jeweiligen Subgruppen immer die gleiche Tendenz bezüglich der Ausprägung der Mittelwerte. Die Ähnlichkeit der Werte der Moderatorvariablen läßt wiederum geringere Unterschiede in den Response-Shift-Prozessen und Response-Shift-Effekten erwarten.

In einem weiteren Schritt werden die Ausprägungen der Indikatorvariablen innerhalb der jeweiligen Subgruppen zu beiden MZPen beschrieben. Dabei werden einmal die Unterschiede innerhalb einer Gruppe zu beiden MZPen (Intragruppenvergleich) dargestellt. Somit kann die Veränderung der Indikatorvariablen pro Gruppe betrachtet werden. Gleichzeitig wird ein Vergleich zwischen den Werten der jeweiligen Vergleichsgruppen zu einem MZP möglich, um auf Unterschiede zwischen den Subgruppen hinzuweisen (Intergruppenvergleich). Die Angaben werden in Tabelle 55 dargestellt.

Hervorzuheben ist an diesen Ergebnissen, dass sich die Werte innerhalb der Subgruppen bei jeder Moderatorvariablen mit Ausnahme der Kognitiven Funktionsfähigkeit (Benefit-Finding, Selbstwirksamkeit, positive Affektivität) sowie der Emotionalen Funktionsfähigkeit (negative Affektivität, Gruppe > Md) zwischen den MZPen bei jedem Prä-Post-Vergleich statistisch signifikant unterscheiden. Das heißt in beinahe jeder Subgruppe steigen die Werte bei jedem subgruppenbezogenen Prä-Post-Vergleich statistisch signifikant an. Die Veränderung der Werte sprechen insgesamt für eine Besserung der Funktionsfähigkeiten in den Vergleichsgruppen.

Tabelle 55: Mittelwerte und Standardabweichungen der Indikatorvariablen in den Subgruppen

Gruppe	n	SF <sub>1</sub>	SF <sub>2</sub>	PF <sub>1</sub>	PF <sub>2</sub>	RF <sub>1</sub>	RF <sub>2</sub>	EF <sub>1</sub>	EF <sub>2</sub>	CF <sub>1</sub>	CF <sub>2</sub>
BF ≤ Median	81	<b>53,5</b> (28,1)	68,9 (26,3)	72,5 (19,7)	83,1 (14,2)	37,4 (33,2)	66,3 (26,3)	61,6 (23,4)	72,1 (22,6)	79,2 (23,5)	81,5 (18,3)
BF > Median	85	<b>65,1</b> (29,8)	74,5 (24,5)	74,4 (17,8)	83,5 (14,6)	41,8 (33,8)	71,4 (24,5)	67,7 (25,3)	77,2 (19,3)	83,9 (19,5)	84,9 (18,5)
SW ≤ Median	88	55,9 (29,0)	<b>64,6</b> (25,7)	71,5 (17,0)	<b>81,0</b> (13,5)	38,2 (33,1)	<b>62,1</b> (26,5)	62,1 (23,6)	<b>69,0</b> (23,0)	79,4 (20,8)	<b>79,5</b> (19,7)
SW > Median	78	64,5 (28,8)	<b>79,9</b> (22,7)	75,6 (20,2)	<b>86,0</b> (14,9)	41,2 (34,1)	<b>76,3</b> (21,9)	68,2 (24,9)	<b>81,3</b> (16,5)	84,2 (22,3)	<b>87,4</b> (15,9)
PA ≤ Median	105	57,0 (29,6)	<b>62,9</b> (27,0)	70,1 (16,9)	78,9 (13,4)	37,0 (33,3)	<b>63,8</b> (26,6)	<b>61,1</b> (23,9)	<b>67,8</b> (22,1)	79,8 (21,1)	<b>78,3</b> (19,5)
PA > Median	102	<b>66,0</b> (29,5)	<b>82,2</b> (19,2)	<b>78,0</b> (19,0)	<b>88,6</b> (13,2)	<b>47,7</b> (35,3)	<b>77,0</b> (21,0)	<b>71,0</b> (25,0)	<b>82,3</b> (18,1)	84,4 (20,9)	<b>87,4</b> (17,0)
NA ≤ Median	109	<b>66,8</b> (29,4)	<b>79,8</b> (24,0)	<b>76,3</b> (19,0)	85,9 (14,2)	<b>46,2</b> (35,5)	<b>75,2</b> (23,2)	<b>73,3</b> (22,1)	<b>86,4</b> (14,4)	<b>86,1</b> (19,0)	<b>90,1</b> (14,2)
NA > Median	99	55,4 (29,1)	<b>63,8</b> (24,4)	<b>71,3</b> (17,2)	81,1 (13,6)	<b>37,9</b> (33,1)	<b>64,8</b> (25,4)	<b>57,8</b> (25,2)	<b>62,1</b> (20,7)	<b>77,1</b> (22,4)	<b>74,7</b> (20,0)

Anmerkungen: BF = Benefit Finding, SW = Selbstwirksamkeit, PA = Positive Affektivität, NA = Negative Affektivität, n = Anzahl der Personen in der Stichprobe, SF = Soziale Funktionsfähigkeit, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, <sub>1</sub> = MZP 1, <sub>2</sub> = MZP 2, kursiv = statistisch signifikanter Unterschied zwischen zwei Messzeitpunkten; fett = statistisch signifikanter Unterschied zwischen zwei Vergleichsgruppen zu einem Messzeitpunkt

Hinsichtlich des Subgruppenvergleichs aufgrund der Moderatorvariable Benefit Finding zeigt sich, dass mit Ausnahme der Indikatorvariable Soziale Funktionsfähigkeit zum MZP 1 keine statistisch signifikanten Unterschiede zwischen den beiden Subgruppen zu den jeweiligen MZP vorliegen. Bezüglich des Subgruppenvergleichs anhand der Variable Selbstwirksamkeit zeigt sich jeweils zum MZP 2 ein statistisch signifikanter Gruppenunterschied. Hierbei ist jedoch darauf hinzuweisen, dass die Personen der Gruppe Selbstwirksamkeit > Md jeweils tendenziell positivere Einschätzungen zum MZP 1 aufweisen, sich also dieser positive Trend bis zum MZP 2 weiter verstärkt. Der Subgruppenvergleich anhand der Moderatorvariablen positive und negative Affektivität zeigt mit zwei Ausnahmen (CF<sub>1</sub> positive Affektivität und PF<sub>2</sub> negative Affektivität) bei jedem Vergleich zu jedem MZP statistisch signifikante Unterschiede zwischen den Gruppen. D.h. die Personen mit einer geringeren positiven Affektivität beurteilen ihre Funktionsfähigkeiten zum MZP 1 (mit Ausnahme von CF<sub>1</sub>) statistisch bedeutsam negativer und zeigen einen schwächeren Trend in Richtung einer positiven Veränderung auf den Indikatoren. Das Gleiche gilt für den Subgruppenvergleich auf der Moderatorvariablen negative Affektivität (mit Ausnahme von PF<sub>2</sub>), wobei die Gruppe mit einer ausgeprägteren negativen Affektivität ihre Funktionsfähigkeiten jeweils schlechter beurteilten.

### 7.3.2 Überprüfung der Modellpassung in den einzelnen Substichproben

Trotz der geringen Stichprobenumfänge der einzelnen Untersuchungsgruppen wurden verschiedene Fitmaße zur Beurteilung der Passung zwischen den empirischen Daten und den postulierten theoretischen Modellen berechnet. Auf eine Ausreißeranalyse sowie auf die Überprüfung von Schiefe und Kurtosis wird aufgrund des explorativen Charakters der Fragestellungen verzichtet. Die Ausprägungen der Güteindizes des Basismodells in Tabelle 56 weisen auf eine akzeptable bis gute Passung zwischen modelltheoretischen Annahmen und empirischen Daten hin.

Tabelle 56: Passung des Basismodells bei verschiedenen Substichproben

Variable	Ausprägung	Fit-Maße			
		$\chi^2/df$	RMSEA	CFI	TLI
Benefit Finding	≤ Median	1.45	.07 [.00;.13]	.96	.92
	> Median	1.10	.04 [.00;.10]	.99	.98
Selbstwirksamkeit	≤ Median	1.25	.05 [.00;.11]	.97	.95
	> Median	1.03	.02 [.00;.09]	1.00	1.00
PANAS – Positive Affektivität	≤ Median	1.22	.05 [.00;.10]	.98	.97
	> Median	1.19	.04 [.00;.09]	.98	.97
PANAS – Negative Affektivität	≤ Median	1.29	.05 [.00;.10]	.98	.96
	> Median	1.00	.00 [.00;.08]	1.00	1.01

Anmerkung:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Außerdem wurde die Passung des Response-Shift-Modells für alle Subgruppen nochmals geprüft. Die Auswertungen zeigen für alle Subgruppen eine gute bis akzeptable Passung. Ziel der vorliegenden Analyse ist der Vergleich von Parameterschätzungen zwischen Gruppen bei gleichem Messmodell. Veränderungen von Beziehungen innerhalb des Modells werden der Vergleichbarkeit untergeordnet. Die Ergebnisse sind in Tabelle 57 dargestellt.



Tabelle 57: Passung des Response-Shift-Modells bei verschiedenen Substichproben

Variable	Ausprägung	Fit-Maße			
		$\chi^2/df$	RMSEA	CFI	TLI
Benefit Finding	≤ Median	1.33	.06 [.00;.13]	.96	.94
	> Median	1.00	.00 [.00;.08]	1.00	.99
Selbstwirksamkeit	≤ Median	1.21	.05 [.00;.10]	.97	.96
	> Median	.99	.00 [.00;.08]	1.00	1.00
PANAS – Positive Affektivität	≤ Median	1.19	.04 [.00;.09]	.98	.97
	> Median	1.21	.04 [.00;.09]	.98	.96
PANAS – Negative Affektivität	≤ Median	1.35	.06 [.00;.10]	.97	.95
	> Median	.90	.00 [.00;.07]	1.00	1.02

Anmerkung:  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

### 7.3.3 Beziehung zwischen Moderatorvariablen und Response-Shift

Trotz guter Passung zwischen modelltheoretischen Annahmen und den empirischen Daten wird in einem ersten Schritt überprüft, ob die in der Gesamtstichprobe identifizierten Response-Shift-Prozesse für alle vorliegenden Substichproben in gleicher Weise gelten. Aus diesem Grund wurde für jede Substichprobe eine gruppenspezifische Response-Shift-Ableitung vorgenommen (wie in Fragestellung 2). Finden in den einzelnen Subgruppen Response-Shift-Prozesse nicht in der gleichen Weise wie in der Gesamtstichprobe statt, würde eine Berechnung der Effektstärkeparameter entsprechend der Ergebnisse der Gesamtstichprobe zu verzerrten Response-Shift-Kalkulationen führen. Dies wäre einerseits darauf zurückzuführen, dass durch überflüssige Parameterfreisetzungen Response-Shift-Effekte berechnet würden, die für die jeweilige Subgruppe nicht von Bedeutung sind. Andererseits beeinflusst jede Freisetzung von Parametern die Berechnung der gesamten Kovarianzmatrix, so dass bei Fehlspezifikationen Verzerrungen in der Parameterausprägung resultieren würden. Letztlich ist dem Ansatz des Vergleichs der Modellpassung zwischen geschachtelten Modellen das Modell zu bevorzugen, das am sparsamsten mit Parameterfreisetzungen umgeht. Je mehr Response-Shift-Parameter restringiert bleiben, umso spezifischer ist das Modell definiert. Ziel dieses Vorgehens ist, dass das Modell als Berechnungsbasis zur Effektgrößenkalkulation herangezogen wird, das die höchste Parameter- und Modellspezifikation aufweist.

Zur Beurteilung der Effektstärkeparameter wird einerseits auf die Einteilung der Effektstärke nach Cohen (1988) Bezug genommen. Ein Wert für ES  $\sim 0.2$  wird als kleiner Effekt, ein Wert von  $\sim 0.5$  als mittlerer und ein Wert von  $\sim 0.8$  als großer Effekt bezeichnet. Andererseits wird die Beurteilung von Effekten nach Norman und Kollegen (2003) verwendet, wonach Effekte ab ES  $\sim 0.5$  als klinisch bedeutsam bezeichnet werden. Die Differenzierung der Effekte in ‚wahre‘ und Response-Shift-Effekte erfolgt anhand der Formel von Oort (2005a).

### *Fragestellung 3.1: Response-Shift – Ableitung nach Ausmaß des Benefit Finding*

Welchen Einfluss hat *Benefit Finding* auf einen Response-Shift im Zusammenhang mit der Erfassung der GLQ bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Im Folgenden werden zuerst die gruppenspezifischen Parameterfreisetzungen in Abhängigkeit von der Ausprägung der *Moderatorvariable Benefit Finding (BF)* mit ihren Auswirkungen auf die Modellpassung dargestellt. Wie bereits erwähnt wurden zwei Subgruppen gebildet, die sich aus der Berechnung und Einteilung aufgrund des Medians der Skala BF ergaben. Es zeigt sich, dass in beiden Subgruppen Response-Shift-Prozesse auftreten, diese jedoch voneinander abweichen.

Die Passung des Response-Shift-Modells in der Gruppe  $BF \leq Md$  weist zwei Parameterfreisetzung auf der Ebene der Intercepts sowie drei auf der Ebene der Messfehlervarianzen auf. In der Vergleichsgruppe  $BF > Md$  sind drei Intercepts sowie vier Messfehlervarianzen freizusetzen. *Somit sind in der Gruppe mit einer höheren Ausprägung an Benefit Finding mehr Response-Shift-Parameter freizusetzen als in der Vergleichsgruppe.* Die beiden Response-Shift-Modelle unterscheiden sich auch inhaltlich: In der Gruppe  $BF \leq Md$  ist die Messfehlervarianz der Skala Kognitive Funktionsfähigkeit freizusetzen; in der Gruppe  $BF > Md$  sind die Messfehlervarianzen der Skalen Emotionale und Soziale Funktionsfähigkeit freizusetzen. Außerdem ist die Restriktion des Intercepts der Skala Kognitive Funktionsfähigkeit in dieser Gruppe nicht haltbar. Die Ableitungen sind in Tabelle 58 aufgeführt.

Tabelle 58: Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß des Benefit Finding

Gruppe	Ausprägung	Fit-Maße			
		$\chi^2/df$	RMSEA	CFI	TLI
BF $\leq$ Md	Nullmodell	2.46	.14 [.10;.17]	.79	.75
	Modell 1 – IN_RF	2.04	.11 [.08;.15]	.85	.82
	Modell 2 – e_PF	1.73	.10 [.05;.14]	.90	.88
	Modell 3 – e_CF	1.65	.09 [.05;.13]	.91	.89
	Modell 4 – e_RF	1.41	.07 [.00;.12]	.95	.93
	Modell 5 – IN_PF	1.33	.06 [.00;.11]	.96	.94
BF > Md	Nullmodell	2.76	.15 [.11;.18]	.76	.72
	Modell 1 – IN_RF	2.34	.13 [.09;.16]	.82	.79
	Modell 2 – e_EF	2.01	.11 [.07;.15]	.87	.84
	Modell 3 – e_RF	1.71	.09 [.05;.13]	.91	.89
	Modell 4 – e_SF	1.43	.07 [.01;.11]	.95	.93
	Modell 5 – IN_CF	1.28	.06 [.00;.11]	.97	.96
	Modell 6 – e_PF	1.16	.04 [.00;.10]	.98	.98
	Modell 7 – IN_PF	1.03	.02 [.00;.08]	.99	.99

Anmerkung: BF = Benefit Finding, e = Messfehler, IN = Intercept, SF = Soziale Funktionsfähigkeit, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, Md = Median, RS = Response-Shift,  $\chi^2/df$  = Discrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Den vorliegenden Angaben zu Folge finden in der Gruppe BF > Md mehr Response-Shift-Prozesse statt, was als eine Bestätigung der Hypothese gesehen wird. Eine vergleichende Darstellung der Parameterfreisetzungen und -restriktionen in den beiden Gruppen bietet Tabelle 59. Wie bereits aus der Modellableitung ersichtlich, findet in beiden Gruppen kein Response-Shift auf der Ebene der Faktorladungen beziehungsweise Kovarianzen zwischen dem latenten Konstrukt GLQ und den einzelnen Skalen statt. Alle weiteren Angaben zu den berechneten Parametern der Subgruppenanalyse sind der Tabelle 59 zu entnehmen.

Tabelle 59: Schätzung der Response-Shift-Parameter – Benefit Finding

	Benefit Finding $\leq$ Md					Benefit Finding $>$ Md				
	Faktorladungen									
	Korrelationen mit GLQ		Kovarianzen mit GLQ		Korrelationen mit GLQ		Kovarianzen mit GLQ			
	Prä	Post	Prä	Post	Prä	Post	Prä	Post	Prä	Post
PF	.46	.62	8.99	8.99	.48	.58	8.19	8.19		
RF	.49	.62	16.16	16.16	.51	.71	17.33	17.33		
SF	.80	.79	22.16	22.16	.63	.80	19.22	19.22		
EF	.70	.69	16.13	16.13	.63	.80	15.73	15.73		
CF	.34	.40	7.67	7.67	.54	.54	10.06	10.06		
	Intercepts									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>72.51</b>	<b>37.69</b>	53.92	61.56	77.38	<b>74.43</b>	<b>41.63</b>	64.21	68.35	<b>83.92</b>
Posttest	<b>77.21</b>	<b>55.47</b>	53.92	61.56	77.38	<b>78.93</b>	<b>61.76</b>	64.21	68.35	<b>79.35</b>
	Residual-Varianzen									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>305.02</b>	<b>848.73</b>	269.95	273.68	<b>456.37</b>	<b>225.27</b>	<b>859.50</b>	<b>557.59</b>	<b>377.01</b>	249.07
Posttest	<b>119.42</b>	<b>400.25</b>	269.95	273.68	<b>273.68</b>	<b>132.28</b>	<b>288.35</b>	<b>204.21</b>	<b>141.88</b>	249.07

Anmerkungen: Md = Median, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität, fett = Parameterfreisetzung

Die Response-Shift-Prozesse in beiden Gruppen lassen sich inhaltlich wie folgt interpretieren: Da in der Gruppe  $BF \leq Md$  in den Indikatoren Physische und Rollenfunktionsfähigkeit eine uniforme *und* eine non-uniforme Rekalibrierung stattgefunden hat, ist davon auszugehen, dass die uniforme Rekalibrierung, d.h. die Freisetzung der Intercepts, eine Folge der Varianzveränderungen des spezifischen Faktors der jeweiligen Skalen ist. In diesem Sinne ist anzunehmen, dass auf den genannten Skalen keine uniforme, sondern eine non-uniforme Rekalibrierung stattgefunden hat. In der Gruppe  $BF \leq Md$  gilt hinsichtlich der non-uniformen Rekalibrierung mit Interceptbeeinflussung, dass die befragten Personen ihre Physische und Rollenfunktionsfähigkeit zum MZP 2 positiver bewerten als zum MZP 1, ohne dass diese Umbewertung durch die Veränderung der GLQ erklärt werden kann. Die non-uniforme Rekalibrierung auf den Skalen Physische, Kognitive und Rollenfunktionsfähigkeit weist auf eine Veränderung einzelner Skalenabstufungen der jeweiligen Indikatoren hin. Teilbereiche der Skalen werden zum MZP 2 quantitativ anders gewertet als zum MZP 1.

In der Gruppe  $BF > Md$  ist die Rekalibrierung auf den Skalen Physische und Rollenfunktionsfähigkeit aufgrund der Messfehlerfreisetzungen dieser Skalen als non-uniform zu

bezeichnen. Die Messfehlerfreisetzung auf den Skalen Physische, Soziale, Emotionale und Rollenfunktionsfähigkeit weist darauf hin, dass die befragten Personen Teile der Skalenabstufungen quantitativ anders bewerten, ohne dass diese Veränderung durch die Veränderung der GLQ bedingt ist. Somit ist auch für diese Subgruppe davon auszugehen, dass die Freisetzung der Intercepts der Skalen Physische und Rollenfunktionsfähigkeit eine Folge der deutlichen Varianzänderung in den spezifischen Faktoren dieser Skalen ist, also nicht von einer einheitlichen Veränderung der Skalierung der zugrunde liegenden Bewertungsstufen auszugehen ist. Dies gilt jedoch nicht für die Skala Kognitive Funktionsfähigkeit. Hier findet eine uniforme Rekalibrierung statt, so dass für diesen Indikator davon auszugehen ist, dass sich deren gesamte Skalierung der Bewertungsstufen gleichmäßig verschoben hat. Auffallend ist, dass diese Veränderung interner Standards in der Weise stattfand, dass zum MZP 2 der (objektiv) gleiche Zustand (subjektiv) schlechter bewertet wurde als zum MZP 1.

In einem weiteren Schritt soll geklärt werden, welchen Einfluss das Ausmaß an BF auf die identifizierten Response-Shift-Prozesse hat. In Tabelle 60 wird die Differenzierung der beobachteten Effekte auf den entsprechenden Skalen dargestellt.

In der Gruppe  $BF \leq Md$  ergibt sich ein großer beobachteter Effekte auf der Skala Rollenfunktionsfähigkeit, wobei dieser durch einen mittleren Response-Shift-Effekt und einen (eher) kleinen Effekt bedingt durch die wahre Veränderung der GLQ verursacht wird. Auf der Skala Physische Funktionsfähigkeit zeigt sich ein mittlerer beobachteter Effekt, der durch zwei kleine Effekte aus Response-Shift und wahrer Veränderung der GLQ resultiert.

In der Gruppe  $BF > Md$  treten in den Skalen Physische und Rollenfunktionsfähigkeit in etwa vergleichbare beobachtete Effekte auf, wie in der Vergleichsgruppe  $BF \leq Md$ . Auf der Skala Kognitive Funktionsfähigkeit zeigt sich ein zu vernachlässigender beobachteter Effekt, wobei dieser jedoch durch einen kleinen Response-Shift-Effekt bedingt ist, der durch einen ebenfalls kleinen Effekt aufgrund der wahren Veränderung der GLQ überdeckt wird.

Auf der Skala Rollenfunktionsfähigkeit zeigt sich in beiden Gruppen in gleichem Ausmaß ein Response-Shift-Effekt, der nach Norman und Kollegen (2003) als klinisch bedeutsam einzuschätzen ist. Die anderen aufgedeckten Response-Shift-Effekte erreichten in ihrem Ausmaß dieses Kriterium nicht.

Insgesamt legen die Ergebnisse zur Differenzierung der beobachteten Effekte in Response-Shift- und Wahre-Veränderungs-Effekte in den untersuchten Gruppen den Schluss nahe, dass das Ausmaß eines Response-Shift-Effektes auf einer der untersuchten Skalen nicht durch das Ausmaß an Benefit Finding beeinflusst wird.

Tabelle 60: Benefit Finding als Einflussfaktor auf Response-Shift

Skala	Effektstärke					
	Benefit Finding $\leq$ Md			Benefit Finding $>$ Md		
	beobachtet	Response-Shift	‘wahre Veränderung’	beobachtet	Response-Shift	‘wahre Veränderung’
SF	0.59	-----	0.59	0.30	-----	0.30
RF	<b>0.94</b>	<b>0.59</b>	<b>0.35</b>	<b>0.82</b>	<b>0.56</b>	<b>0.26</b>
PF	<b>0.58</b>	<b>0.26</b>	<b>0.32</b>	<b>0.62</b>	<b>0.31</b>	<b>0.31</b>
EF	0.43	-----	0.43	0.64	-----	0.64
CF	0.09	-----	0.09	<b>0.05</b>	<b>-0.25</b>	<b>0.30</b>

Anmerkung: Md = Median, SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit; fett = Skalen mit Response-Shift Einfluss

### Fragestellung 3.2: Response-Shift – Ableitung nach Ausmaß der Selbstwirksamkeit

Welchen Einfluss hat die wahrgenommene *Selbstwirksamkeit* auf einen Response-Shift im Zusammenhang mit der Erfassung der gesundheitsbezogenen Lebensqualität bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Als zweite Moderatorvariable wird die Selbstwirksamkeit (SW) auf ihren potenziellen Einfluss auf einen Response-Shift überprüft. Die Modellableitungen sind in Tabelle 61 dargestellt. Auch hier werden die zwei Gruppen anhand des Medians der Skala SW gebildet und anschließend auf gruppenspezifische Parameterfreisetzungen überprüft. Wie zur Variable Benefit Finding zeigen sich auch hinsichtlich der Selbstwirksamkeit Unterschiede zwischen den Subgruppen in der Anzahl der response-shift-beeinflussten Parameter. Die Passung des Response-Shift-Modells in der Gruppe  $SW \leq Md$  weist insgesamt vier Parameterfreisetzungen auf, wobei jeweils zwei auf der Ebene der Intercepts und auf der Ebene der Messfehlervarianzen vorzunehmen sind. In der Gruppe  $SW > Md$  sind ebenfalls zwei Intercepts sowie alle Messfehlervarianzen freizusetzen. *Demnach gilt für die Variable Selbstwirksamkeit, dass in der Gruppe mit einer deutlich ausgeprägten Selbstwirksamkeit mehr Response-Shift-Parameter freizusetzen sind als in der Vergleichsgruppe.*

Tabelle 61: Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß der Selbstwirksamkeit

Gruppe	Ausprägung	Fit-Maße			
		$\chi^2/df$	RMSEA	CFI	TLI
SW $\leq$ Md	Nullmodell	2.22	.12 [.08;.15]	.79	.75
	Modell 1 – IN_RF	1.95	.11 [.07;.14]	.84	.81
	Modell 2 – IN_PF	1.71	.09 [.05;.13]	.88	.85
	Modell 3 – e_RF	1.46	.07 [.02;.11]	.93	.91
	Modell 4 – e_PF	1.20	.04 [.00;.10]	.97	.96
SW $>$ Md	Nullmodell	3.24	.17 [.14;.21]	.73	.68
	Modell 1 – IN_RF	2.61	.14 [.11;.18]	.81	.77
	Modell 2 – e_EF	2.15	.12 [.08;.16]	.87	.84
	Modell 3 – e_CF	1.94	.11 [.07;.15]	.90	.87
	Modell 4 – e_RF	1.64	.09 [.04;.13]	.93	.91
	Modell 5 – e_SF	1.38	.07 [.00;.12]	.96	.95
	Modell 6 – IN_CF	1.16	.05 [.00;.10]	.98	.98
	Modell 7 – e_PF	0.96	.00 [.00;.08]	1.00	1.00

Anmerkung: SW = Selbstwirksamkeit, e = Messfehler, IN = Faktorladung, Md = Median, RS = Response-Shift,  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Zur Veranschaulichung der Veränderungen in den response-shift-relevanten Parametern werden die freigesetzten und restringierten Parameter in Tabelle 62 vorgestellt. Auf der Ebene der Faktorladungen findet in den Subgruppen kein Response-Shift statt. Auffällig ist die geringe Faktorladung der Skala Kognitive Funktionsfähigkeit in der Gruppe SW  $\leq$  Md. Diese Variable scheint in dieser Gruppe für die GLQ nicht von wesentlicher Bedeutung zu sein. Alle anderen Faktorladungen – auch für die Vergleichsgruppe – weisen akzeptable bis gute Werte für die Faktorladungen auf.

Aus den Unterschieden in der Anzahl der freizusetzenden Response-Shift-Parameter zwischen den Subgruppen läßt sich ableiten, dass in der Gruppe SW  $>$  Md mehr Response-Shift-Prozesse stattfinden, was als weiterer Hinweis auf eine Bestätigung der formulierten Annahme zu werten ist.

Inhaltlich lassen sich die Parameterfreisetzungen in beiden Gruppen folgendermaßen interpretieren: In der Gruppe SW  $\leq$  Md gilt hinsichtlich der Rekalibrierung, dass die befragten Personen ihre Physische und Rollenfunktionsfähigkeit zum MZP 2 positiver bewerten als zum MZP 1, ohne dass diese Umbewertung auf die GLQ zurückgeführt werden kann. Die non-uni-

forme Rekalibrierung auf den Skalen Physische und Rollenfunktionsfähigkeit weist auf eine Veränderung einzelner Skalenabstufungen der jeweiligen Indikatoren hin. Teilbereiche der Skalen werden zum MZP 2 quantitativ anders gewertet als zum MZP 1. Somit ist für die genannte Subgruppe davon auszugehen, dass die Interceptfreisetzung eine Folge der Varianzveränderung der spezifischen Faktoren der Indikatorvariablen ist. Die deutliche Varianzreduktion führt in der Konsequenz auch zu einer statistisch signifikanten Interceptveränderung.

Tabelle 62: Schätzung der Response-Shift-Parameter - Selbstwirksamkeit

	Selbstwirksamkeit $\leq$ Md					Selbstwirksamkeit $>$ Md				
	Faktorladungen									
	Korrelationen mit GLQ		Kovarianzen mit GLQ		Korrelationen mit GLQ		Kovarianzen mit GLQ			
	Prä	Post	Prä	Post	Prä	Post	Prä	Post	Prä	Post
PF	.43	.56	7.53	7.53	.63	.65	12.32	12.32		
RF	.49	.66	17.14	17.14	.57	.69	19.39	19.39		
SF	.84	.84	23.03	23.03	.67	.73	20.17	20.17		
EF	.65	.65	15.36	15.36	.66	.79	16.50	16.50		
CF	.26	.25	5.15	5.15	.62	.65	13.40	13.40		
-----										
	Intercepts									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>71.52</b>	<b>38.27</b>	55.72	62.57	78.44	75.97	<b>41.12</b>	63.91	68.06	<b>84.19</b>
Posttest	<b>78.44</b>	<b>55.40</b>	55.72	62.57	78.44	75.97	<b>60.53</b>	63.91	68.06	<b>76.62</b>
-----										
	Residual-Varianzen									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>248.63</b>	<b>919.39</b>	220.44	317.17	378.86	<b>235.98</b>	<b>798.60</b>	<b>494.20</b>	<b>346.60</b>	<b>284.10</b>
Posttest	<b>120.55</b>	<b>368.74</b>	220.44	317.17	378.86	<b>127.49</b>	<b>252.48</b>	<b>217.50</b>	<b>99.99</b>	<b>153.57</b>

Anmerkungen: Md = Median, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität

In der Gruppe SW  $>$  Md gilt für die Rekalibrierung auf der Skala Rollenfunktionsfähigkeit, dass die Subgruppe diese Funktionsfähigkeit zum MZP 2 positiver bewertet als zum MZP 1, wobei diese Umbewertung nicht durch die GLQ erklärt werden kann. Auf der Skala Kognitive Funktionsfähigkeit findet eine Umbewertung der Skalierung in Richtung einer negativeren Bewertung dieser Fähigkeit zum MZP 2 statt, wobei auch diese Skalierungsänderung nicht mit dem Einfluss der GLQ erklärt werden kann. Die non-uniforme Rekalibrierung auf allen Indikatorskalen weist darauf hin, dass die befragten Personen Teile der Skalenabstufungen quantitativ anders bewerten, ohne dass diese Veränderung durch die GLQ bedingt ist. Somit ist auch für diese Subgruppe davon auszugehen, dass die formal aufgedeckte uniforme



Rekalibrierung statistisch gesehen eine Folge der non-uniformen Rekalibrierung ist. Die Varianzveränderung in den spezifischen Faktoren der Indikatorvariablen Rollenfunktionsfähigkeit und Kognitive Funktionsfähigkeit sind in ihrem Ausmaß so groß, dass dadurch die Intercepts der Indikatoren ebenfalls verändert werden.

Abschließend wird auch für die beiden Subgruppen berechnet, in wie weit es Unterschiede im Ausmaß der Response-Shift-Effekte zwischen den Gruppen  $SW \leq Md$  versus  $SW > Md$  gibt. Tabelle 63 sind Angaben über die Differenzierung der beobachteten Effekte auf den entsprechenden Skalen zu entnehmen.

Tabelle 63: Selbstwirksamkeit als Einflussfaktor auf Response-Shift

Skala	Effektstärke					
	Selbstwirksamkeit $\leq Md$			Selbstwirksamkeit $> Md$		
	beobachtet	Response-Shift	‘wahre Veränderung’	beobachtet	Response-Shift	‘wahre Veränderung’
SF	0.30	-----	0.30	0.58	-----	0.58
RF	<b>0.72</b>	<b>0.52</b>	<b>0.20</b>	<b>1.08</b>	<b>0.60</b>	<b>0.48</b>
PF	<b>0.54</b>	<b>0.37</b>	<b>0.17</b>	0.70	-----	0.70
EF	0.28	-----	0.28	0.58	-----	0.58
CF	0.01	-----	0.01	<b>0.18</b>	<b>-0.42</b>	<b>0.60</b>

Anmerkung: Md = Median, SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit; fett = Skalen mit Response-Shift Einfluss

In der Gruppe  $SW \leq Md$  zeigen sich auf allen Skalen eher kleine Effekte hinsichtlich der wahren Veränderungen durch die GLQ. Der Response-Shift-Effekt auf der Skala Rollenfunktionsfähigkeit im Ausmaß eines mittleren Effektes führt dazu, dass auf dieser Skala ein großer Effekt beobachtet wird. Auf der Skala Physische Funktionsfähigkeit ergibt sich durch den kleinen Effekt aufgrund der wahren Veränderung durch die GLQ und eines Response-Shift-Effekts, der als klein bis mittel zu beurteilen ist, insgesamt ein mittlerer beobachteter Effekt.

In der Gruppe  $SW > Md$  zeigen sich auf allen Skalen mit Ausnahme der Physischen Funktionsfähigkeit (tendenziell großer Effekt) eher mittlere Effekte auf der Ebene der wahren Veränderung. Für die Skala Rollenfunktionsfähigkeit wird ein Response-Shift-Effekt ebenfalls in mittlerer Höhe berechnet, so dass insgesamt ein großer beobachtbarer Effekt resultiert. Auffällig ist, dass auf der Skala Kognitive Funktionsfähigkeit der mittlere Effekt, der durch

die wahre Veränderung der GLQ bedingt ist, durch den Response-Shift-Effekt relativiert wird, so dass letztlich ein kleiner Effekt auf dieser Skala zu beobachten ist.

Die Response-Shift-Effekte auf der Skala Rollenfunktionsfähigkeit erfüllen in beiden Gruppen das Kriterium der klinischen Bedeutsamkeit nach Norman und Kollegen (2003). Die jeweils anderen Response-Shift-Effekte erreichen dieses Kriterium nicht.

Die Anzahl der Parameterfreisetzungen in der Gruppe  $SW > Md$ , wo sieben Parameter freizusetzen waren und insgesamt zwei Response-Shift-Prozesse mehr auftreten als in der Vergleichsgruppe  $SW \leq Md$ , spricht für einen Einfluss der Selbstwirksamkeit auf einen Response-Shift im Sinne der formulierten Annahmen. Aus den vorliegenden Effektstärke-Differenzierungen lässt sich jedoch ableiten, dass in beiden Gruppen auf der Skala Rollenfunktionsfähigkeit in etwa gleich große Response-Shift-Effekte aufgedeckt werden können, so dass das Ausmaß der wahrgenommenen Selbstwirksamkeit darauf keinen Einfluss zu haben scheint. Gegen die formulierte Annahme spricht ebenso, dass in der Gruppe  $SW \leq Md$  auf der Skala Physische Funktionsfähigkeit der Response-Shift-Effekt deutlich größer ist als der Effekt durch die wahre Veränderung.

### *Fragestellung 3.3: Response-Shift – Ableitung nach Ausmaß der positiven Affektivität*

Welchen Einfluss hat die *positive Affektivität* auf einen Response-Shift im Zusammenhang mit der Erfassung der gesundheitsbezogenen Lebensqualität bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Anhand des Medians der Verteilung der Einzelwerte auf der Skala Positive Affektivität (PA) innerhalb der Gesamtstichprobe wurden zwei Gruppen gebildet. Getrennt voneinander wurde jeweils eine Response-Shift-Ableitung vorgenommen. Es ergeben sich Unterschiede zwischen den Gruppen in der Anzahl der Parameterfreisetzungen. In der Gruppe  $PA \leq Md$  findet ein Response-Shift auf zwei Intercepts sowie auf zwei Messfehlern statt. Im Vergleich dazu werden in der Gruppe  $PA > Md$  ebenfalls zwei Intercepts, jedoch insgesamt vier Messfehlervarianzen freigesetzt. *Somit sind in der Gruppe  $PA > Md$  mehr response-Shift-Parameter freizusetzen als in der Vergleichsgruppe.* Die Angaben zur Modellpassung und inhaltliche Angaben zu den Parameterfreisetzungen sind der Tabelle 64 zu entnehmen.

Tabelle 64: Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß der positiven Affektivität

Gruppe	Ausprägung	Fit-Maße			
		$\chi^2/df$	RMSEA	CFI	TLI
PA $\leq$ Md	Nullmodell	2.45	.12 [.09;.15]	.83	.80
	Modell 1 – IC_RF	1.92	.09 [.06;.13]	.89	.87
	Modell 2 – IC_PF	1.54	.07 [.03;.11]	.94	.92
	Modell 3 – e_RF	1.40	.06 [.00;.10]	.96	.94
	Modell 4 – e_PF	1.26	.05 [.00;.09]	.97	.96
PA $>$ Md	Nullmodell	3.94	.17 [.14;.20]	.57	.49
	Modell 1 – e_RF	3.28	.15 [.12;.18]	.68	.61
	Modell 2 – e_SF	2.65	.13 [.10;.16]	.77	.72
	Modell 3 – e_PF	2.08	.10 [.07;.14]	.85	.81
	Modell 4 – IC_RF	1.70	.08 [.04;.12]	.91	.88
	Modell 5 – IC_CF	1.53	.07 [.03;.11]	.93	.91
	Modell 6 – e_EF	1.22	.05 [.00;.09]	.97	.96

Anmerkung: PA = Positive Affektivität, Md = Median, e = Messfehler, IN = Faktorladung,  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Die Werte der response-shift-relevanten Parameter in den entsprechenden gruppenspezifischen Response-Shift-Modellen werden in Tabelle 65 vorgestellt. Auch hier waren auf der Ebene der Faktorladungen beziehungsweise Kovarianzen zwischen den Indikatoren und der GLQ in keiner der Subgruppen Response-Shift-Einflüsse erkennbar. Hinsichtlich der Faktorladung der Skala Kognitive Funktionsfähigkeit resultiert in der Gruppe PA  $\leq$  Md zum MZP 1 eine grenzwertige Ladung, die zum MZP 2 nochmals etwas geringer ausfällt.

Die gruppenspezifische Response-Shift-Ableitung weist darauf hin, dass in der Gruppe PA  $>$  Md mehr Response-Shift-Prozesse stattfinden, was der formulierten Annahme entspricht.

Tabelle 65: Schätzung der Response-Shift-Parameter – positive Affektivität

	Positive Affektivität $\leq$ Md					Positive Affektivität $>$ Md				
	Faktorladungen									
	Korrelationen mit GLQ		Kovarianzen mit GLQ		Korrelationen mit GLQ		Kovarianzen mit GLQ			
	Prä	Post	Prä	Post	Prä	Post	Prä	Post	Prä	Post
PF	.39	.44	6.42	6.42	.53	.64	10.56	10.56		
RF	.60	.69	20.21	20.21	.38	.54	14.06	14.06		
SF	.79	.76	22.95	22.95	.59	.73	17.69	17.69		
EF	.70	.66	16.55	16.55	.61	.68	15.12	15.12		
CF	.37	.34	7.45	7.45	.62	.52	12.04	12.04		
-----										
	Intercepts									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>70.10</b>	<b>36.98</b>	56,82	62.21	78.04	79.17	<b>48.63</b>	66.50	69.42	<b>84.15</b>
Posttest	<b>77.18</b>	<b>58.37</b>	56,82	62.21	78.04	79.17	<b>64.94</b>	66.50	69.42	<b>76.80</b>
-----										
	Residual-Varianzen									
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>225.54</b>	<b>732.37</b>	310.21	291.10	352.66	<b>281.32</b>	<b>1140.49</b>	<b>581.10</b>	<b>379.43</b>	238.71
Posttest	<b>140.70</b>	<b>361.84</b>	310.21	291.10	352.66	<b>101.32</b>	<b>293.77</b>	<b>172.62</b>	<b>167.70</b>	238.71

Anmerkungen: Md = Median, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität

Die in Tabelle 65 dargestellten Response-Shift-Prozesse lassen sich inhaltlich folgendermaßen beschreiben: In der Gruppe  $PA \leq Md$  gilt bezüglich der Rekalibrierung, dass die befragten Personen ihre Physische und Rollenfunktionsfähigkeit zum MZP 2 positiver bewerten als zum MZP 1. Diese geänderte Bewertung kann nicht durch den Einfluss der GLQ erklärt werden. Die non-uniforme Rekalibrierung auf diesen beiden Skalen drückt eine Umbewertung einzelner Skalenabstufungen der beiden Indikatoren aus. Teilbereiche der Skalen werden zum MZP 2 quantitativ anders gewertet als zum MZP 1. Da die Interceptfreisetzung auf den Skalen stattfindet, auf denen auch eine non-uniforme Rekalibrierung auftritt, ist der Schluss nahe liegend, dass dies eine Konsequenz der Varianzveränderung der spezifischen Faktoren der jeweiligen Indikatoren ist. Wesentlich ist in diesem Sinne also die non-uniforme Rekalibrierung als Response-Shift-Prozess.

In der Gruppe  $PA > Md$  bedeutet die Rekalibrierung auf der Skala Rollenfunktionsfähigkeit eine positivere Bewertung dieser Funktionsfähigkeit zum MZP 2 im Vergleich zum MZP 1, wobei diese Skalierungsänderung nicht durch die GLQ als latente Variable begründet werden kann. Die non-uniforme Rekalibrierung auf den Skalen Physische, Soziale, Emotionale

und Rollenfunktionsfähigkeit weist darauf hin, dass die befragten Personen Teile der Skalenabstufungen zum MZP 2 quantitativ anders skalieren als zum MZP 1, ohne dass diese Veränderung durch die GLQ bedingt ist. Für die Interceptfreisetzung der Skala Rollenfunktionsfähigkeit ist davon auszugehen, dass diese eine Folge der statistisch bedeutsamen Veränderung der Varianz des spezifischen Faktors dieses Indikators ist, also eine Folge der non-uniformen Rekalibrierung. Dafür ist auf der Skala Kognitive Funktionsfähigkeit von einer einheitlichen Veränderung der internen Bewertungsstufen des Indikators auszugehen. Die gleiche kognitive Funktionsfähigkeit wird zum MZP 2 negativer bewertet als zum MZP 1. Dieser Trend zu einer negativeren Bewertung gleicher Zustände findet somit auf allen Skalenabstufungen gleichermaßen statt.

Für die beiden Subgruppen werden Effektstärke-Kalkulationen vorgenommen, um Hinweise auf einen Einfluss der positiven Affektivität auf die Response-Shift-Effekte zu erhalten. Der Tabelle 66 sind Angaben über die Differenzierung der beobachteten Effekte auf den entsprechenden Skalen zu entnehmen.

Tabelle 66: Positive Affektivität als Einflussfaktor auf Response-Shift

Skala	Effektstärke					
	Positive Affektivität $\leq$ Md			Positive Affektivität $>$ Md		
	beobachtet	Response-Shift	‘wahre Veränderung’	beobachtet	Response-Shift	‘wahre Veränderung’
SF	0.22	-----	0.22	0.57	-----	0.57
RF	<b>0.85</b>	<b>0.68</b>	<b>0.17</b>	<b>0.75</b>	<b>0.42</b>	<b>0.33</b>
PF	<b>0.56</b>	<b>0.45</b>	<b>0.11</b>	0.64	-----	0.64
EF	0.26	-----	0.26	0.48	-----	0.48
CF	0.07	-----	0.07	<b>0.16</b>	<b>-0.37</b>	<b>0.53</b>

Anmerkung: MD = Median, SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit; fett = Skalen mit Response-Shift Einfluss

In der Gruppe  $PA \leq Md$  zeigen sich insgesamt sehr heterogene beobachtete Effekte. Auf den Skalen Soziale und Emotionale Funktionsfähigkeit, die nicht durch einen Response-Shift beeinflusst werden, zeigen sich kleine Effekte. Der Effekt auf der Skala Kognitive Funktionsfähigkeit ist zu vernachlässigen. Auf der Skala Rollenfunktionsfähigkeit ist ein großer Effekt zu beobachten, der sich aus einem mittleren Response-Shift-Effekt und einem kleinen Effekt bedingt durch die Veränderung des wahren Wertes zusammensetzt. Auf der Skala Physische Funktionsfähigkeit resultiert ein mittlerer beobachteter Effekt aus der Summe eines mittleren

Response-Shift-Effektes und eines zu vernachlässigenden Effekts durch die wahre Veränderung der GLQ.

In der Gruppe  $PA > Md$  zeigen sich auf allen Skalen, die nicht durch einen Response-Shift beeinflusst sind, tendenziell eher mittlere beobachtete Effekte. Die Skala Rollenfunktionsfähigkeit weist einen großen beobachteten Effekt auf, der sich aus einem tendenziell mittleren Response-Shift-Effekt und einem eher kleinen Effekt durch die Veränderung der GLQ zusammensetzt. Auf der Skala Kognitive Funktionsfähigkeit zeigt sich wiederum das Bild, dass ein Response-Shift-Effekt im Ausmaß eines kleinen bis mittleren Effektes den mittleren Effekt der wahren Veränderung teilweise überdeckt, so dass der beobachtete Effekt als eher klein zu beurteilen ist. Lediglich der Response-Shift-Effekt auf der Skala Rollenfunktionsfähigkeit in der Gruppe  $PA \leq Md$  erfüllt das Kriterium der klinischen Bedeutsamkeit nach Norman und Kollegen (2003).

Die Anzahl der freizusetzenden Response-Shift-Parameter sowie die damit einhergehende Anzahl der Response-Shift-Prozesse sprechen für die Annahme, dass ein höheres Ausmaß an positiver Affektivität einen Response-Shift fördert. Gegen die Annahme spricht jedoch, dass in der Gruppe  $PA \leq Md$  auf der Skala Rollenfunktionsfähigkeit der Response-Shift-Effekt größer ist als in der Gruppe  $PA > Md$ . Auch der mittlere Response-Shift-Effekt auf der Skala Physische Funktionsfähigkeit in der Gruppe  $PA \leq Md$  kann als Hinweis auf die Ablehnung der Annahme gewertet werden.

#### *Fragestellung 3.4: Response-Shift – Ableitung nach Ausmaß der negativen Affektivität*

Welchen Einfluss hat die *negative Affektivität* auf einen Response-Shift im Zusammenhang mit der Erfassung der gesundheitsbezogenen Lebensqualität bei Patienten mit PCa im Rahmen der onkologischen Rehabilitation?

Wie auch bei den vorangegangenen potenziellen Moderatorvariablen wurde eine Subgruppenbildung der Gesamtstichprobe anhand des Medians der Variable negative Affektivität (NA) vorgenommen. Für beide Gruppen wurden getrennt Response-Shift-Ableitungen vorgenommen. Unterschiede zwischen den Gruppen ergaben sich in der Anzahl der Parameterfreisetzungen. In der Gruppe  $NA \leq Md$  fand ein Response-Shift auf einem Intercept sowie auf allen Messfehlern statt. In der Gruppe  $NA > Md$  wurden drei Intercepts und vier Messfehler-

varianzen freigesetzt. In der Gruppe  $NA > Md$  waren somit mehr Response-Shift-Parameter freizusetzen als in der Vergleichsgruppe. Die Ergebnisse sind in Tabelle 67 für beide Subgruppen dargestellt.

Tabelle 67: Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß der negativen Affektivität

Gruppe	Ausprägung	Fit-Maße			
		$\chi^2/df$	RMSEA	CFI	TLI
NA $\leq$ Md	Nullmodell	3.32	.15 [.12;.18]	.70	.64
	Modell 1 – e_EF	2.66	.12 [.10;.15]	.79	.75
	Modell 2 – e_RF	2.21	.11 [.07;.14]	.85	.82
	Modell 3 – IN_RF	1.90	.09 [.06;.12]	.89	.86
	Modell 4 – e_PF	1.58	.07 [.03;.11]	.93	.91
	Modell 5 – e_CF	1.41	.06 [.00;.10]	.95	.94
	Modell 6 – e_SF	1.31	.05 [.00;.09]	.97	.95
NA $>$ Md	Nullmodell	2.60	.13 [.10;.16]	.76	.72
	Modell 1 – IN_CF	2.40	.12 [.09;.15]	.80	.75
	Modell 2 – IN_RF	2.04	.10 [.07;.14]	.85	.82
	Modell 3 – IN_PF	1.74	.09 [.05;.12]	.90	.87
	Modell 4 – e_RF	1.51	.07 [.02;.11]	.93	.91
	Modell 5 – e_PF	1.23	.05 [.00;.09]	.97	.96
	Modell 6 – e_EF	1.10	.02 [.00;.08]	.99	.99
	Modell 7 – e_SF	0.90	.00 [.00;.07]	1.00	1.02

Anmerkung: NA = Negative Affektivität, Md = Median, e = Messfehler, IN = Faktorladung,  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation [CI 95%], CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

In Tabelle 68 werden die Werte der response-shift-relevanten Parameter in den entsprechenden gruppenspezifischen Response-Shift-Modellen vorgestellt. Wie aus der Tabelle ersichtlich ist, war auch bei diesen Subgruppen kein Response-Shift auf der Ebene der Faktorladungen zu verzeichnen. Bezüglich der Skala Kognitive Funktionsfähigkeit resultiert in der Gruppe  $NA \leq Md$  zu beiden MZPen eine grenzwertig akzeptable Faktorladung. Für die Interceptfreisetzungen der Skalen Physische und Rollenfunktionsfähigkeit in der Gruppe  $NA > Md$  ist nach Berücksichtigung der Freisetzung der Messfehlervarianz davon auszugehen, dass es sich bei diesen Prozessen um eine Folge der non-uniformen Rekalibrierung auf den entsprechenden Skalen handelt, also der Varianzveränderung der spezifischen Faktoren eine wesentliche Rolle an der Interceptfreisetzung zukommt. Anders ist dies bei der Interceptfreisetzung der Skala Kognitive Funktionsfähigkeit in dieser Gruppe. Hier ist von einer uniformen Reka-

librierung auszugehen, da die Varianz des spezifischen Faktors über beide MZPe gleich bleibt. Somit ist für die Skala Kognitive Funktionsfähigkeit von einer einheitlichen Veränderung der Bewertung der Skalierung des Indikators auszugehen.

Tabelle 68: Schätzung der Response-Shift-Parameter – Negative Affektivität

	Negative Affektivität $\leq$ Md				Negative Affektivität $>$ Md					
	Korrelationen mit GLQ		Kovarianzen mit GLQ		Korrelationen mit GLQ		Kovarianzen mit GLQ			
	Prä	Post	Prä	Post	Prä	Post	Prä	Post	Post	
PF	.57	.62	10.76	10.76	.43	.54	7.46	7.46		
RF	.48	.61	17.19	17.19	.55	.71	18.46	18.46		
SF	.63	.65	18.53	18.53	.69	.80	20.11	20.11		
EF	.60	.73	13.29	13.29	.60	.69	14.96	14.96		
CF	.35	.36	6.32	6.32	.43	.41	9.04	9.04		
-----										
Intercepts										
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	76.84	<b>45.68</b>	65.16	74.87	85.17	<b>71.31</b>	<b>38.01</b>	56.09	56.95	<b>77.10</b>
Posttest	76.84	<b>76.84</b>	65.16	74.87	85.17	<b>78.37</b>	<b>58.00</b>	56.09	56.95	<b>71.39</b>
-----										
Residual-Varianzen										
	PF	RF	SF	EF	CF	PF	RF	SF	EF	CF
Prätest	<b>241.16</b>	<b>1000.11</b>	<b>514.69</b>	<b>320.45</b>	<b>292.43</b>	<b>240.92</b>	<b>786.02</b>	<b>439.09</b>	<b>402.53</b>	368.55
Posttest	<b>122.07</b>	<b>318.25</b>	<b>307.53</b>	<b>102.67</b>	<b>174.82</b>	<b>123.42</b>	<b>316.72</b>	<b>214.90</b>	<b>226.35</b>	368.55

Anmerkungen: Md = Median, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, SF = Soziale Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, GLQ = Gesundheitsbezogene Lebensqualität

Die Response-Shift-Prozesse in den beiden dargestellten Subgruppen lassen sich inhaltlich derart interpretieren, dass in der Gruppe  $NA \leq Md$  bezüglich der Rekalibrierung gilt, dass die befragten Personen ihre Rollenfunktionsfähigkeit zum MZP 2 positiver bewerten als zum MZP 1, wobei diese Umbewertung nicht durch die GLQ als latentes Konstrukt erklärt werden kann. Die non-uniforme Rekalibrierung auf allen Skalen verweist auf eine geänderte Bewertung einzelner Abstufungen der Indikatorskaliierungen hin. Teilbereiche der Skalen werden zum MZP 2 quantitativ anders gewertet als zum MZP 1. Auch für diese Subgruppe ist anzunehmen, dass die Interceptfreisetzung auf der Skala Rollenfunktionsfähigkeit eine Konsequenz der deutlichen Varianzreduktion des spezifischen Faktors des Indikators zum MZP 2 ist.



In der Gruppe  $NA > Md$  bedeutet die uniforme Rekalibrierung auf den Skalen Physische und Rollenfunktionsfähigkeit eine positivere Bewertung dieser Funktionsbereiche zum MZP 2 im Vergleich zum MZP 1, wobei diese Änderung der Skalierung nicht durch die GLQ begründet werden kann. Die non-uniforme Rekalibrierung auf den Skalen Physische, Soziale, Emotionale und Rollenfunktionsfähigkeit weist darauf hin, dass die befragten Personen Teile der Skalenabstufungen zum MZP 2 quantitativ anders skalieren als zum MZP 1, ohne dass diese Veränderung durch die GLQ bedingt ist. Da die Messfehlervarianz der Skala Kognitive Funktionsfähigkeit restringiert bleiben konnte, handelt es sich bei der Interceptfreisetzung auf dieser Skala um eine uniforme Rekalibrierung. Somit weist die gruppenspezifische Response-Shift-Ableitung darauf hin, dass in beiden Subgruppen gleich viele Response-Shift-Prozesse stattgefunden haben.

Um Hinweise auf einen Einfluss der NA auf das Ausmaß der Rekalibrierungen zu erhalten, werden für beide Subgruppen Effektstärke-Differenzierungen vorgenommen. Der Tabelle 69 sind Angaben über die Kalkulation der beobachteten Effekte auf den entsprechenden Skalen zu entnehmen.

Tabelle 69: Negative Affektivität als Einflussfaktor auf Response-Shift

Skala	Effektstärke					
	Negative Affektivität $\leq$ Md			Negative Affektivität $>$ Md		
	beobachtet	Response-Shift	‘wahre Veränderung’	beobachtet	Response-Shift	‘wahre Veränderung’
SF	0.45	-----	0.45	0.31	-----	0.31
RF	<b>0.83</b>	<b>0.43</b>	<b>0.40</b>	<b>0.81</b>	<b>0.60</b>	<b>0.21</b>
PF	0.62	-----	0.62	<b>0.58</b>	<b>0.42</b>	<b>0.16</b>
EF	0.59	-----	0.59	0.16	-----	0.16
CF	0.23	-----	0.23	<b>0.16</b>	<b>-0.37</b>	<b>0.53</b>

Anmerkung: Md = Median, SF = Soziale Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, PF = Physische Funktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit; fett = Skalen mit Response-Shift Einfluss

In der Gruppe  $NA \leq Md$  ergeben die Berechnungen der beobachteten Effekte auf den Skalen Soziale, Physische und Emotionale Funktionsfähigkeit eher mittlere Effekte. Auf der response-shift-beeinflussten Skala Rollenfunktionsfähigkeit ist ein großer Effekt beobachtbar, der in etwa zu gleichen Anteilen auf einen Response-Shift und auf die wahre Veränderung durch die GLQ zurückzuführen ist. Der beobachtete Effekt auf der Skala Kognitive Funktionsfähigkeit ist als klein zu bewerten.

In der Gruppe NA > Md ergeben die Berechnungen für die Skalen Soziale, Emotionale und Kognitive Funktionsfähigkeit eher kleine Effekte. Auf der zuletzt genannten Skala findet jedoch ein Response-Shift statt, der das Ausmaß eines tendenziell mittleren Effektes aufweist und einen mittleren Effekt in der Kognitiven Funktionsfähigkeit, bedingt durch die Veränderungen in der GLQ, teilweise überdeckt. Auf der Skala Physische Funktionsfähigkeit resultiert für diese Gruppe ein beobachteter Effekt im Ausmaß eines mittleren Effektes, der zum größeren Teil auf einen Response-Shift zurückzuführen ist. Auf der Skala Rollenfunktionsfähigkeit zeigt sich ein ähnliches Bild, wobei der beobachtete Effekt vom Ausmaß her als groß zu bezeichnen ist. Als klinisch bedeutsam ist lediglich der Response-Shift-Effekt auf der Skala Rollenfunktionsfähigkeit in der Gruppe NA > Md einzustufen (nach Norman et al. 2003).

Insgesamt sprechen die Ergebnisse nicht dafür, dass ein geringes Ausmaß an NA einen Response-Shift bei PCa-Patienten in der onkologischen Rehabilitation im Rahmen einer AHB eher fördert als ein hohes Ausmaß dieser Persönlichkeitseigenschaft. Die verglichenen Subgruppen unterscheiden sich nicht bezüglich der Anzahl der Response-Shift-Prozesse. Außerdem ist der Response-Shift-Effekt auf der Skala Rollenfunktionsfähigkeit in der Gruppe NA > Md um einen kleinen Effekt größer als in der Vergleichsgruppe, was ebenfalls gegen die vorher formulierte Annahme spricht.

#### 7.3.4. Zusammenfassung der Ergebnisse der Fragestellung 3

Im Rahmen der Fragestellung 3 der vorliegenden Arbeit wurde überprüft, ob die aus dem Response-Shift-Modell nach Sprangers und Schwartz (1999) abgeleiteten Variablen Benefit Finding, Selbstwirksamkeit sowie positive und negative Affektivität einen moderierenden Einfluss auf das Auftreten und das Ausmaß eines Response-Shift haben.

Zur Überprüfung des Einflusses der vier Variablen wurde die Gesamtstichprobe anhand jeder der vier Variablen in jeweils zwei Subgruppen eingeteilt. Diese Einteilung erfolgte anhand des Medians der entsprechenden Variablen. Aufgrund fehlender Werte setzten sich die Subgruppen nur aus einer kleinen Anzahl PCa-Patienten zusammen, so dass die Aussagekraft der vorliegenden Ergebnisse deskriptiv zu werten sind.

Für die Variable Benefit Finding wurde angenommen, dass ein höheres Ausmaß an Benefit Finding einen Response-Shift fördert. Die Auswertungen anhand der Modellableitung

und Parameterfreisetzungen sprechen für diese Annahme. In der Gruppe mit einer höheren Ausprägung an Benefit Finding waren mehr Response-Shift-Parameter freizusetzen als in der Vergleichsgruppe. Gegen die formulierte Annahme spricht jedoch das Ergebnis, dass Benefit Finding auf die Größe des Response-Shift-Effektes keinen Einfluss zu haben scheint.

Hinsichtlich der wahrgenommenen Selbstwirksamkeit wurde angenommen, dass ein hohes Ausmaß dieser Variable response-shift-förderlich ist. Die Invarianztestung zwischen den Subgruppen resultierte in dem Ergebnis, dass in der Subgruppe der Personen mit einer stärkeren Ausprägung an Selbstwirksamkeit mehr Response-Shift-Parameter freizusetzen waren und mehr Response-Shift-Prozesse auftraten. Jedoch zeigte sich auch hier, dass in den vorliegenden Subgruppen die Größe der Response-Shift-Effekte nicht durch das Ausmaß an Selbstwirksamkeit beeinflusst wird.

Die Auswertungen zur positiven Affektivität zeigten für die Subgruppe der Personen mit einer hohen Ausprägung auf der entsprechenden Variablen zwei Parameterfreisetzungen und drei Response-Shift-Prozesse mehr als in der Vergleichsgruppe. Dieses Ergebnis ist als Hinweis für die Annahme zu sehen, dass ein hohes Ausmaß an positiver Affektivität Response-Shift-Prozesse fördert. Die Höhe der berechneten Response-Shift-Effekte auf den einzelnen Skalen zeigte jedoch keine Unterschiede zwischen den Gruppen, so dass dieses Ergebnis die formulierte Annahme nicht bestätigt.

Die Response-Shift-Ableitung in den Gruppen, die anhand des Medians der Variable negative Affektivität gebildet wurden, zeigte das Ergebnis, dass in beiden Subgruppen gleich viele Response-Shift-Prozesse stattfanden. Die Auswertungen zu den Response-Shift-Effekten weisen darauf hin, dass ein höheres Ausmaß an negativer Affektivität einen Response-Shift zu fördern scheint. Auf der Skala Rollenfunktionsfähigkeit zeigt sich in der Gruppe mit einem höheren Ausmaß an negativer Affektivität ein im Ausmaß eines kleinen Effekts größerer Response-Shift-Effekt als in der Vergleichsgruppe.

## 8. Diskussion

Gesundheitsbezogene Lebensqualität (GLQ) ist neben organisch-medizinischen Parametern ein wesentliches Outcome-Kriterium zur Bewertung von Behandlungsmaßnahmen für Menschen mit Krebserkrankungen. Dies gilt sowohl für Betroffene mit Krebserkrankungen mit günstiger Prognose beziehungsweise einem frühen Krankheitsstadium als auch für diejenigen, deren Erkrankung fortgeschritten ist beziehungsweise eine ungünstige Prognose besitzt (Penson et al. 2003). Es existieren zahlreiche Instrumente zur Erfassung der GLQ bei Menschen mit Krebserkrankungen. Diese lassen sich unterteilen in generische, krebs- und krankheitsspezifische Verfahren (Ravens-Sieberer & Cieza 2000). Generische Instrumente dienen in erster Linie dem Vergleich der GLQ bei unterschiedlichen Grunderkrankungen oder von (krebs-)kranken und gesunden Menschen. Krebspezifische Verfahren wurden entwickelt, um Lebensqualität bei Personen mit Krebserkrankungen allgemein zu erfassen sowie diese bei Menschen mit verschiedenen Krebserkrankungen untersuchen und vergleichen zu können. Krankheitsspezifische Instrumente dienen der Erfassung der Lebensqualität bei Personen mit einer bestimmten Krebserkrankung.

GLQ ist nicht direkt erfassbar. Sie wird durch verschiedene Indikatoren operationalisiert. Aus den Beurteilungen dieser Indikatoren, die in der Regel durch den Betroffenen selbst, also subjektiv, erfolgen, wird auf das Ausmaß der GLQ geschlossen. Zur Erfassung der Veränderungen der GLQ bei Krebspatienten werden normalerweise Messungen zu verschiedenen MZPen vorgenommen. Aus der Differenz der Messwerte zu zwei MZPen werden Veränderungen der GLQ abgeleitet. Zur Bewertung von Interventionen bei Krebspatienten werden insbesondere randomisierte kontrollierte Studien (RCTs) gefordert. Diese sind jedoch nicht immer anwendbar. Zwar gelten RCTs als zu erreichender Standard in der Interventionsforschung. Jedoch sind in empirischen Studien – insbesondere im Rahmen der Wirksamkeitsstudien in der onkologischen Rehabilitation – RCTs aus ethischen Gründen nicht umsetzbar. Häufiger wird die klassische Prä-Post-Messung verwendet, bei der vor und nach einer Maßnahme die GLQ erhoben und daraus ein Differenzwert berechnet wird, der an spezifischen stichprobenbezogenen Verteilungsmaßen relativiert wird, um einen Effekt der Maßnahme zu kalkulieren. Kritisch ist diese Strategie der Veränderungsmessung, wenn sich die Bewertungsgrundlagen der befragten Personen zur Einschätzung der Fragen zur GLQ zu beiden MZPen unterscheiden, so dass möglicherweise nicht mehr von einem vergleichbaren Hinter-

grund zur Beantwortung der Items zu den beiden MZPen ausgegangen werden kann. Im Extremfall würde aus zwei qualitativ unterschiedlichen Zuständen eine rein quantitative Differenz berechnet, die jedoch nicht valide ist (Schwartz & Sprangers 2002).

Die Veränderung des Bedeutungsgehalts eines internen Konstrukts, wie beispielsweise der GLQ, im Rahmen der Auseinandersetzung mit einem kritischen Lebensereignis, zum Beispiel einer Krebserkrankung, wird nach Sprangers und Schwartz (1999) als Response-Shift bezeichnet. Ursächlich für diesen Response-Shift können drei verschiedene Prozesse sein, die als Rekalibrierung, Repriorisierung und Neukonzeptualisierung bezeichnet werden. Diese drei Prozesse stellen unterschiedliche Formen des Response-Shifts dar: Eine Rekalibrierung bedeutet eine Veränderung der internen-subjektiven Metrik zur Beurteilung eines Konstrukts beziehungsweise dessen Komponenten; die Repriorisierung bezeichnet eine Änderungen der Bedeutsamkeit einzelner Komponenten eines Konstrukts, ohne dass sich das Konstrukt selbst qualitativ verändert; die Neukonzeptualisierung umfasst die Neudefinition eines Konstrukts. Response-Shift tritt vor allem dann auf, wenn sich Menschen mit einer die eigene Identität bedrohenden kritischen Lebenssituation auseinandersetzen (Sprangers 2002). Die Mitteilung der Diagnose Krebs beziehungsweise die Behandlung einer schwerwiegenden, chronischen oder auch lebensbedrohlichen Erkrankung kann solch ein kritisches Ereignis sein.

Die vorliegende Arbeit beschäftigt sich mit Response-Shift in mehrfacher Weise. Zum Einen wurde untersucht, ob in einer Stichprobe von 212 PCa-Patienten in einem Zeitraum von Beginn der Anschlussheilbehandlung in einer onkologischen Rehabilitationsklinik bis drei Monate nach Beginn der Maßnahme ein Response-Shift nachweisbar ist (Fragestellung 1). Zum Anderen sollte bei Auftreten eines Response-Shifts in der untersuchten Stichprobe weiter geklärt werden, ob aus einer Veränderung der Globalen Lebensqualität Aussagen über einen Response-Shift abgeleitet werden können (Fragestellung 2). Abschließend sollten verschiedene potenzielle Moderatorvariablen bezüglich ihres Einflusses auf einen Response-Shift überprüft werden (Fragestellung 3). Die Fragestellung 1 ist die Hauptfragestellung der vorliegenden Arbeit. Die Fragestellungen 2 und 3 werden als Nebenfragestellungen angesehen.

In der Literatur werden verschiedene methodische Strategien zur Response-Shift-Erfassung vorgeschlagen. In den meisten Studien werden entweder mehrere Erfassungsstrategien für die drei von Sprangers und Schwartz genannten Response-Shift-Prozesse verwendet oder sie untersuchen lediglich die Rekalibrierung mit dem Then-Test. Dabei ist es jedoch nötig,

zum zweiten MZP eine retrospektive Einschätzung der GLQ für den Zeitpunkt der ersten Messung durchzuführen. Daher hat die genannte Strategie den Nachteil, dass sie mit einem hohen Aufwand betrieben werden muss. Die Verwendung des Thentests, der in der Literatur am häufigsten zur Response-Shift-Erfassung eingesetzt wird, ist jedoch nicht unumstritten (Schwartz et al. 2006). Der Einsatz der Konfirmatorischen Faktorenanalyse (KFA) als statistisches Verfahren ermöglicht eine Überprüfung der Response-Shift-Prozesse nach Sprangers und Schwartz (1999) ohne zusätzliche Messungen (Oort 2004). Dabei wird anhand von Parameterveränderungen im Längsschnitt überprüft, ob Response-Shift stattgefunden hat oder nicht (Oort 2005a). Somit kombiniert der Ansatz der Response-Shift-Erfassung mittels KFA eine hohe Ökonomie mit einer Differenzierung verschiedener Response-Shift-Prozesse. In der vorliegenden Arbeit wurde zur Response-Shift-Erfassung die Konfirmatorische Faktorenanalyse eingesetzt.

Im Folgenden werden die wesentlichen Ergebnisse der vorliegenden Arbeit zusammengefasst und diskutiert. Dabei wird die Diskussion in verschiedene Abschnitte gegliedert. Es werden zuerst die Ergebnisse der Fragestellung 1 (Hauptfragestellung) zusammengefasst und kritisch bewertet. Anschließend werden methodische Aspekte zum Ansatz nach Oort (2004) im Rahmen der Response-Shift-Erfassung diskutiert. Danach wird auf die Bedeutung des Response-Shifts für die Veränderungsmessung und Effektkalkulation sowie die inhaltliche Interpretation des Response-Shifts in der vorliegenden Gesamtstichprobe eingegangen. Ein weiteres Thema ist die Bedeutung von Response-Shift für die Messung von GLQ als Outcome-Kriterium in der onkologischen Rehabilitation. Daran anschließend werden in einem weiteren Abschnitt die Ergebnisse der Fragestellungen 2 und 3 (Nebenfragestellungen) diskutiert, wobei Bezug genommen wird auf die Modellbildung von GLQ und Response-Shift nach Sprangers und Schwartz (1999). Abschließend werden allgemeine methodische Einschränkungen der Arbeit erörtert, um den Interpretationsrahmen der vorgestellten Ergebnisse zu begrenzen.

## **8.1 Response-Shift bei PCa-Patienten in der onkologischen Rehabilitation – Fragestellung 1**

Im folgenden Kapitel werden die Ergebnisse der Fragestellung 1 (Hauptfragestellung) vorgestellt und diskutiert. Die Fragestellung bildet den Kern der Arbeit und ist der Ausgangspunkt für die Diskussion der wesentlichen inhaltlichen und methodischen Aspekte und Aussagen, die sich aus der vorliegenden Arbeit ableiten lassen. Es wird mit einer kurzen Zusam-

menfassung der Ergebnisse der Fragestellung begonnen. Diese werden anschließend inhaltlich interpretiert. Aussagen über die Bedeutung des Response-Shifts für die Veränderungsmessung und Änderungssensitivität sowie die Effektkalkulation werden kritisch betrachtet. Die Möglichkeiten und Grenzen des Ansatzes der Response-Shift-Erfassung mit Hilfe der Konfirmatorischen Faktorenanalyse sind weitere Diskussionsthemen.

Die Hauptfragestellung der vorliegenden Arbeit lautet: Hat Response-Shift einen Einfluss auf die Erfassung der gesundheitsbezogenen Lebensqualität in der onkologischen Rehabilitation von PCa-Patienten, und falls ja, in welcher Form?

Zur Aufdeckung eines Response-Shifts nach Sprangers und Schwartz (1999) wurde die methodische Strategie der Invarianztestung spezifischer Modell-Parameter durch deren Restriktion zu zwei MZPen nach Oort (2004, 2005a, 2005b) angewendet. Dem Ansatz zufolge lässt sich Response-Shift operationalisieren, indem die Intercepts beziehungsweise Messfehlervarianzen der einzelnen GLQ-Indikatoren zu beiden MZPen gleich gesetzt werden. Auf diese Weise lässt sich eine uniforme beziehungsweise non-uniforme Rekalibrierung untersuchen. Weiterhin wurden die Faktorladungen eines Indikators der GLQ zu beiden MZPen gleich gesetzt, wodurch eine Repriorisierung geprüft werden soll. Zusätzlich lässt sich über die Invarianz des gesamten Messmodells der GLQ zu beiden MZPen eine Aussage über eine Neukonzeptualisierung treffen. Somit lässt sich neben der ersten Teilfrage, ob ein Response-Shift in der vorliegenden Stichprobe auftritt oder nicht, aufgrund der Operationalisierungen der drei verschiedenen Response-Shift-Prozesse auch die zweite Teilfrage nach der Form des Response-Shifts beantworten.

#### 8.1.1 Kurzbeschreibung der Ergebnisse der Fragestellung 1

Das Ziel der vorliegenden Arbeit war es, bei einer Stichprobe von 212 PCa-Patienten im Rahmen der AHB zu untersuchen, ob ein Response-Shift im Zusammenhang mit der Erfassung gesundheitsbezogener Lebensqualität bei Prä-Post-Messungen auftritt, und wenn ja, in welcher Form sich dieser Response-Shift darstellt. Weiterhin wurde untersucht, wie ein möglicher Response-Shift die herkömmlich berechneten Effekte von Prä-Post-Messungen beeinflussen kann. In diesem Unterkapitel der Diskussion werden das Vorgehen und die Ergebnisse zur Beantwortung der Hauptfragestellung zusammengefasst und kritisch diskutiert.

## *Beschreibung des Vorgehens*

Um einen Response-Shift nachzuweisen, wurde ein Modell der GLQ, gemessen mit dem EORTC-QLQ-C30 (Fayers et al. 1995), aus der Literatur sowie anhand einer ersten Modellvalidierung in einer Substichprobe abgeleitet. Zur Modellableitung und Überprüfung sowie der Prüfung auf verschiedene Response-Shift-Prozesse wurde das Statistikprogramm AMOS 17 (Arbuckle 2005) verwendet. Ausgehend von der inhaltlichen Definition von Response-Shift als Phänomen, das sich in drei unterschiedlichen Formen, einer Rekalibrierung, Repriorisierung und Neukonzeptualisierung, zeigen kann, wurde eine Testung geschachtelter Modelle vorgenommen, um diese Prozesse auf Parameterebene aufzudecken. Das Vorgehen zur Response-Shift-Erfassung mittels KFA orientiert sich an dem Vorgehen nach Oort (2005a).

Ein Response-Shift nach Sprangers und Schwartz (1999) lässt sich mittels KFA in folgender Art und Weise operationalisieren (Oort 2005a): Eine non-uniforme Rekalibrierung zeigt sich durch statistisch signifikant unterschiedliche Messfehlervarianzen eines Indikators zu zwei MZPen. Eine uniforme-Rekalibrierung äußert sich in Form einer statistisch signifikanten Veränderung des Intercepts eines Indikators zwischen zwei MZPen. Die Veränderung der Intercepts kann jedoch auch verursacht werden durch die Unterschiede in der Messfehlervarianz eines Indikators zu beiden MZPen. Somit kann sich die Messfehlerinvarianz in Form einer Intercept- *und* einer Messfehlerinvarianz ausdrücken. Eine Repriorisierung ist nach dem Vorgehen mittels KFA gekennzeichnet durch eine statistisch signifikante Veränderung der Faktorladungen eines Indikators zwischen zwei MZPen, ohne dass die Faktorladung zu einem der beiden MZPe einen Wert nahe Null annimmt. Die Neukonzeptualisierung zeichnet sich aus durch eine statistisch signifikante Veränderung des gesamten Faktorladungsmusters eines Messmodells zwischen zwei MZPen. Der Freisetzungsprozess der Parameter wird begonnen von einem vollständig restringierten Modell. Das bedeutet, dass einerseits bestimmte Parameter definiert werden müssen, um das Modell identifizierbar zu machen, und andererseits alle Response-Shift-Parameter zu beiden MZPen gleichgesetzt werden. Mit Hilfe der Modifikationsindizes wurde überprüft, ob die Gleichsetzung der Response-Shift-Parameter für beide MZPe beizubehalten oder aufzuheben war. Im Sinne geschachtelter Modelltestungen wurden die Parameterfreisetzungen in der Reihenfolge überprüft, die aufgrund der Modifikationsindizes vorzunehmen waren. Es wurde demnach jeweils der Response-Shift-Parameter freigesetzt, der die Diskrepanzfunktion mittels  $\chi^2/df$ -Test am deutlichsten und statistisch bedeutsam minimiert.



Inhaltlich bedeuten die Response-Shift-Prozesse Folgendes: Bei der Rekalibrierung verändert sich die Metrik beziehungsweise Bedeutung der vorgegebenen Item- beziehungsweise Skalenabstufungen. Beispielsweise verwenden die befragten Teilnehmer für objektiv gleichbleibende Zustände zu zwei MZPen unterschiedliche quantitative Bewertungen. Ein und derselbe Zustand (bspw. der körperlichen Belastbarkeit) wird zu verschiedenen Zeitpunkten als unterschiedlich belastend bewertet. Bei der uniformen Rekalibrierung ist davon auszugehen, dass sich alle Item- beziehungsweise Skalenstufen eines Instrumentes homogen in die gleiche Richtung verändern, sich also die Bewertungsstufen parallel zueinander verschieben. Eine non-uniforme Rekalibrierung führt zu einer *uneinheitlichen* Veränderung der Item- beziehungsweise Skalenabstufungen (einige Stufen ändern sich, andere nicht). Eine Repriorisierung zeigt an, dass ein für die Bewertung der GLQ relevanter Indikator an Bedeutung für die GLQ zu- oder abnimmt. Die Neukonzeptualisierung bedeutet, dass die GLQ zu beiden MZPen durch unterschiedliche Indikatorenmuster definiert wird. Die GLQ verändert sich demnach *qualitativ*.

#### *Ergebnisse: Response-Shift bei Patienten mit PCa in der AHB*

In der vorliegenden Stichprobe konnte ein Response-Shift an einer Stichprobe aus 212 PCa-Patienten in der AHB nachgewiesen werden. Das Basismodell der GLQ, das anhand der Indikatorvariablen des EORTC-QLQ-C30 abgeleitet und über zwei MZPe definiert wurde, zeigte Werte für eine gute Passung sowie für alle Fitmaße ( $\chi^2/df = 1.50$ , RMSEA = 0.05, CFI = 0.98, TLI = 0.97). Die Berechnung der Passung des Nullmodells zur Prüfung der Response-Shift-Prozesse, in dem alle relevanten Parameter restringiert wurden, ergab Werte, die auf eine inakzeptable Passung zwischen den empirischen Daten und dem theoretischen Modell hinweisen ( $\chi^2/df = 5.30$ , RMSEA = 0.14, CFI = 0.76, TLI = 0.71). Dieser erste Modellvergleich stellt die Voraussetzung für die weitere parameterbezogene Response-Shift-Erfassung dar. Der Unterschied zwischen den Passungen beider Modelle (Basismodell, Nullmodell) wird als Beleg dafür angesehen, dass zwischen den MZPen ein Response-Shift stattgefunden hat. Nach Abschluss der Prüfung aller Response-Shift-Parameter wurde ein Modell definiert, in dem insgesamt acht Parameter freigesetzt wurden. Für dieses Response-Shift-Modell lassen sich Werte für die Passung berechnen, die wiederum als gut zu bezeichnen sind ( $\chi^2/df = 1.31$ , RMSEA = 0.04, CFI = 0.99, TLI = 0.98). Die acht Response-Shift-Parameter sind aufgrund der Art der Parameter als uniforme und non-uniforme Rekalibrierung zu bezeichnen. *Die Intercepts der Skalen Physische, Kognitive und Rollenfunktionsfähigkeit sowie die Messfeh-*

*lervarianzen aller Skalen sind freizusetzen.* In der Gesamtstichprobe sind die Restringierungen der Faktorladungen beizubehalten. *Demgemäß findet keine Repriorisierung statt.* Das bedeutet gleichzeitig, dass *eine Neukonzeptualisierung ebenfalls nicht nachzuweisen war.*

Ein weiteres Ergebnis ist darin zu sehen, dass sich die GLQ zwischen den beiden MZPen statistisch signifikant verbessert, wobei die Veränderung der Varianz der GLQ zwischen beiden MZPen nicht statistisch bedeutsam war (Mittelwert der GLQ zum MZP 1 = 0.00, zum MZP 2 = 0.56;  $p < .001$ ; Varianz GLQ zum MZP 1 = 1.00, zum MZP 2 = 0.95, Vergleich der Modelle:  $C_{min} = 0.22$ ,  $p = 0.64$ ).

Findet eine Rekalibrierung statt, ist für den entsprechenden Indikator davon auszugehen, dass sich der beobachtete Effekt aufgrund der herkömmlichen Mittelwertsberechnung aus zwei Komponenten zusammensetzt: Einem Anteil, der auf die Veränderung der GLQ als latentem Konstrukt zurückzuführen ist, sowie einem zweiten Anteil, der aufgrund der Rekalibrierung zustande kommt. Anhand der Effektkalkulation nach Oort (2005a) lassen sich diese beiden Anteile konkret berechnen. Die beobachteten Effekte (aufgrund der herkömmlichen Kalkulation) werden auf den Skalen Physische, Kognitive und Rollenfunktionsfähigkeit in deutlichem Umfang durch einen Response-Shift mit bedingt. Auf der Skala Rollenfunktionsfähigkeit setzt sich der beobachtete Effekt von 0.81 zusammen aus einem Response-Shift-Effekt in Höhe von 0.55 und einem Effekt durch die Veränderung der GLQ von 0.26. Auf der Skala Physische Funktionsfähigkeit lässt sich ein beobachteter Effekt von 0.60 bestimmen, der zu einem Anteil von 0.33 auf einen Response-Shift sowie von 0.27 auf die tatsächliche Veränderung der GLQ zurückzuführen ist. Auf der Skala Kognitive Funktionsfähigkeit zeigt sich ein beobachtbarer Effekt von 0.05, wobei dieser durch einen negativen Response-Shift-Effekt in Höhe von -0.22 sowie einen Effekt aufgrund der wahren Veränderung der GLQ von 0.27 zustande kommt. Auf den Skalen Physische und Rollenfunktionsfähigkeit führt der Response-Shift somit zu einer Überschätzung der Veränderungen beziehungsweise Effekte, die auf Veränderungen durch die GLQ zurückzuführen sind. Auf der Skala Kognitive Funktionsfähigkeit resultiert aufgrund des Response-Shifts eine Unterschätzung der Veränderung bedingt durch die GLQ.

### 8.1.2 Diskussion der Ergebnisse der Fragestellung 1

Aufgrund der Darstellung der Ergebnisse ist die Fragestellung 1, ob ein Response-Shift in der vorliegenden Stichprobe aufgetreten ist, zu bejahen. Die Freisetzung der genannten Response-Shift-Parameter wird als Beleg für das Auftreten dieses Phänomens gewertet, wobei die statistische Bedeutung der Parameter gleichzeitig die Art des Response-Shifts definiert. Im Folgenden soll zunächst beschrieben werden, wie die Rekalibrierung statistisch zu erklären ist. Anschließend werden inhaltliche Erklärungen für die Rekalibrierung vorgestellt.

#### *Diskussion von Response-Shift in der vorliegenden Studie im Zusammenhang mit allgemeinen methodischen und statistischen Aspekten*

Nach dem Linear Latent Variable Model (LLVM) von Oort (2004) wird eine Rekalibrierung erfasst durch die Veränderung der Intercepts beziehungsweise Messfehlervarianzen. Diese beiden Parameterklassen operationalisieren den Einfluss der spezifischen Faktoren, die unabhängig vom latenten Konstrukt der GLQ sind und auf den Indikator der GLQ beziehungsweise die manifeste Variable einen Einfluss haben.

Anders als in der Klassischen Testtheorie (KTT) geht Oort davon aus, dass sich in den Messfehlern nicht nur alle zufälligen Einflussfaktoren auf die Bewertung eines Indikators eines latenten Konstrukts auswirken, wie beispielsweise situative Konzentrationsfähigkeit, gegenwärtiges Befinden oder Merkmale wie Tageszeit. Nach dem LLVM ist die Reduktion der Messfehlervarianz nicht, wie nach der KTT, eine Reduktion der zufälligen Einflussfaktoren auf die manifeste Variable, sondern ein Hinweis auf eine Veränderung spezifischer Einflussfaktoren, die relativ hohe Stabilität über die Zeit haben *können*, aber nicht zwangsläufig haben müssen. Statistisch gesehen bedeutet eine Rekalibrierung nach Oort eine Veränderung der Relation zwischen diesen spezifischen Faktoren und der GLQ als latentem Konstrukt, die sich auf die Beantwortung von Items auswirkt. Das Messmodell der GLQ ist die Grundlage der Berechnung der empirischen Varianz/Kovarianzmatrix. Das latente Konstrukt soll durch die Indikatoren angemessen erfasst werden. Aufgrund der Varianz/Kovarianzmatrix als Verteilungsgrundlage der Parameterschätzung ergeben sich Varianzverhältnisse zwischen spezifischen Faktoren und der GLQ. Diese unterscheiden sich zwischen den MZPen für die Beantwortung eines Items. Die tatsächlich vorgegebene Skalenabstufung (das Instrument) für die Stichprobe ist jedoch identisch. Somit ist davon auszugehen, dass sich die Metrik des Items

beziehungsweise der Skala dieses Indikators *für die GLQ* verändert. Die Frage nach der Art des Response-Shifts lässt sich aufgrund der vorliegenden Daten beantworten. Nach dem LLVM nach Oort (2004) zeigt sich eine non-uniforme Rekalibrierung *entweder* nur in der Form der Freisetzung der Messfehlervarianz *oder* in der Freisetzung der Messfehlervarianz und des Intercept einer Skala. Eine uniforme Rekalibrierung zeigt sich darin, dass *lediglich* eine Interceptfreisetzung vorliegt. Die Messfehlervarianz bleibt in diesem Fall restringiert. In der vorliegenden Arbeit ist auf den Indikatoren gleichzeitig eine Freisetzung der Messfehlervarianzen und der Intercepts vorzunehmen. Damit ist davon auszugehen, dass eine non-uniforme Rekalibrierung entscheidend ist für die Freisetzung der Intercepts.

Aufgrund dieser Annahmen ist von folgendem Response-Shift in der Gesamtstichprobe auszugehen: Die Veränderung der Bewertungen einzelner Skalen- beziehungsweise Itemabstufungen findet bei allen Indikatoren statt. Auf den beiden Skalen Soziale und Emotionale Funktionsfähigkeit findet eine non-uniforme Rekalibrierung statt, die sich jedoch nicht auf die Intercepts und somit auf die Mittelwertebene auswirkt. Der Response-Shift ist statistisch nachweisbar, bleibt jedoch ohne Auswirkungen auf die Kalkulation der Effekte. Aufgrund der erheblichen Varianzreduktion auf diesen Skalen, die jedoch die Intercepts nicht beeinflusst, ist möglicherweise davon auszugehen, dass die Änderungen der Itemabstufungen eher im Extrembereich der Skalen (Tendenz zu Extremwerten) stattgefunden haben und bei der zweiten Erhebung nach drei Monaten für einen vergleichbaren Zustand gemäßigte Einschätzungen abgegeben wurden. Ein großer Teil der Stichprobe bewertet jedoch (objektiv) vergleichbare Zustände (wahrscheinlich im mittleren Bereich der Skala) auch subjektiv mit dem gleichen Skalenwert. Auf den Skalen Physische, Kognitive und Rollenfunktionsfähigkeit ist die non-uniforme Rekalibrierung in einem solchen Ausmaß aufgetreten, dass auch die Intercepts der Indikatoren betroffen und frei zu setzen sind. Die non-uniforme Rekalibrierung auf diesen drei Skalen hat somit für die weitere Berechnung von Veränderungswerten statistische Relevanz.

Von Bedeutung für die beschriebenen Ergebnisse bezüglich der Strategie der Response-Shift-Erfassung ist die Frage, ob auch andere Prozesse eine Veränderung der Messfehlervarianz bedingen können. Hier sei aus statistischer Perspektive insbesondere auf die Bedeutung der *Regression zur Mitte* als statistisches Artefakt hingewiesen. Der vorliegenden Arbeit liegt die Operationalisierung des Messfehlers als Varianz des spezifischen Faktors zugrunde, der unabhängig vom latenten Konstrukt GLQ auf die manifeste Variable wirkt. Wenn die Mess-

fehlervarianzen die Varianzen der spezifischen Faktoren der einzelnen Skalen ausdrücken, bedeutet dies, dass die Stichprobe bezüglich der spezifischen Faktoren der einzelnen Skalen zum MZP 2 homogener wird, da die Variation dieser Faktoren abnimmt. Aufgrund der vorliegenden Ergebnisse ist zu überlegen, ob diese Varianzreduktion tatsächlich durch veränderte Bewertungsprozesse oder durch eine Regression zur Mitte beeinflusst wird. Nach Rogosa und Willett (1985) ist die Regression zur Mitte empirisch dadurch zu prüfen, indem eine Korrelation zwischen dem Prä-Wert und dem Differenzwert aus dem Prä-Post-Vergleich berechnet wird. Bei Vorliegen einer negativen Korrelation ist von einer Regression zur Mitte auszugehen. In der vorliegenden Arbeit sind alle Korrelationen zwischen den jeweiligen manifesten Skalenwerten und den Differenzwerten dieser Skalenwerte im Prä-Post-Vergleich negativ (siehe Anhang A3). Dieses Ergebnis spricht für einen Effekt der Regression zur Mitte, der die Auswertungen beeinflusst. Grundsätzlich ist jedoch nach Zwingmann (2003) ein solcher Effekt in empirischen Untersuchungen sehr häufig nachgewiesen und auch zu erwarten. Ebenfalls ist der Nachweis eines Regressionseffekts dieser Art nicht gleichbedeutend mit der Unwirksamkeit der Intervention beziehungsweise der geringen Bedeutsamkeit von Response-Shift-Einflüssen. Die Berechnung der Korrelationen sowie die Höhe der Ausgangswerte der Prämessung sprechen dafür, dass ein Effekt der Regression zur Mitte an der Veränderung der Werte nicht auszuschließen ist. Jedoch ist zwischen den Regressionseffekten und den Response-Shift-Prozessen in der vorliegenden Stichprobe kein eindeutiger Zusammenhang erkennbar. Beispielsweise sind auf den Skalen Physische und Soziale Funktionsfähigkeit die Effekte der Regression zur Mitte in etwa vergleichbar, jedoch findet bei diesen beiden Skalen eine Interceptveränderung nur auf der Skala Physische Funktionsfähigkeit statt.

Weiterhin ist die Definition der Messfehlervarianz als Varianz des spezifischen Faktors eines Indikators zu betrachten. Einerseits ist diese Annahme gerade bei Longitudinaldaten gut nachvollziehbar, da zu beiden MZPen dieselben Personen mit demselben Instrument untersucht werden. Dass in solchen Fällen nicht mehr von einer Unabhängigkeit der ‚Messfehler‘ ausgegangen werden kann, ist nachvollziehbar. Auch drückt sich diese Annahme in der praktischen Umsetzung der Invarianztestung aus, da die Messfehler eines Indikators über zwei Messzeitpunkte miteinander korreliert sind (Bollen 1989). Dennoch stellt sich die Frage nach dem Einfluss des klassischen Messfehlers auf die Erfassung einer non-uniformen Rekalibrierung, da die Messfehler zwar nach den Axiomen der KTT und dem LLVM unkorreliert sind, aber dennoch bei einer Messung grundsätzlich auftreten. In der Berechnung der Varianzan-

teile sind diese Einflüsse enthalten. Die Größe des Anteils des klassischen Messfehlers auf eine non-uniforme Rekalibrierung kann bisher nicht genau kalkuliert werden.

Die Messfehlerkorrelationen zwischen zwei Indikatorvariablen zu *einem* MZP lassen sich vor diesem Hintergrund ebenfalls inhaltlich auswerten. Dieser Korrelation liegt nach dem LLVM (Oort 2004) die Annahme zugrunde, dass es einen spezifischen Faktor gibt, der für zwei Indikatorvariablen zu einem MZP relevant ist. Eine Response-Shift-Prüfung kann auch in Bezug auf diese Messfehlerkorrelation erfolgen. Mittels der Restriktion der Messfehlerkorrelation für beide MZPe kann geprüft werden, ob zwischen den MZPen diesbezüglich ein Response-Shift auftritt. Es würde sich dabei um eine Repriorisierung auf ‚unterer‘ Ebene handeln (im Gegensatz zu einer Repriorisierung auf höherer Ebene, wenn die restringierte Kovarianz zwischen zwei latenten Konstrukten zwischen zwei MZPen freizusetzen ist). Die im Messmodell zugelassenen Messfehlerkorrelationen wurden auf Invarianz zwischen den MZPen getestet. Das Ergebnis dieser zusätzlichen Auswertung ergab, dass diese zwischen den MZPen nicht variierten (verglichen wurde jeweils der Unterschied zwischen dem Response-Shift-Modell und dem Modell auf Testung der Messfehler-Restriktion; für die Messfehlerkorrelation der Skalen Physische und Rollenfunktionsfähigkeit ergab sich ein Unterschied zwischen beiden Modellen (mit und ohne Messfehlerrestriktion) von  $CMIN = 0.627$ ,  $p = .428$ ; für die Messfehlerkorrelation der Skalen Soziale und Rollenfunktionsfähigkeit ergab sich ein Unterschied zwischen beiden Modellen (mit und ohne Messfehlerrestriktion) von  $CMIN = 3.462$ ,  $p = .063$ ). Dieses Vorgehen weist auf die Möglichkeit hin, auch die spezifischen Faktoren nach dem LLVM statistisch und inhaltlich zu untersuchen. Im vorliegenden Fall bieten sich diese Messfehlerkorrelationen / spezifischen Faktoren als weiter zu analysierende Variablen an, die auf die Lebensqualitätsmessung mittels des EORTC-QLQ-C30 bei PCa-Patienten in der AHB einen Einfluss haben. Zur weiteren Klärung dieser Frage sind inhaltliche Präzisierungen der spezifischen Einflussfaktoren von Bedeutung, um in zukünftige Modelltestungen als Variablen aufgenommen zu werden.

Ebenso ist für die Anwendung der KFA zur Response-Shift-Erfassung die Güte der Passung zwischen Modell und empirischer Datenbasis kritisch zu betrachten. Bei der KFA handelt es sich um einen konfirmatorischen Ansatz, bei dem ein zuvor definiertes Modell und die Varianz/Kovarianz-Struktur einer empirischen Datenmatrix auf Passung geprüft werden. Die vorliegende Arbeit orientiert sich hinsichtlich der Beurteilungen von Messmodellen an allgemeinen Richtlinien (Schermelleh-Engel et al. 2003). Wie sich eine mangelnde Passung des

Messmodells auf die Überprüfung von Response-Shift auswirken kann, ist am Beispiel einer Untersuchung von Ahmed und Kollegen (2005b) zu sehen. Die Autoren berichten von einer mangelnden Übereinstimmung zwischen den Ergebnissen der Response-Shift-Erfassung verschiedener Strategien. Verglichen wurden die KFA-Methode mit dem Then-Test und einer individualisierten Response-Shift-Erfassung. Anhand der KFA-Methode konnte kein Response-Shift nachgewiesen werden, jedoch mit den beiden anderen Strategien. Dieses Ergebnis steht im Widerspruch zu einer vergleichbaren Untersuchung von Visser und Kollegen (2005), in der eine angemessene Übereinstimmung zwischen Thentest und KFA-Methode nachgewiesen wurde. Eine mögliche Ursache für das Studienergebnis ist darin zu sehen, dass eine akzeptable Passung des Messmodells des SF-36 und der empirischen Datenbasis in den untersuchten Stichproben von Ahmed (Schlaganfall und Kontrollgruppe) bereits auf Gruppenebene nicht gegeben ist (Satorra  $\chi^2/df = 4$ , RMSEA = .12, CFI = .92, Ahmed et al 2005b). Dennoch wurde die Methode zur Response-Shift-Erfassung eingesetzt. Zwar weist das Basismodell (Modell aus 2 MZPen ohne Restriktionen) insgesamt eine leicht bessere Passung auf (Satorra  $\chi^2/df = 2.4$ , RMSEA = .08, CFI = .93), jedoch sind diese Werte noch immer als kritische Passung des Gesamtmodells zu bezeichnen. Grundsätzlich sollte bei der Verwendung der KFA-Methode zur Überprüfung von Response-Shift-Prozessen von einem Modell mit akzeptabler bis guter Passung ausgegangen werden (Borsboom et al. 2007). Unklar ist gegenwärtig noch die Bedeutung des Einflusses der Passung zwischen theoretischem Modell und empirischen Daten. In der vorliegenden Arbeit ist die Passung als gut bis akzeptabel zu bezeichnen, so dass die Frage, ob das Modell geeignet ist, auf Response-Shift zu testen, bejaht werden kann. Wie Response-Shift-Prozesse in Abhängigkeit von der Gesamtpassung des Modells variieren, ist jedoch bisher nicht geprüft.

In diesem Zusammenhang ist ebenfalls interessant, welche Strategie gewählt wird, um Parameterfreisetzungen zu prüfen. Die in der vorliegenden Arbeit eingesetzte Strategie mit Hilfe der Modifikationsindizes ist inhaltlich und statistisch begründbar sowie praktikabel. In diesem Vorgehen zeigt sich der wesentliche Unterschied zu dem Vorgehen bei der Prüfung von Datenmatrizen auf Invarianz von Messungen bei Gruppenvergleichen, da in einem solchen Fall die Reihenfolge der Parameterfreisetzung vorher definiert ist. Es wird nicht nach freisetzbaren Parametern ‚gesucht‘. Schmitt (1982) überträgt diesen Ansatz auf längsschnittliche Testungen von Messmodellen. Die beiden Vorgehen haben jedoch unterschiedliche Auswirkungen auf die Berechnung der Varianz/Kovarianzmatrix. Eine Vergleichsstudie zur Untersuchung von Response-Shift nach der Strategie von Oort (2004) und Schmitt (1982) hat zu

vollkommen unterschiedlichen Ergebnissen geführt (Ahmed et al. 2009). Während nach dem Oort-Ansatz Response-Shift-Prozesse identifiziert werden konnten, war dies nach dem Schmitt-Ansatz nicht der Fall. Dies kann auf einen Unterschied im Vorgehen der Response-Shift-Erfassung zurückgeführt werden: Schmitt beginnt seine Response-Shift-Analyse mit einem Modell, in dem alle Parameter freigesetzt sind und setzt *anschließend* Restriktionen auf einzelnen Parameterebenen ein, die inhaltlich als unterschiedlich bedeutsam für die Invarianztestung angesehen werden. Der Ansatz nach Oort (2004) ist genau entgegengesetzt. Zu Beginn der Response-Shift-Testung steht ein vollständig restringiertes Modell und es werden die Parameter freigesetzt, die aus statistischen Gründen zu einer Verbesserung der Modellpassung führen. Daraus resultiert, dass der Ansatz nach Schmitt weniger sensitiv für Response-Shift ist, da spezifische Parameterebenen getestet werden. Andererseits ist am Ansatz nach Oort zu bedenken, dass durch die höhere Sensitivität des Ansatzes für Parameteränderungen auch die Typ-I-Fehlerwahrscheinlichkeit steigt. Das bedeutet, dass nach dem Oort-Ansatz die Gefahr erhöht ist, einen Response-Shift zu identifizieren, obwohl dieser gar nicht vorhanden ist, weil die Messwertdifferenzen auf andere Einflussfaktoren zurückzuführen sind (bspw. Regression zur Mitte im Falle einer non-uniformen Rekalibrierung).

Ein weiterer Aspekt, der kritisch zu diskutieren ist, ist die Analyse des Response-Shifts auf Gruppenebene. Dies wird von einigen Autoren als wesentlicher Nachteil der in dieser Arbeit verwendeten methodischen Strategie der Response-Shift-Erfassung bezeichnet (bspw. Mayo et al. 2008). Die Vertreter der Oort-Methode im Rahmen der Response-Shift-Erfassung gehen von der Annahme aus, dass wenn bei vielen Personen innerhalb einer untersuchten Stichprobe eine vergleichbare Veränderung stattfindet, diese auch auf Gruppenebene abbildbar ist. Wenn beispielsweise eine Repriorisierung erfolgt, also die Bedeutung der latenten Variable für den Indikator zu oder abnimmt, dann ist dies nur dann mittels KFA statistisch nachweisbar, wenn dieser Prozess bei möglichst vielen Personen der untersuchten Stichprobe auftritt. Diese Art der Response-Shift-Erfassung ist nur dann zielführend, wenn angenommen werden kann, dass eine möglichst große Anzahl der Studienteilnehmer diesen Prozess erleben. Die Methode ist auf Gruppenebene durchaus sinnvoll anwendbar, ermöglicht jedoch keine Aussagen über das Ausmaß eines individuellen Response-Shift. Es lassen sich auch Szenarien konstruieren, in denen Response-Shift auf Gruppenebene auftritt und nicht mit der KFA-Methode identifiziert werden kann. Einerseits ist es möglich, dass einzelne Personen innerhalb der Stichprobe zwischen den MZPen beispielsweise eine Rekalibrierung produzieren, diese jedoch in unterschiedliche Richtungen ausfallen (Über- und Unterschätzungen von



Werten). Im Mittel würde sich diese Veränderung wiederum nivellieren, so dass diese Prozesse mittels Gruppenstatistik nicht identifizierbar wären. Andererseits ist ein Response-Shift ebenfalls dann nicht erfassbar, wenn auf jeder Indikatorvariablen eine Rekalibrierung in gleichem Ausmaß stattfindet. In diesem Fall ließe sich die Rekalibrierung nicht mehr von der Veränderung des latenten Konstrukts trennen, da diese eben aus der Kovarianz der Indikatoren beziehungsweise der Kovarianz der Indikatoren mit dem latenten Konstrukt abgeleitet wird.

In diesem Zusammenhang ist auf die Unterscheidung zwischen der methodischen und der inhaltlichen Perspektive auf einen Response-Shift hinzuweisen. Bei dem Vorgehen nach Oort (2004) handelt es sich um einen rein statistischen Ansatz. Die Veränderungen der Response-Shift-Parameter sind *aufgrund ihrer statistischen Definition* Operationalisierungen für Response-Shift-Prozesse. Eine Reprioritisierung ist Kennzeichen eines Umbewertungsprozesses, weil dies aufgrund der Bedeutung der statistischen Parameter der Faktorladungen angenommen wird: Das Gewicht der latenten Variablen auf den Indikator verändert sich zwischen zwei MZPen. Die *Ursache* für die Parameterveränderung ist jedoch unklar. Fraglich ist, ob diese Veränderung auch durch andere Einflussfaktoren als einen Response-Shift-Prozess zustande kommen kann. Hierin zeigt sich die unterschiedliche Definition von Response-Shift aus methodischer und inhaltlicher Perspektive: Unter der methodischen Perspektive ist ein Response Shift ein Phänomen, das die Messung von Veränderungen beeinflusst beziehungsweise Parameterveränderungen beschreibt. Im Vordergrund dieser Perspektive steht die Beeinflussbarkeit einer Messung anhand eines Items oder einer Skala durch potenziell veränderbare subjektive Bewertungskriterien. Die inhaltliche Perspektive auf einen Response-Shift nach Sprangers und Schwartz (1999) betrachtet die Gründe für die Veränderung eines Bewertungshintergrundes eines zentralen Konstrukts, die sich in Form der Rekalibrierung, Reprioritisierung und Neukonzeptualisierung ausdrücken kann. Der Unterschied zwischen den beiden Perspektiven ist darin zu sehen, dass sich der methodische Ansatz mit dem Phänomen der *Beeinflussung von Messungen* beschäftigt, der inhaltliche mit der *Beeinflussung von latenten Konstrukten beziehungsweise manifesten Faktoren*, die latente Konstrukte definieren. Der methodische Zugang prüft Messungen auf Invarianz, der inhaltliche sucht *theoretische Erklärungen* für einen Response-Shift. Sollte im Verlauf einer PCa-Erkrankung und Behandlung bei einem Betroffenen ein Response-Shift auftreten, dann ist es wegen der eben beschriebenen methodischen Definition unerheblich, ob dies durch das individuelle Erleben und Verarbeiten der Erkrankung mit allen Bewältigungsaspekten an sich bedingt ist oder ob bei-

spielsweise eine Schulungsmaßnahme dazu beiträgt, die eigene Lebenssituation und -perspektive anders zu bewerten. Für die inhaltliche Perspektive ist dies jedoch das entscheidende Kriterium. In beiden Fällen verändert sich der Bewertungshintergrund für die Beurteilung eines zentralen Konzepts. Deshalb handelt es sich in beiden Fällen um einen Response-Shift. Die Veränderung ist inhaltlich unterschiedlich erklärbar, jedoch in beiden Fällen mit Hilfe der Oort-Methode operationalisierbar.

### *Diskussion von Response-Shift im Zusammenhang mit Ansätzen der Veränderungsmessung*

Nach dem LLVM (Oort 2004) bedeutet die Freisetzung der Messfehlervarianz eine non-uniforme Rekalibrierung. Ist diese non-uniforme Rekalibrierung deutlich ausgeprägt, beeinflusst diese den Intercept eines Indikators und hat nach der Formel von Oort (2005a) Auswirkungen auf die Kalkulation der Effekte. Dies ist in der vorliegenden Arbeit bei den Indikatoren Physische, Kognitive und Rollenfunktionsfähigkeit der Fall. Die Darstellung der Response-Shift-Effekte aufgrund der non-uniformen Rekalibrierung macht deutlich, dass die Berechnung der beobachteten Effekte anhand herkömmlicher Effektkalkulationen durch einen Response-Shift beeinflusst ist. Der große beobachtete Effekt auf der Skala Rollenfunktionsfähigkeit wird zu einem großen Teil (mittlerer Effekt) durch einen Response-Shift verursacht und nur zu einem kleineren Teil (kleiner Effekt) durch die Veränderung der GLQ. An diesem Ergebnis wird die Bedeutung der Response-Shift-Erfassung für die Veränderungsmessung deutlich. Die häufigste Art, Veränderungen mittels eines Prä-Post-Designs zu erfassen, ist die, anhand der (Mittel-)Werte einer Variablen zu zwei MZPen einen Differenzwert zu berechnen und diese Differenz an bestimmten Streukoeffizienten zu relativieren. In der Praxis wird davon ausgegangen, dass sich auf diese Art Effekte, beispielsweise einer Interventionsmaßnahme, bestimmen lassen. In einem weiteren Schritt wird häufig diskutiert, wann diese Veränderung beziehungsweise dieser Effekt auch praktische Bedeutsamkeit beziehungsweise klinische Relevanz besitzt. Aufgrund der vorliegenden Ergebnisse zur Kalkulation der Effekte unter Berücksichtigung des Response-Shifts zeigt sich, dass neben der Bestimmung des allgemeinen quantitativen Ausmaßes eines Effektes auch eine differenzierte Aussage über Veränderung verschiedener Faktoren möglich ist. Somit ist die KFA-Methode zur Response-Shift-Untersuchung auch im Kontext der Änderungssensitivität zu diskutieren. Darauf soll im Folgenden nochmals eingegangen werden.

Um Aussagen über Veränderungen von Variablen aufgrund von Messungen treffen zu können, sollten die eingesetzten Instrumente neben den klassischen Gütekriterien auch das der Änderungssensitivität erfüllen (Kirshner & Guyatt 1985). Diese ist für die Verwendung von Instrumenten im Rahmen von evaluativen Studien ein wichtiges Gütekriterium und wird in der Regel über die Kalkulation von Effekten operationalisiert (Guyatt et al. 1992). Dabei kann Änderungssensitivität unterschiedlich weit gefasst werden (Terwee et al. 2003). Das Spektrum bezieht sich auf (a) Veränderungen allgemein, (b) klinisch bedeutsame Veränderungen für eine Person beziehungsweise eine Gruppe von Personen sowie auf (c) die Erfassung der Veränderung eines zugrundeliegenden *true scores*. Ist es das Ziel zu klären, *ob* eine Veränderung (bspw. auf der Skala Physische Funktionsfähigkeit) stattgefunden hat, dann genügen die bekannten Effektstärkekalkulationen (Gruppe a und b, bspw. ES, SRM, GRM). Soll jedoch eine Aussage *über* die Veränderung der Werte einer Skala getroffen werden, dann sind diese herkömmlichen Strategien der Berechnung der Änderungssensitivität nicht geeignet, da sie keine differenzierte Aussagen über die erfassten Veränderungen auf einer Skala zulassen. Die von Oort (2005a) vorgestellte Effektkalkulation im Rahmen der Response-Shift-Erfassung ermöglicht jedoch eine Zerlegung der Effekte, insbesondere bezüglich einer Veränderung bedingt durch den *true score* (durch das latente Konstrukt) und einer Veränderung *aufgrund des Einflusses spezifischer Faktoren auf den Indikator, die unabhängig sind vom true score*. Die Response-Shift-Erfassung mittels KFA ist daher nicht nur geeignet, Aussagen über die Vergleichbarkeit zweier Messungen im Längsschnitt zu treffen, sondern bietet auch die Möglichkeit, beobachtbare Effekte genauer zu spezifizieren. Der besondere Vorteil für die Erfassung der Veränderung eines latenten Konstrukts mittels des hier durchgeführten KFA-Ansatzes liegt darin, dass kein weiteres Außenkriterium einbezogen werden muss, um einen *true score* abzuleiten. Die Veränderungen durch den *true score* und durch Response-Shift lassen sich statistisch aus zwei einfachen Messungen berechnen. Da der KFA-Ansatz beobachtbare Veränderungen in Beziehung setzt zu Veränderungen des latenten Konstrukts, kann die KFA-Methode zur Response-Shift-Erfassung von Oort (2005a, Oort et al. 2005) der nach Terwee und Kollegen (2003) beschriebenen Gruppe der Strategien zugeordnet werden, die Änderungssensitivität auf der Ebene der wahren Veränderung eines *true scores* erfassen. Aufgrund der Ergebnisse der Effektzerlegungen im Rahmen der Auswertungen der Fragestellung 1 wird deutlich, dass die Indikatorkalen Physische, Kognitive und Rollenfunktionsfähigkeit des EORTC-QLQ-C30 nur zum Teil durch die GLQ als latentem Konstrukt beeinflusst werden. Die Änderungssensitivität der Indikatorkalen, im Sinne der Erfassung von Veränderungen durch

die GLQ, ist von daher durch die üblichen Effektkalkulationen nur ungenau abbildbar. Dies gilt im vorliegenden Fall nicht im selben Maße für die beiden Skalen Emotionale und Soziale Funktionsfähigkeit. Die Veränderungen auf diesen Skalen scheinen eher durch die Veränderung der GLQ bedingt zu sein (Faktorladungen der Skalen zu beiden MZPen  $\lambda > .66$ ). Somit können diese Skalen – im Vergleich zu den drei response-shift-beeinflussten Skalen – auch als änderungssensitiver in Bezug auf die Veränderung der GLQ als latentes Konstrukt angesehen werden.

Der Response-Shift auf der Skala Kognitive Funktionsfähigkeit weist auf einen weiteren Aspekt der Erfassung von Veränderungen hin. In der vorliegenden Arbeit hätte die Veränderung der GLQ zu einer positiveren Bewertung der Kognitiven Leistungsfähigkeit zum MZP 2 geführt, wenn diese nicht durch einen Response-Shift überdeckt würde. Somit ist auf der Ebene der GLQ ein Effekt nachweisbar. Daran wird ersichtlich, dass die Response-Shift-Erfassung mittels KFA dazu beitragen kann, Effekte aufzudecken, die auf der Ebene beobachteter Effekte nicht nachweisbar sind. Dies ist in erster Linie deshalb von Bedeutung, da weiter hinterfragt werden kann, warum eine solche ‚Kompensation‘ von Effekten stattgefunden hat. Wenn die Ursachen für den Response-Shift auf der Skala Kognitive Funktionsfähigkeit genauer expliziert und durch eine Intervention beeinflusst werden könnte, wäre es durchaus möglich, einen ‚beobachtbaren‘ Effekt auf diesem Indikator zu erzielen. Somit ist die statistische Bedeutung der Differenzierung von Effekten auch für praktische Fragestellungen relevant.

#### *Diskussion inhaltlicher Aspekte im Zusammenhang mit Response-Shift in der vorliegenden Studie*

Bei einer Rekalibrierung ist nach der Definition von Sprangers und Schwartz (1999) von einer quantitativen Umbewertung einzelner oder aller Skalenabstufungen und eine damit einhergehende Veränderung der Metrik der Skala auszugehen. Gleiche subjektive Belastungen beziehungsweise selbstbewertete Beeinträchtigungen werden zu zwei MZPen unterschiedlich skaliert. Ursächlich dafür können Erfahrungen sein, die eine Person dazu veranlassen, die vormals getroffene Einschätzung der Schwere eines Belastungsgrads beziehungsweise die subjektiv wahrgenommene Belastung aufgrund von neuen Beeinträchtigungen zu relativieren. Im Sinne eines Response-Shifts wird oftmals beschrieben, dass die Bewertung eines Zustands, beispielsweise der Einschätzung von Schmerzen, dadurch geändert werden kann, dass

eine Person in der Zwischenzeit (schmerzhafte) Erfahrungen macht, die sie vorher nicht kannte, und somit die in der Vergangenheit erlebten Erfahrungen vor dem neuen Bewertungshintergrund als weniger ausgeprägt beziehungsweise belastend beurteilen würde. Derselbe Zustand wird somit mit einem anderen – im beschriebenen Fall geringeren – Wert quantifiziert. Die inhaltliche Ursache für den Prozess der Rekalibrierung kann entweder das Selbsterleben (wie oben beschrieben) oder die Beobachtung anderer sein, die wiederum belasteter oder weniger belastet sein können (Soziale Vergleichsprozesse) (Norman 2003).

Es stellt sich die Frage, worin die inhaltlichen Ursachen für die Veränderung der Intercepts beziehungsweise Bewertungshintergründe bei den Patienten der vorliegenden Stichprobe gesehen werden können. Denkbar erscheint, dass die befragten PCa-Patienten zum MZP 1 ihren Zustand kritischer beurteilen, als dieser objektiv war. Kurz nach der Akutbehandlung prägen möglicherweise physische Beeinträchtigungen und körperliche Belastungen die Einschätzungen der Einzelitems stärker als zum MZP 2, wenn die Personen mit diesen Einschränkungen beziehungsweise Beeinträchtigungen besser zu Recht kommen. Möglicherweise sehen die Patienten jedoch nach einigen Monaten eher die positiven Entwicklungen seit Beginn der Behandlung und rücken diese für die Bewertungen zum MZP 2 stärker in den Vordergrund. Denkbar wäre auch, dass die Befragten im Verlauf der AHB einen Vergleich mit Patienten ziehen, denen es schlechter geht, und somit zum MZP 2 den eigenen ‚schlechten‘ Zustand als positiver bewerten (sozialer Abwärtsvergleich, Norman 2003). Andererseits ist ebenfalls denkbar, dass sich die Referenzpunkte für eine Beurteilung der Fragen qualitativ verändert haben. Möglicherweise liegen den beiden Einschätzungen unterschiedliche Vergleichsprozesse zugrunde. Zum MZP 1 hat eher ein individueller Vergleich mit der eigenen Vergangenheit (und den damit zusammenhängenden Möglichkeiten, Empfindungen und Belastungen) den Bewertungshintergrund für die Beantwortung der Fragen geprägt (individueller Vergleichsmaßstab). Zum MZP 2 war möglicherweise der soziale Vergleich mit anderen Patienten während der Reha, mit Menschen, die an anderen oder schwereren Erkrankungen leiden, entscheidend, deren Zustand noch schlechter eingeschätzt wird (sozialer Vergleichsmaßstab). Auch ist denkbar, dass in beiden Fällen eine Bewertung vor dem individuellen Hintergrund vorgenommen, jedoch zu beiden MZP eine andere Metrik verwendet wird. Dies soll an einem Beispiel verdeutlicht werden: Eine Person könnte zum MZP 1 den eigenen körperlich-funktionalen Zustand als ‚stark beeinträchtigt‘ und mit einem subjektiven Skalenwert von schätzungsweise 7 (10 = maximale Beeinträchtigung) bewerten. Der individuellen Beurteilung liegt demnach eine Art numerische Beeinträchtigungsskala von (möglicherweise)

‚gar nicht‘ bis ‚extrem‘ zu Grunde. Zum MZP 2 befindet sich die Person im selben objektiven Zustand und bewertet diesen jedoch nicht mehr vor dem subjektiven Hintergrund einer *Beeinträchtigungsskala*, sondern vor dem einer persönlichen *Prozentskala*, wie aktiv sie zum aktuellen Zeitpunkt Aktivitäten im täglichen Leben bewältigen kann. Möglicherweise schätzt die Person für die körperlich-funktionalen Beeinträchtigungen nun ein, dass sie ihre optimale Leistungsfähigkeit nur zu ‚50 Prozent‘ im Alltag erreicht. Obwohl es ihr objektiv zu beiden MZPen gleich geht, würde sich die Person zum MZP 1 einen Wert im negativen Drittel der Skala geben. Zum MZP 2 kreuzt die Person den Mittelwert der Skala an.

Auch eine Rolle könnte die Anforderungssituation der jeweiligen Umwelten zu beiden MZPen spielen, wobei auch diese die Bewertungshintergründe für die Items beeinflussen können. Dass der aufgedeckte Response-Shift dadurch erklärbar wäre, dass die Personen einmal in der Klinik und später in der eigenen Wohnumwelt Auskunft zur Physischen beziehungsweise Rollenfunktionsfähigkeit geben, somit also die mit der jeweiligen Umgebung verbundenen Anforderungen eine besondere Relevanz bei der Beantwortung der Items gespielt haben, ist möglich. So ist es vorstellbar, dass sich Patienten in der Ausübung alltäglicher Aktivitäten beziehungsweise von Freizeitbeschäftigungen in der Rehaklinik deutlicher beeinträchtigt sehen, als im eigenen häuslichen Umfeld, und das ‚unabhängig‘ vom gesundheitlichen Zustand. Diese Erklärung ist für die Skala Rollenfunktionsfähigkeit weniger auszuschließen, als für die Physische Funktionsfähigkeit. Hier ist es eher zweifelhaft anzunehmen, dass die Patienten aufgrund der räumlichen Gegebenheiten der Klinik mehr Schwierigkeiten erleben, beispielsweise kurze Strecken zu gehen beziehungsweise sich anzukleiden. Hinsichtlich dieser Items ist es naheliegender, die Bewertung des gesundheitlichen Zustandes und den damit verbundenen Beeinträchtigungen als maßgeblich für die Itembeantwortung zu beiden MZPen anzusehen.

Anders verhält es sich mit der non-uniformen Rekalibrierung auf der Skala Kognitive Funktionsfähigkeit. Hier weisen die Ergebnisse darauf hin, dass ein großer Teil der Stichprobe zum MZP 2 die eigene kognitive Funktionsfähigkeit durch den Einfluss spezifischer Faktoren auf den Indikator deutlich schlechter einschätzt als zum MZP 1. Die gleiche kognitive Funktionsfähigkeit wird zum MZP 2 negativer bewertet als zum MZP 1. Diese Umbewertung vergleichbarer Zustände ist im Verhältnis zum Einfluss der Veränderungen in der GLQ auf die kognitive Funktionsfähigkeit sogar so ausgeprägt, dass ein kleiner Effekt durch die Verbesserung der GLQ nach Cohen (1988) zu einem Nulleffekt auf der Beobachtungs-

ebene führt. Es findet durch die Auswertung auf der Ebene der beobachtbaren Effekte eine Unterschätzung der Veränderung statt. Inhaltlich sind auch hier verschiedene Sichtweisen vorstellbar. Zum Einen könnte angenommen werden, dass die Patienten dazu neigen, ihre durch die Erkrankung und Behandlung beeinflusste kognitive Leistungsfähigkeit zum MZP 1 positiver einzuschätzen, als diese tatsächlich ist, da sie mit *mehr* Beeinträchtigungen der kognitiven Leistungsfähigkeit gerechnet hatten und ihren tatsächlichen Zustand dadurch *überschätzen*. Diese Einschätzung wird jedoch zum MZP 2, mit mehr Abstand zur Akutbehandlung und der AHB wiederum relativiert. Als individueller Orientierungspunkt wird dann nicht mehr die erwartete Beeinträchtigung herangezogen, sondern die tatsächliche. Ein weiterer Grund kann auch hier ein möglicher Perspektivenwechsel sein. Die Einschätzung zum MZP 1 erfolgt möglicherweise vor dem Hintergrund der eigenen Erfahrungen mit der Erkrankung, der Behandlung und der Einschränkungen der kognitiven Leistungsfähigkeit während dieser Zeit. Diese wurden möglicherweise positiv bewertet, da die kognitive Funktionsfähigkeit zum MZP 1 wieder positiver erlebt wurde, da die Beeinträchtigungen während der Akutversorgung nachgelassen haben (individueller Vergleichsmaßstab). Während der AHB findet jedoch eine Orientierung an anderen Betroffenen statt (sozialer Vergleichsmaßstab). Diese Veränderung bleibt auch nach der AHB für die Bewertung maßgeblich und beeinflusst die Einschätzung eines (objektiv) vergleichbaren Zustands negativ.

Die inhaltlichen Ursachen für die statistisch identifizierten Rekalibrierungen können im Rahmen der vorliegenden Arbeit nicht spezifiziert und müssen somit durch zukünftige Untersuchungen geklärt werden. Wesentlich bleibt jedoch, dass alle genannten Gründe zu einer Veränderung des Bewertungshintergrunds einer Skala führen und somit den Effekt durch die Veränderung der GLQ verzerren können.

#### *Diskussion von Response-Shift im Zusammenhang mit der onkologischen Rehabilitation von PCa-Patienten*

Nach Lübbe (1998) kommt der onkologischen Rehabilitation aus verschiedenen Gründen eine Sonderrolle in der medizinischen Rehabilitation zu. Das abrupte Auftreten der Erkrankung, aber auch die langandauernde Beeinträchtigung, der ungewisse Verlauf einer Krebserkrankung, die Beeinträchtigung durch Therapiefolgen, die unsichere Prognose und die allgemeine psychische Beanspruchung der Patienten und deren Angehörigen sind nur einige Faktoren, die bei Krebspatienten in der Rehabilitation zu beachten sind. Dementsprechend steht

auch nicht die berufliche Wiedereingliederung im Vordergrund der Behandlung, sondern die Wiederherstellung des körperlichen und psychischen Wohlbefindens.

Ziele von Maßnahmen der onkologischen Rehabilitation umfassen verschiedene Bereiche. Die AWMF (2002) formuliert als wesentliche Ziele der onkologischen Rehabilitation die Veränderung funktionaler Aspekte, die Verbesserung der Lebensqualität und die soziale Integration. Alle drei Aspekte sind für die Einschätzung eines Rehabilitationserfolges mitentscheidend. Die Verbesserung der Lebensqualität ist eines der vorrangigen Ziele der onkologischen Rehabilitation (Haaf 2005, Koch 2000). Wie den Angaben über die Veränderung der GLQ als latentes Konstrukt zu entnehmen ist, wurde dieses Ziel auch bei der untersuchten Stichprobe erreicht: Die GLQ der untersuchten PCa-Patienten verbesserte sich statistisch signifikant vom MZP 1 zum MZP 2. Über objektive Veränderungen funktionaler Aspekte kann aufgrund der angewandten Datenerfassung keine Aussage getroffen werden, jedoch belegen die vorliegenden Ergebnisse, dass die verschiedenen subjektiv eingeschätzten Funktionsfähigkeiten überwiegend positiver bewertet werden.

Vor dem Hintergrund des Ziels der Verbesserung der GLQ im Rahmen der onkologischen Rehabilitation lassen sich die vorliegenden Ergebnisse folgendermaßen bewerten: Das Ziel der Veränderung beziehungsweise Verbesserung der GLQ im Rahmen der AHB bei PCa-Patienten lässt sich mit Hilfe der Response-Shift-Erfassung genauer abbilden. Die nach den herkömmlichen Methoden der Effektkalkulation beziehungsweise Veränderungsmessungen dargestellten Ergebnisse der onkologischen Rehabilitation bezüglich der Veränderung der GLQ sind aufgrund der vorliegenden Ergebnisse kritisch zu hinterfragen, denn ein Response-Shift kann zu einer Verzerrung im Sinne einer Überschätzung der Effekte bedingt durch die GLQ beitragen. Allerdings weist die Beeinflussung der Effekte durch einen Response-Shift in Form der Rekalibrierung ebenso darauf hin, dass ein Response-Shift nicht nur dazu führt, die Veränderungen aufgrund der GLQ zu überschätzen (wie auf den Skalen Physische und Rollenfunktionsfähigkeit), sondern dass ein tatsächlicher ‚wahrer‘ positiver Effekt auch überlagert und auf der Beobachtungsebene reduziert werden kann (Skala Kognitive Funktionsfähigkeit). Vor diesem Hintergrund ist für Evaluationsstudien zur onkologischen Rehabilitation bei PCa-Patienten im Speziellen, aber auch bei der onkologischen Rehabilitation im Allgemeinen nicht auszuschließen, dass möglicherweise auch die als eher kleine beziehungsweise Nulleffekte eingestuften Mittelwertsunterschiede durch einen Response-Shift ‚kleingehalten‘ sein könnten. Das Response-Shift nicht nur ein Phänomen in der onkologischen Rehabilitation ist,



zeigen die Studienergebnisse von Kohlmann und Raspe (1998) sowie Nübling und Kollegen (2004), die Response-Shift auch in anderen Indikationsbereichen der medizinischen Rehabilitation nachweisen konnten.

Die besondere Bedeutung eines Response-Shifts für die Evaluation von Maßnahmen im Rahmen der onkologischen Rehabilitation ist jedoch bereits aus der Definition ihrer Aufgabenfelder bei PCa-Patienten und den daraus abgeleiteten medizinisch-therapeutischen Maßnahmen ersichtlich. Die Neubewertung lebensqualitätsrelevanter Zusammenhänge wird durch eine Vielzahl einzelner Behandlungsmaßnahmen, aber auch die Gesamtausrichtung der Rehabilitation bei Krebserkrankungen in vielen Fällen angestrebt. Hierbei sind in erster Linie die psychosozialen Ziele und die darauf abzielenden Maßnahmen zu benennen (Lübbe 1998, Heim & Schwerte 2006). Es sei beispielhaft auf die psychologische Intervention und Psychotherapie als Behandlungsmaßnahmen in der onkologischen Rehabilitation verwiesen, die direkt zu einer Umbewertung von Zuständen beziehungsweise Perspektiven im Sinne einer ‚kognitiven Umstrukturierung‘ führen können oder im entsprechenden Einzelfall in der Regel auch sollen. Auch andere Maßnahmen können indirekt eine Perspektivenänderung bedingen, da beispielsweise bestimmte körperlich-organische Beeinträchtigungen nicht mehr vollständig ‚heilbar‘ sind und eine Auseinandersetzung mit bleibenden Einschränkungen und Beeinträchtigungen auch ohne psychologische Interventionen beziehungsweise Psychotherapie sinnvoll und realistisch ist. Es wird deutlich, dass vor allem in der onkologischen Rehabilitation Response-Shift nicht nur ein kritisch zu betrachtender Methodenaspekt, sondern auch ein potenzielles inhaltliches *Ziel* der rehabilitativen Behandlungen von Menschen mit Krebserkrankungen sein kann. Umso entscheidender ist die Differenzierung der beobachteten Effekte in einen Effektanteil aufgrund der wahren Veränderung und einen Anteil bedingt durch Response-Shift. Das ist insbesondere dann von Bedeutung, wenn der Response-Shift, wie in der vorliegenden Arbeit auf der Skala Rollenfunktionsfähigkeit, einen mittleren Effekt ( $ES = 0.55$ ) aufweist. Nach Norman und Kollegen (2003) ist bei diesem Response-Shift davon auszugehen, dass er klinisch bedeutsam, also nicht nur ein statistisch zu berücksichtigender Faktor ist. Aus diesem Grund sind weitere Studien zu response-shift-verursachenden Einflüssen durchzuführen. Die nach Oort (2004) definierten spezifischen Faktoren bieten dazu einen guten Ausgangspunkt.

Die spezifischen Einflussfaktoren auf die beobachtbaren Effekte auf Indikatorebene können in zukünftigen Modelltestungen und empirischen Untersuchungen als konkrete Variablen

integriert und geprüft werden. So ist es beispielsweise denkbar, den Einfluss von bestimmten Einzelmaßnahmen auf die beobachteten Effekte auf der Ebene der Indikatoren und auf der Ebene der GLQ dadurch zu untersuchen, dass Gruppenvergleiche kontrolliert und randomisiert durchgeführt werden. Ob eine Maßnahme einen von ihr erwarteten und nachweisbaren Einfluss auf die beobachteten Effekte beziehungsweise den zugrunde liegenden Anteilen dieser beobachteten Effekte besitzt, kann somit geprüft werden. Ein Therapieangebot zur Verbesserung der Lebensqualität in der onkologischen Rehabilitation sollte auf der Ebene des latenten Konstrukts GLQ und dessen Veränderung geprüft werden, eine Maßnahme zum Umgang mit Beeinträchtigungen der physischen Kapazitäten eher anhand der spezifischen Faktoren des entsprechenden Indikators *und* der GLQ. Darüber hinaus ist es erstrebenswert, verschiedene andere Einflussfaktoren während der onkologischen Rehabilitation (zur Förderung körperlicher Fitness, Schmerzbewältigung, Inkontinenz etc.), aber auch den Schweregrad der Erkrankung, die Art der medizinischen Behandlung (AHB, Heilverfahren) und Ähnliches in Modelle der Veränderung der GLQ beziehungsweise der Indikatoren aufzunehmen, um weitere Varianz der verschiedenen Indikatoren beziehungsweise der spezifischen Faktoren aufzuklären und gleichzeitig Response-Shift zu testen. Nach Hergert und Kollegen (2009) wurden bisher nur wenige Studien durchgeführt, die einzelinterventionsbezogene Evaluationen von Maßnahmen in der onkologischen Rehabilitation von PCa-Patienten untersuchen. Hier bietet der KFA-Ansatz eine gute Möglichkeit, die methodische Qualität solcher Studien in Zukunft auf ein höheres Niveau zu heben.

## **8.2 Einfluss der Veränderung der Lebensqualität auf Response-Shift – Fragestellung 2**

Im folgenden Unterkapitel werden die Ergebnisse der Fragestellung 2 zusammengefasst und diskutiert. Die Fragestellung lautet: Hat das Ausmaß der Veränderung der Globalen Lebensqualität einen Einfluss auf den Response-Shift und falls ja, in welcher Form?

### **8.2.1 Kurzbeschreibung der Ergebnisse der Fragestellung 2**

Im Response-Shift-Modell von Sprangers und Schwartz (1999) führt die Veränderung des Gesundheitszustands über ein Zusammenwirken aus Vorbedingungen und Mechanismen zu einem Response-Shift, der wiederum die Wahrnehmung der GLQ beeinflusst. Die Autorinnen beschreiben zudem einen Rückkopplungsprozess zwischen der wahrgenommenen Lebensqualität auf die Verarbeitung des Gesundheitszustands, so dass insgesamt von einem zirkulären Prozess auszugehen ist. Bisher ist mit der KFA-Methode nicht untersucht worden, ob das Ausmaß der Veränderung der wahrgenommenen Lebensqualität von Bedeutung ist für das

Auftreten eines Response-Shifts. In der Literatur gibt es nur eine Studie von Lepore und Eton (2000), in der sich die Veränderung der Lebensqualität mittels Hierarchischer Regressionsanalyse als einziger statistisch bedeutsamer Prädiktor für die Kriteriumsvariable Rekalibrierung, gemessen durch den Thentest, erwies.

Die Stichprobe wurde anhand des Medians der Veränderung der *Globalen Lebensqualität*, gemessen mit der gleich lautenden Skala des EORTC-QLQ-C30, in zwei Subgruppen unterteilt. Anschließend wurde für beide Gruppen eine eigene Response-Shift-Ableitung vorgenommen. Zusätzlich wurden für beide Gruppen die entsprechenden Response-Shift-Effekte kalkuliert. Auf diese Weise sollte untersucht werden, ob eine deutliche Verbesserung der wahrgenommenen Lebensqualität mit mehr Response-Shift-Prozessen beziehungsweise größeren Response-Shift-Effekten einhergehen als eine geringe Verbesserung der Lebensqualität.

Die Auswertungen ergaben, dass sich beide Gruppen hinsichtlich ihrer Response-Shift-Prozesse und dem Ausmaß der Response-Shift-Effekte zum Teil deutlich unterscheiden. Gemeinsam ist beiden Gruppen eine non-uniforme Rekalibrierung auf den Skalen Physische und Rollenfunktionsfähigkeit, die zu Interceptveränderungen auf den entsprechenden Skalen führt. Die weiteren Response-Shift-Prozesse unterscheiden sich jedoch zwischen den Gruppen. Zum Einen findet in der Gruppe mit den deutlichen Verbesserungen der Lebensqualität eine *Repriorisierung* auf der Skala Rollenfunktionsfähigkeit statt, was ein Hinweis dafür ist, dass sich auch die GLQ in dieser Gruppe verändert hat. Weiterhin treten in der Gruppe mit deutlichen Verbesserungen der globalen Lebensqualität non-uniforme Rekalibrierungen auf den Skalen Soziale und Emotionale Funktionsfähigkeit auf. In der Vergleichsgruppe findet kein solcher Response-Shift statt. In der Subgruppe der Personen mit geringen Verbesserungen der Lebensqualität ist zudem eine *uniforme Rekalibrierung* auf der Skala Kognitive Funktionsfähigkeit nachweisbar.

Für einen besonderen Einfluss der Veränderung der Globalen Lebensqualität auf einen Response-Shift spricht, dass in der Gruppe mit den deutlicheren Veränderungen auf insgesamt sieben Parametern ein Response-Shift stattfindet. In der Vergleichsgruppe sind fünf Parameter freizusetzen. Jedoch ist für die Gruppe mit deutlichen Veränderungen der Globalen Lebensqualität davon auszugehen, dass fünf Response-Shift-Prozesse stattgefunden haben (eine Repriorisierung, zwei non-uniforme Rekalibrierungen mit Auswirkungen auf die Intercepts und zwei non-uniforme Rekalibrierungen ohne Interceptbeeinflussung). In der Gruppe mit ge-

ringen Veränderungen der Globalen Lebensqualität ist von drei Response-Shift-Prozessen auszugehen (zwei non-uniforme Rekalibrierungen mit Interceptveränderung und eine uniforme Rekalibrierung). Ebenfalls für die Annahme des Einfluss der Veränderung der Globalen Lebensqualität auf einen Response-Shift spricht, dass der Response-Shift-Effekt auf der Skala Rollenfunktionsfähigkeit in der Gruppe mit den deutlicheren Lebensqualitätsveränderungen als groß zu bewerten ist, wohingegen der Response-Shift-Effekt in der Gruppe mit geringeren Lebensqualitätsveränderungen als klein bis mittel anzusehen ist. Gegen die Annahme spricht jedoch, dass der Response-Shift-Effekt auf der Skala Physische Funktionsfähigkeit in beiden Gruppen in etwa gleich ausfällt und auf der Skala Kognitive Funktionsfähigkeit lediglich in der Gruppe mit geringen Veränderungen der Lebensqualität ein Response-Shift-Effekt berechenbar ist.

### 8.2.2 Diskussion der Ergebnisse der Fragestellung 2

Die Subgruppenanalyse zur Variable ‚Veränderung der Globalen Lebensqualität‘ ergibt eine Repriorisierung aufgrund der Veränderung der Faktorladung der Skala Rollenfunktionsfähigkeit in der Gruppe mit deutlichen Veränderung der globalen Lebensqualität. Die Rollenfunktionsfähigkeit steht zum MZP 2 in einer engeren Beziehung zur GLQ als zum MZP 1. Aufgrund der relativ geringen Faktorladung des Indikators zum MZP 1 ( $\lambda=.23$ , Varianzaufklärung  $\sim 5\%$ ) ist zu diskutieren, ob es sich bei diesem Response-Shift um eine Neukonzeptualisierung handelt. Dafür sprechen die Höhe der Faktorladung zum MZP 1 und die damit einhergehende geringe Varianzaufklärung des Indikators. Im vorliegenden Fall wird jedoch nicht von einer Neukonzeptualisierung ausgegangen. Zwei Gründe sprechen für diese Entscheidung: Zum Einen ist eine Neukonzeptualisierung dann anzunehmen, wenn die Bedeutung des latenten Konstrukts für einen Indikator zu einem der MZPe *gleich Null* ist. Inhaltlich würde das bedeutet, dass *keine* Beziehung zwischen dem Indikator und dem latenten Konstrukt nachweisbar ist. Die ermittelte Faktorladung zum MZP 1 weist jedoch auf eine – zwar schwache – aber dennoch gegebene Beziehung zwischen der Rollenfunktionsfähigkeit und der GLQ zum MZP 1 hin. Zum Zweiten ist auf den geringen Stichprobenumfang und die Anfälligkeit der Berechnungen für Stichprobeneffekte zu verweisen. Aus diesem Grund wird hinsichtlich der Bewertung des Response-Shifts eher konservativ entschieden. Deshalb wird die Veränderung der Faktorladung des genannten Indikators zum MZP 2 als Repriorisierung gewertet.

Die Ursache für die Repriorisierung kann anhand der vorliegenden Daten nicht weiter untersucht werden. Weitgehend auszuschließen ist der Einfluss relevanter soziodemographischer Variablen, da diesbezüglich von Vergleichbarkeit zwischen den Gruppen auszugehen ist. Möglicherweise ist die Repriorisierung auf Unterschiede in der gesundheitlichen Beeinträchtigung der Gruppen zurückzuführen. Vor diesem Hintergrund wäre anzunehmen, dass die Bewertung der GLQ in erster Linie durch die Belastungen in basalen Aktivitäten, beispielsweise erfasst durch die Physische Funktionsfähigkeit, von größerer Bedeutung für die Patienten zum MZP 1 waren als Freizeit und Soziale Aktivitäten. Durch eine Besserung des Gesundheitszustandes zum MZP 2 könnte die Rollenfunktion eine gesteigerte Bedeutung erlangt haben. Statistisch betrachtet ist bei diesem Subgruppenvergleich auch die Tendenz der Regression zur Mitte zu berücksichtigen. Die von Rogosa und Willett (1985) vorgeschlagene Prüfung der Korrelation zwischen den Prä-Werten und Differenzwerten der Indikatoren erbrachte für beide Gruppen einen Hinweis auf eine Regression zur Mitte (siehe Anhang 4). Dieser Effekt ist in der Gruppe mit deutlichen Veränderungen der globalen Lebensqualität tendenziell stärker ausgeprägt. Jedoch ist der Unterschied zwischen den Regressionseffekten der Subgruppen auf der Skala Rollenfunktionsfähigkeit eher zu vernachlässigen. Somit ist zumindest zweifelhaft, ob der vorhandene Effekt der Regression zur Mitte für das Auftreten der Repriorisierung verantwortlich zu machen ist.

Eine weitere Erklärung könnte sein, dass die Personen der Gruppe mit Repriorisierung zum MZP 1 insgesamt psychisch stärker beeinträchtigt gewesen sind und dies die Beantwortung der Items beeinflusst hat. Zum MZP 2 hat sich die psychische Beeinträchtigung wiederum relativiert und die Rollenfunktionsfähigkeit entsprechend wieder die ihr ‚tatsächlich‘ zukommende Bedeutung für die GLQ der Personen bekommen. Einen Hinweis darauf lässt sich aus den Vergleichen der Mittelwerte der Indikatorvariablen zum MZP 1 sowie zum MZP 2 in beiden Gruppen ableiten. Die Beurteilung der Physischen Funktionsfähigkeit unterscheidet sich zum MZP 1 zwischen beiden Gruppen nicht statistisch signifikant, dafür jedoch die der Emotionalen, Sozialen, Kognitiven und Rollenfunktionsfähigkeit. Zum MZP 2 unterscheiden sich die Mittelwerte aller Indikatoren nicht statistisch signifikant, wobei die Mittelwertsunterschiede zwischen den Gruppen hinsichtlich der Physischen ( $p = .054$ ) und Rollenfunktionsfähigkeit ( $p = .052$ ) knapp das Kriterium der statistischen Signifikanz verfehlen. Sollte diese Annahme zutreffend sein, dann wäre die Frage von Bedeutung, auf welche weiteren Variablen sich die Unterschiede im psychischen Erleben beziehungsweise der psychischen Verarbeitung zurückführen lassen. Denkbar wären Einflussfaktoren wie depressive Bewältigungs-

strategien, das Ausmaß sozialer Unterstützung oder andere Einflussfaktoren auf interne Bewältigungsprozesse.

Ein weiterer Grund dafür kann in deutlicheren Einschränkungen der Funktionsfähigkeiten der Gruppe mit der ausgeprägteren Veränderung der Globalen Lebensqualität zum MZP 1 gesehen werden. Die beiden Subgruppen unterscheiden sich zum MZP 1 auf allen Indikatorvariablen mit Ausnahme der Physischen Funktionsfähigkeit. Im Sinne des Response-Shift-Modells ist denkbar, dass die Gruppe mit deutlichen Veränderungen der Globalen Lebensqualität zum MZP 1 auch tatsächlich körperlich stärker beeinträchtigt beziehungsweise schwerer erkrankt gewesen ist und aufgrund der gravierenderen Veränderung durch die Erkrankung mehr Response-Shift-Prozesse sowie einen größeren Response-Shift-Effekt entwickelt hat. Diese Erklärung ist auch nach dem Modell von Sprangers und Schwartz (1999) denkbar, die in dem Ausmaß des Ereignisses (Katalysators) eine entscheidende Einflussgröße auf einen Response-Shift sehen. Allerdings kann diese Erklärung nicht mit entsprechenden objektiven Daten unterlegt werden, so dass darin in erster Linie eine Anregung für weitere Forschungen zum Thema zu sehen ist. Gegen diese Annahme spricht jedoch auch das Ergebnis des Vergleichs der Response-Shift-Effekte auf den Skalen Physische und Kognitive Funktionsfähigkeit.

Die reine uniforme Rekalibrierung in der Subgruppe der Personen mit geringen Veränderungen der Lebensqualität auf der Skala Kognitive Funktionsfähigkeit bedeutet nach dem LLVM nach Oort (2004), dass in dieser Subgruppe die Personen ihre Bewertungen der kognitiven Funktionsfähigkeit zum MZP 2 über alle Stufungen hinweg einheitlich verändert haben. Die Werte weisen darauf hin, dass die Personen zum MZP 2 ihre kognitive Funktionsfähigkeit negativer einschätzen als zum MZP 1. Inhaltlich können hier die selben potenziellen Ursachen angeführt werden wie zur Diskussion der Ergebnisse zum Response-Shift bei der Kognitiven Funktionsfähigkeit in der Gesamtgruppe. Zum Einen könnte eine subjektive Überschätzung der Kognitiven Funktionsfähigkeit zum MZP 1 vorliegen, weil die eigene Leistungsfähigkeit aufgrund der Behandlung über der erwarteten liegt und somit das tatsächliche Ausmaß der Funktionsfähigkeit eher überschätzt wird. Zum Anderen könnte es sein, dass durch die AHB und die Zeit danach die Bewertung der Kognitiven Funktionsfähigkeit vor allem durch einen sozialen Vergleich mit anderen relativiert wird. Persönlich geht es zum MZP 1 also besser als erwartet (individueller Vergleichsmaßstab). Zum MZP 2 ist die Kogni-

tive Funktionsfähigkeit jedoch nicht besser geworden. Lediglich die Bewertung der Funktionsfähigkeit hat sich an einem allgemeinen Niveau orientiert (sozialer Vergleichsmaßstab).

Des Weiteren soll Bezug genommen werden auf das Ergebnis, dass in *beiden* Gruppen Response-Shift-Prozesse und zum Teil auch vergleichbare Response-Shift-Effekte aufgetreten sind. Hier ist wiederum darauf hinzuweisen, dass in der onkologischen Rehabilitation die Verbesserung der Lebensqualität durch verschiedene Behandlungsmaßnahmen wie Psychotherapie, Psychosoziale Beratung etc. als Zielkriterium definiert wird. Dass in beiden Gruppen eine Veränderung der Bewertungshintergründe festzustellen ist, ist möglicherweise darauf zurückzuführen, dass sich die Patienten in beiden Gruppen mit ihren Beeinträchtigungen aufgrund des PCa auseinandersetzen und ihre eigene Lebenssituation vor dem Hintergrund dieser Erkrankung umbewerten. In beiden Gruppen kann dies eine Folge der psychosozialen Belastungen sowie der darauf ausgerichteten Angebote im Rahmen der onkologischen Rehabilitation sein, die neben den sozialen Vergleichen oder den veränderten subjektiven Vergleichswerten aufgrund der Erfahrung mit dem PCa Einfluss auf einen Response-Shift haben können.

Wie bereits in der Diskussion der Hauptfragestellung beschrieben, ist der Response-Shift für beide Gruppen auch durch die Maßnahmen der onkologischen Rehabilitation angestrebt, wenn die Beeinträchtigungen durch die Erkrankung deutlich sind. Beide Subgruppen weisen zum MZP 1 Werte in den Indikatoren auf, die im Vergleich zu den Werten der gesunden Normstichprobe als unterdurchschnittlich angesehen werden können (Schwarz & Hinz 2001). Die Beeinträchtigungen sind zum MZP 1 demnach in *beiden* Gruppen gegeben. Wenn weiterhin davon ausgegangen wird, dass sich der objektive gesundheitliche Zustand der Patienten *nicht* beziehungsweise *kaum verändert*, dann ist aus Response-Shift-Perspektive davon auszugehen, dass dieses Phänomen stattfinden kann, um die bleibenden Beeinträchtigungen neu zu bewerten und den eigenen Bewertungshintergrund vor dem Hintergrund der fortbestehenden Beeinträchtigungen neu zu adjustieren. Dieser Prozess wäre vor allem für die Gruppe der Personen mit geringeren Veränderungen der Globalen Lebensqualität zu erwarten, wenn die subjektive Bewertung der globalen Lebensqualität als Hinweis auf den tatsächlichen objektiven Gesundheitszustand der Patienten gewertet wird. Aus diesen Gründen ist nachvollziehbar, dass in beiden Gruppen Response-Shift-Prozesse und -Effekte nachweisbar sind.

### **8.3 Einfluss von Benefit Finding, Selbstwirksamkeit, positiver und negativer Affektivität auf Response-Shift – Fragestellung 3**

Im Folgenden sollen die wesentlichen Ergebnisse der Fragestellung 3 zusammengefasst und interpretiert werden. Dabei wird für jede untersuchte Moderatorvariable eine separate Darstellung präsentiert. Es werden Unterschiede in den Angaben zur Anzahl von freigesetzten Response-Shift-Parametern und Response-Shift-Prozessen beschrieben. Nach der Diskussion der Ergebnisse der Fragestellung 3 wird auf methodische Kritikpunkte der Subgruppenanalysen eingegangen, die auch für die Auswertungen der Fragestellung 2 gelten.

#### **8.3.1 Benefit Finding**

Als erste Moderatorvariable wurde der Einfluss des *Benefit Finding* auf einen potenziellen Response-Shift untersucht. Benefit Finding bezeichnet nach Affleck und Tennen (1996) einen Prozess, aufgrund dessen widrige Umstände und kritische Situationen umbewertet und als Möglichkeiten für positive persönliche Entwicklungen angesehen werden. Benefit Finding geht unter anderem einher mit dem Setzen neuer relevanter Ziele und Werte, einer Zunahme an persönlicher Stärke und Kraft sowie der Wahrnehmung neuer Entwicklungsmöglichkeiten (Tedeschi et al. 1998). Somit stellt Benefit Finding eine mögliche inhaltliche Erklärung für einen Response-Shift dar. In der vorliegenden Arbeit wurde Benefit Finding als Operationalisierung eines kognitiven Bewältigungsprozesses angesehen, der als Mechanismus-Variable im Response-Shift-Modell von Sprangers und Schwartz (1999) einen Einfluss auf einen Response-Shift haben kann.

Die gruppenspezifischen Auswertungen in Abhängigkeit vom Ausmaß des Benefit Finding (hoch vs. gering) ergaben, dass in der Gruppe mit höherer Ausprägung an Benefit Finding mehr Response-Shift-Prozesse stattfinden. In der Gruppe mit höherem Ausmaß an Benefit Finding ist davon auszugehen, dass insgesamt fünf Response-Shift-Prozesse stattgefunden haben. In der Vergleichsgruppe sind es drei. In der Gruppe mit höherem Benefit Finding sind sieben Response-Shift-Parameter freizusetzen, in der Gruppe mit geringerem Benefit Finding fünf.

Die Intercepts der Skalen Physische und Rollenfunktionsfähigkeiten waren in beiden Gruppen freizusetzen. Es handelt sich bei den zugrundeliegenden Response-Shift-Prozessen jedoch um eine non-uniforme Rekalibrierung, da in beiden Gruppen auf den entsprechenden



Skalen auch die Messfehlervarianzen freizusetzen waren. Die Intercepts der Skala Kognitive Funktionsfähigkeit war nur in der Gruppe mit deutlichem Benefit Finding freizusetzen. Da die Messfehlervarianz der Skala in dieser Gruppe restringiert blieb, ist davon auszugehen, dass es sich hierbei um eine uniforme Rekalibrierung handelt.

Außerdem waren in der Gruppe mit deutlichem Benefit Finding die Restriktionen der Messfehlervarianzen der Skalen Emotionale und Soziale Funktionsfähigkeit nicht aufrechtzuerhalten. In der Vergleichsgruppe mit geringerem Benefit Finding war die Messfehlervarianz der Skala Kognitive Funktionsfähigkeit freizusetzen. Diese non-uniformen Rekalibrierungen hatten jedoch in beiden Gruppen keine Auswirkungen auf die Intercepts und somit auch nicht auf die Effektkalkulation.

Bezüglich der Kalkulation der Effekte zeigte sich, dass in beiden Gruppen auf den Skalen Physische und Rollenfunktionsfähigkeit vergleichbare Response-Shift-Effekte nachgewiesen werden konnten. Außerdem fand auf der Skala Kognitive Funktionsfähigkeit in der Gruppe mit deutlicherem Benefit-Finding ein ‚kleiner‘ Response-Shift statt, der einen Effekt aufgrund der Veränderung der GLQ relativiert, so dass der beobachtete Effekt als unbedeutsam zu bezeichnen ist.

Wesentlich für die Bewertung der aufgedeckten Response-Shift-Prozesse und -effekte hinsichtlich der Variable Benefit Finding ist, dass es sich bei der Einteilung der Gruppen nicht um zwei Gruppen mit Extremwerten handelt, sondern dass aufgrund der relativ kleinen Gesamtstichprobe die Gruppe der PCa-Patienten mit Angaben zum Benefit Finding anhand des Medians der Variable in zwei Gruppen geteilt wurde. Die Werte der beiden Gruppen unterscheiden sich zwar deutlich und statistisch signifikant. Es handelt es sich dabei inhaltlich jedoch eher um die Ausprägungen ‚mittelmäßig‘ für die Gruppe mit *geringerem* Benefit Finding und ‚ziemlich‘ für die Gruppe mit *deutlicherem* Benefit Finding. In beiden Gruppen findet somit Benefit Finding in nicht geringem Maße statt, so dass aufgrund des Ergebnisses zu erwarten ist, dass auch in beiden Gruppen Response-Shift stattfindet. So läßt sich die Vergleichbarkeit der Größe der Response-Shift-Effekte in beiden Gruppen erklären. Dennoch weisen die Anzahl der Freisetzungen von Response-Shift-Parametern in der vorliegenden Stichprobe darauf hin, dass Benefit Finding möglicherweise einen Einfluss auf einen Response-Shift hat.

Die Auswertungen der Mittelwertsvergleiche der Indikatorvariablen in beiden Subgruppen zu beiden MZPen ergaben, dass lediglich auf der Skala Soziale Funktionsfähigkeit zum MZP 1 ein statistisch signifikanter Unterschied zwischen den Gruppen besteht. Anhand der multiplen Testung ohne Alpha-Fehler-Adjustierung ist auch dieser Unterschied eher zu vernachlässigen. Dieses Ergebnis ist ein möglicher Hinweis darauf, dass sich beide Gruppen bezüglich der selbsteingeschätzten Funktionsfähigkeiten ähneln und die Unterschiede in der Anzahl der Response-Shift-Prozesse weniger auf Aspekte der Funktionsfähigkeiten, sondern tatsächlich auf Auswirkungen des Benefit Finding zurückführen lassen. Bezüglich soziodemographischer Variablen sind beide Subgruppen vergleichbar. Der objektive Gesundheitszustand kann als Einflussfaktor nicht ausgeschlossen werden.

### 8.3.2 Selbstwirksamkeit

Als weitere Moderatorvariable wurde *Selbstwirksamkeit* untersucht, die nach dem Modell von Sprangers und Schwartz (1999) ebenfalls als Mechanismus-Variable eingestuft wurde. Selbstwirksamkeit wird definiert als Erwartung, schwierige, neue oder anforderungsreiche Aufgaben in kompetenter Weise kontrollieren und bewältigen zu können (Bandura 1997). Studien zur Bedeutung von Selbstwirksamkeit bei Patienten mit Krebserkrankungen beschreiben diese als protektiven Faktor in der Auseinandersetzung mit Krebserkrankungen. (Böhmer et al. 2007, Cunnigham et al. 1991, Lev 1997, Weber et al. 2004). Mit der vorliegenden Arbeit sollte untersucht werden, ob ein höheres Ausmaß an Selbstwirksamkeit mit mehr Response-Shift-Prozessen sowie größeren Response-Shift-Effekten einhergeht als ein geringes Ausmaß.

In der Gruppe mit einem höheren Ausmaß an Selbstwirksamkeit sind mehr Response-Shift-Prozesse nachweisbar als in der Vergleichsgruppe. In der Gruppe der Patienten mit geringerer Selbstwirksamkeitserwartung sind vier Response-Shift-Parameter freizusetzen. In der Vergleichsgruppe sind es sieben. Auswirkungen auf die Effektkalkulation haben in beiden Gruppen jedoch jeweils nur zwei Response-Shift-Prozesse. In der Gruppe mit geringerer Selbstwirksamkeitserwartung verursacht eine non-uniforme Rekalibrierung auf den Skalen Physische und Rollenfunktionsfähigkeit eine Beeinflussung der Intercepts, so dass hier Effekte nachweisbar sind. In der Gruppe mit höherer Selbstwirksamkeitserwartung finden ebenfalls zwei Rekalibrierungsprozesse statt, wobei einer durch eine non-uniforme Rekalibrierung verursacht wird (Skala Rollenfunktionsfähigkeit) und der zweite durch eine uniforme Rekalibrierung (Skala Kognitive Funktionsfähigkeit). Für beide Skalen lassen sich entsprechend

Response-Shift-Effekte kalkulieren, wobei der direkte Vergleich der Response-Shift-Effekte auf der Skala Rollenfunktionsfähigkeit nicht auf Unterschiede zwischen den Gruppen mit hoher und geringerer Ausprägung an Selbstwirksamkeit schließen lässt.

Wie bereits beim Benefit Finding beschrieben, handelt es sich auch bei den Mittelwerten der beiden Subgruppen nicht um Ausprägungen im Bereich der beiden Skalenendpunkte, sondern um Werte, die sich ‚lediglich‘ statistisch signifikant voneinander unterscheiden. In beiden Gruppen wird ein Ausmaß an Selbstwirksamkeit angegeben, dass sich für die Gruppe der PCa-Patienten mit geringeren Werten um den Skalenwert „stimmt eher“ bewegt. Für die Gruppe mit deutlicher Ausprägung der Selbstwirksamkeitserwartung liegt die mittlere Einschätzung bei dem Skalenwert „trifft genau zu“. In beiden Gruppen wird demnach ein positives Ausmaß an Selbstwirksamkeitserwartung angegeben. Die Ausprägungen innerhalb der beiden Gruppen lassen es wiederum erwarten, dass in beiden Gruppen Response-Shift-Prozesse nachweisbar sind.

Fraglich ist jedoch bezüglich der vorliegenden Auswertungen zur Selbstwirksamkeitserwartung, warum die Response-Shift Effekte zum Teil auf unterschiedlichen Skalen auftreten. Eine Erklärung dafür könnte darin zu sehen sein, dass in der Gruppe mit geringerer Selbstwirksamkeitserwartung die gesundheitlichen Beeinträchtigungen zum MZP 2 stärker ausgeprägt sind als in der Vergleichsgruppe, so dass in dieser Gruppe ein Response-Shift eine zusätzliche stabilisierende Wirkung haben könnte. Aufgrund der fehlenden Angaben zur objektiven Gesundheit der Personen, muss die Beantwortung dieser Frage offen bleiben. Ein möglicher Hinweis darauf ist jedoch in den Ergebnissen der Mittelwertvergleiche der Indikatoren zwischen den beiden Subgruppen zu sehen. Zum MZP 1 unterscheiden sich beide Gruppen tendenziell, aber nicht statistisch signifikant voneinander. Zum MZP 2 sind die Unterschiede jedoch auf allen Indikatoren statistisch signifikant. Die Gruppe mit höheren Ausprägungen der Selbstwirksamkeit weist auf allen Mittelwerten zum MZP 2 höhere Werte auf den Indikatoren, das heißt auch eine bessere GLQ, auf.

### 8.3.3 Positive Affektivität

Sprangers und Schwartz (1999) nehmen an, dass positive Affektivität response-shift-förderlich wirkt. Dabei wird diese im Modell der Autorinnen als eine Vorbedingung bezeichnet. In der vorliegenden Arbeit wird davon ausgegangen, dass ein hohes Ausmaß an positiver Affektivität einen Response-Shift fördert, ein geringes Ausmaß eher hemmend auf einen Res-

ponse-Shift wirkt. Ein hohes Ausmaß an positiver Affektivität sollte demnach mit mehr Response-Shift-Prozessen sowie größeren Response-Shift-Effekten einhergehen als eine geringe Ausprägung.

Die wesentlichen Ergebnisse der Subgruppenanalyse sind folgende: In der Gruppe mit geringer positiver Affektivität sind insgesamt vier Response-Shift-Parameter freizusetzen. Diese betreffen zwei Intercepts sowie zwei Messfehler von insgesamt zwei Indikatoren. Somit ist in der Gruppe mit geringer positiver Affektivität in beiden Fällen von einer non-uniformen Rekalibrierung auszugehen. In der Vergleichsgruppe mit deutlicher Ausprägung der positiven Affektivität sind insgesamt sechs Freisetzungen von Response-Shift-Parametern vorzunehmen. Hier sind die Restriktionen bei zwei Intercepts und vier Messfehlern nicht aufrechtzuerhalten. Insgesamt zeigen sich in dieser Gruppe fünf Response-Shift-Prozesse, von denen sich vier auf eine non-uniforme Rekalibrierung zurückführen lassen, wobei sich davon lediglich eine auf die Interceptebene auswirkt (Rollenfunktionsfähigkeit), und eine auf eine uniforme Rekalibrierung (Kognitive Funktionsfähigkeit). Auf der Ebene der Berechnung der Effekte lässt sich ein direkter Vergleich der Response-Shift-Effekte zwischen den beiden Gruppen auf der Skala Rollenfunktionsfähigkeit ziehen: In der Gruppe mit geringerer positiver Affektivität findet ein um einen kleinen Effekt größerer Response-Shift statt. Die anderen beiden Response-Shift-Effekte finden auf verschiedenen Skalen statt und lassen keinen direkten Vergleich zu.

Die Auswertung zur Anzahl der Response-Shift-Prozesse und der Höhe der Effekte sind widersprüchlich. Der Vergleich der Anzahl der Response-Shift-Prozesse spricht für, das Ausmaß der Effektkalkulation gegen die Annahme, dass positive Affektivität einen Einfluss auf einen Response-Shift hat. Dass in beiden Gruppen Response-Shift-Prozesse identifizierbar sind, ist ähnlich zu erklären wie bei den vorangegangenen Moderatorvariablen. Die Mittelwerte der jeweiligen Subgruppen liegen für die Gruppe mit geringer Ausprägung der positiven Affektivität bei 2,69, für die Gruppe mit hoher Ausprägung bei 3,79. Das entspricht für die erste Gruppe einer Skalenabstufung von ‚einigermaßen‘, für die zweite einem Wert von ‚deutlich‘. In beiden Gruppen ist somit ein *relativ hohes* Ausmaß an positiver Affektivität vorhanden. Hinsichtlich des mittleren Response-Shift-Effekts in der Gruppe mit geringer ausgeprägter positiver Affektivität ist möglicherweise anzunehmen, dass hier wiederum die stärkere physische und psychische Beeinträchtigung dazu beiträgt, einen möglicherweise schlechteren Gesundheitszustand durch einen Response-Shift in einer akzeptableren Weise

annehmen zu können. Möglicherweise liegt diesbezüglich ein Interaktionseffekt zwischen Gesundheitszustand und positiver Affektivität vor. Aufgrund der stärkeren Beeinträchtigung der Funktionsfähigkeiten der Gruppe mit geringerer positiver Affektivität kann der ausgeprägtere Response-Shift auf der Skala Rollenfunktionsfähigkeit möglicherweise dadurch erklärt werden, dass bei PCa-Patienten dieser Gruppe mehr intrusive Gedanken stattfinden, die als Zeichen für eine intensivere kognitive Auseinandersetzung mit der Gesundheits- beziehungsweise Lebenssituation gewertet werden könnten. Diese Annahme spräche dafür, dass der Response-Shift-Effekt in der Gruppe mit geringerer Ausprägung der positiven Affektivität größer ist als in der Vergleichsgruppe.

#### 8.3.4 Negative Affektivität

Die Auswirkungen von negativer Affektivität auf einen Response-Shift wurden bisher ebenfalls nicht empirisch untersucht. Sprangers und Schwartz (1999) gehen davon aus, dass Depressivität einen negativen Zusammenhang mit Response-Shift aufweist. In der vorliegenden Arbeit wurde von der Annahme ausgegangen, dass ein geringeres Ausmaß an negativer Affektivität response-shift-förderlich ist.

Hinsichtlich der Anzahl freizusetzender Response-Shift-Parameter unterscheiden sich die beiden Subgruppen insofern, dass in der Gruppe mit geringer negativer Affektivität sechs, in der Vergleichsgruppe sieben Parameter freizusetzen sind. In der Gruppe mit geringer Ausprägung der negativen Affektivität handelt es sich jedoch nur um fünf Response-Shift-Prozesse, da die Interceptfreisetzung der Skala Rollenfunktionsfähigkeit auf die non-uniforme Rekalibrierung der Messfehlervarianz der Skala zurückzuführen ist. In der Vergleichsgruppe finden ebenfalls fünf Response-Shift-Prozesse statt, da die Interceptfreisetzungen der Skalen Physische und Rollenfunktionsfähigkeit auf eine non-uniforme Rekalibrierung zurück zu führen sind. Somit weist dieses Ergebnis nicht darauf hin, dass ein geringes Ausmaß an negativer Affektivität response-shift-förderlich ist.

In der Gruppe mit ausgeprägter negativer Affektivität sind insgesamt drei Intercepts betroffen, wobei sich dies auf die Kalkulation von Effekten auswirkt. Auf der Skala Rollenfunktionsfähigkeit findet in beiden Gruppen ein Response-Shift statt, wobei der in der Gruppe mit hoher Ausprägung der negativen Affektivität um ca. einen kleinen Effekt größer ausfällt. Diese Ergebnisse sprechen ebenfalls gegen die Annahme, dass ein geringeres Ausmaß an negativer Affektivität mit mehr Response-Shift einhergeht.

Die Physische Funktionsfähigkeit unterscheidet sich zum MZP 2 nicht zwischen den beiden Subgruppen, wobei auf allen anderen Indikatoren statistisch signifikante Unterschiede zwischen den Gruppen bestehen. Die Gruppe mit geringer negativer Affektivität weist dabei höhere Werte auf den Funktionsskalen auf. Vor dem Hintergrund der Symptom-Perception-Hypothese von Watson und Pennebaker (1989) führt die ausgeprägtere negative Affektivität möglicherweise zu einer kritischeren Bewertung der Funktionsfähigkeiten. Dennoch ist darauf hinzuweisen, dass das Ausmaß der negativen Affektivität selbst in der Gruppe mit hoher Ausprägung der Skalenstufe ‚ein bisschen‘ entspricht, demnach inhaltlich gesehen noch immer ein eher *geringes* Ausmaß vorliegt. Möglicherweise ist diese mäßige Ausprägung der negativen Affektivität bei den Personen dieser Gruppe – vorausgesetzt die Angaben zur Physischen Funktionsfähigkeit sind für beide Gruppen ein Hinweis auf eine vergleichbare objektive physische Beeinträchtigung – ein Niveau, das die betreffenden Personen zu mehr ‚Neuorientierung‘ motiviert. Die Gruppe mit geringer negativer Affektivität ist in etwa mit der Skalenstufe ‚gar nicht‘ gleichzusetzen. Dies kann als gelassene und ruhige Art der Auseinandersetzung mit der kritischen Lebenssituation gewertet werden. Im vorliegenden Fall scheint diese Haltung damit einherzugehen, den Bewertungshintergrund für die GLQ nur geringfügig zu verändern.

Die deutlich geringere Ausprägung der negativen Affektivität in der entsprechenden Gruppe lässt sich auch nicht eindeutig als *Ergebnis* eines Response-Shifts bewerten, da in diesem Fall Response-Shift in dieser Gruppe – bei vergleichbarer Physischer Funktionsfähigkeit zum MZP 2 – ausgeprägter hätte stattfinden müssen. Andererseits kann in der beinahe ausschließlichen Verneinung negativer Affekte auch die Tendenz zur Vermeidung der Auseinandersetzung mit einer kritischen Lebenssituationen gesehen werden, die wiederum einen Response-Shift verhindert. Wichtig ist jedoch auch die Frage, ob es sich bei diesen Einschätzungen tatsächlich um Persönlichkeitsmerkmale oder um bewältigungsabhängige Beurteilungen handelt. Krohne und Kollegen (1996) untersuchten die Zusammenhänge zwischen den Beurteilungen der Affektivität je nach unterschiedlichem Bewertungszeitraum (‚aktuell‘, ‚in den letzten Tagen‘, bis hin zu ‚Allgemein‘). Die Autoren kommen zu dem Ergebnis, dass die Einschätzung der ‚Allgemeinen Affektivität‘ hoch korreliert mit der Beurteilung der Affektivität ‚in den letzten Tagen‘. In Bezug auf die vorliegenden Ergebnisse könnte dies bedeuten, dass die Beurteilungen der Affektivität weniger durch Persönlichkeitsmerkmale bedingt sind, als durch die Auseinandersetzung mit der Erkrankung. Dies würde eine Erklärung dafür sein, warum die ausgeprägtere negative Affektivität zu mehr und höheren Response-Shift-Effekten

führt. Das immer noch geringe Ausmaß an negativer Affektivität in der Gruppe mit geringer Ausprägung dieser Variable könnte unter der Bewältigungsperspektive dennoch mit einem höheren Ausmaß an intrusiven Gedanken einhergehen. Allerdings bedeutet diese Erklärung auch, dass negative Affektivität als Persönlichkeitseigenschaft zumindest im ermittelten Ausprägungsbereich der Skala eher geringen Einfluss auf einen Response-Shift hat.

Insgesamt ist aufgrund der vorliegenden Ergebnisse der Einfluss der Moderatorvariablen weiter zu untersuchen. Die Anzahl der Response-Shift-Prozesse beziehungsweise freizusetzender Response-Shift-Parameter bestätigt in drei von vier Fällen die formulierten Annahmen. Das Ausmaß der Response-Shift-Effekte in den jeweiligen Vergleichsgruppen spricht jedoch überwiegend gegen die Bedeutsamkeit dieser Moderatorvariablen in der vorliegenden Stichprobe.

#### **8.4 Methodische Einschränkungen der Fragestellungen 2 und 3**

Hinsichtlich der zukünftigen Untersuchung von Moderatorvariablen auf einen Response-Shift mittels KFA ist eine genügend große Stichprobe in die Analysen aufzunehmen, um dann anhand von Extremwerten innerhalb der Moderatorvariablen stärker zu polarisieren. Mit inhaltlich größeren Unterschieden zwischen den Subgruppen werden Aussagen über den Einfluss auf einen Response-Shift klarer. Die vorliegende Untersuchung bietet Hinweise darauf, dass Moderatorvariablen für einen Response-Shift von Bedeutung sein können. Bei genügend großen Substichproben mit Werten im 1. und 4. Quartil sollten sich Unterschiede im Response-Shift deutlicher darstellen. Es ist darauf hinzuweisen, dass Personen im 1. Quartil der Summenwerte ein gewisses Ausmaß der untersuchten Variable zeigen werden, so dass auch in einem solchen Fall mit dem Auftreten von Response-Shift zu rechnen ist. Jedoch sollte das Ausmaß des Response-Shift in beiden Quartilen wesentlich deutlicher voneinander abweichen als in der vorliegenden Auswertung.

Grundsätzlich ist bei der Einteilung der Gesamtstichprobe von 212 PCa-Patienten in Subgruppen davon auszugehen, dass die Anzahl der Personen pro Gruppe für die Verwendung der KFA kritisch einzuschätzen ist. Es handelt sich um Gruppengrößen, die den Vorgaben der allgemeinen Richtlinien zur Durchführung von KFAen nicht entsprechen (Kline 2005). Grundsätzlich ist eine Gruppengröße mit  $n > 200$  eine wesentliche Richtlinie. Aufgrund des explorativen Charakters der Fragestellungen wird dieses Kriterium jedoch ver-

nachlässigt, da aus Gründen des Umfanges der Gesamtstichprobe nur auf diese Gruppenumfänge zurückgegriffen werden kann.

Dennoch sprechen sich Marsh und Kollegen (1996) auch für die Verwendung der KFA bei Stichproben mit geringem  $n$  aus, da sie mittels Simulationsstudien zu dem Schluss kamen, dass ein Stichprobenumfang von  $n = 50$  mit steigendem Verhältnis Indikatoren/Faktor und steigendem Freiheitsgraden ausreichend sein kann, um gute Schätzungen zu liefern, wobei eine Relation von Indikator pro Faktor von 3 zu 1 nicht unterschritten werden darf. In der vorliegenden Arbeit sprechen drei Gründe für die Verwendung der Subgruppen mit einem  $n$  von  $\leq 106$  pro Gruppe: 1. Die Modellpassung wurde an zwei Stichproben mit 212 beziehungsweise 185 Personen überprüft und als akzeptabel bis gut bewertet; 2. Das Verhältnis zwischen Indikator und Faktor beträgt 5 zu 1; 3. Die Werte zur Passung der Modelle in den verschiedenen Subgruppen können insgesamt als akzeptabel angesehen werden. Das gilt für das Basismodell ebenso wie für das Response-Shift-Modell. Dennoch werden die Ergebnisse vor dem Hintergrund der Anfälligkeit von Varianz/Kovarianzmatrizen aufgrund von Stichprobenverzerrungen kritisch bewertet. Die Ausgangsbasis wird jedoch für die Bewertung unterschiedlicher Response-Shift-Prozesse und deren Effekte als akzeptabel angenommen.

Die Ergebnisse lassen sich zwar als Hinweise für oder gegen die entsprechenden Annahmen verwerten, sollen aber nicht im klassischen Sinne hypothesentestend sein. Zum Einen schränkt der geringe Stichprobenumfang den Interpretationsspielraum der Ergebnisse stark ein. Zum Anderen wird auf eine Differenzierung der Alpha-Fehler-Wahrscheinlichkeit aufgrund der multiplen Subgruppenanalysen verzichtet (Fayers & King 2009). Wesentlich für die vorliegende Arbeit ist der explorative Charakter dieser Untersuchung.

## **8.5 Kritische Betrachtung allgemeiner methodischer Aspekte der Studie**

Mit der vorliegenden Arbeit wird Response-Shift im Zusammenhang mit der Erfassung der GLQ bei PCa-Patienten im Rahmen der AHB untersucht. Abschließend sollen verschiedene methodische Aspekte kritisch betrachtet werden. Dazu zählen die Durchführung der Datenerhebung, die Stichprobengewinnung, die Modellableitung und das Messmodell der GLQ.

Die Datenerhebung zur Beantwortung der vorliegenden Fragestellungen fand im Rahmen der Regelversorgung der Patientinnen und Patienten der *Klinik Bad Oexen – Klinik für onko-*



*logische Rehabilitation und Anschlussheilbehandlung* statt. Im Jahr 2005 wurde unter den Patientinnen und Patienten eine Prä-Post-Messung zur Erfassung der GLQ mit dem EORTC-QLQ-C30 durchgeführt, das als Standardinstrument zur Lebensqualitätserfassung in der Klinik eingesetzt wurde. Im Rahmen dieser Erhebung ergab sich für den Autor der vorliegenden Arbeit die Möglichkeit, theoriegeleitet zusätzliche Fragebögen einzusetzen, um response-shift-relevante Fragestellungen zu bearbeiten. Die Fragestellungen der vorliegenden Arbeit wurden anhand der Annahmen aus dem Response-Shift-Modell nach Sprangers und Schwartz (1999) sowie den publizierten Untersuchungen zur methodischen Erfassung von Response-Shift abgeleitet. Um jedoch eine möglichst große Stichprobe zu verwenden, wurden die neu aufgenommenen Fragebögen zur Beantwortung der Nebenfragestellungen in die laufende Datenerhebung integriert. Somit fand die Erfassung der Variablen Selbstwirksamkeit, Benefit Finding sowie positive und negative Affektivität erst zum MZP 2 statt. Aus diesem Grund können diese Zusatzvariablen nicht zu einer differenzierten Drop-out-Analyse verwendet werden, da zu den Non-Respondern zum zweiten MZP keine entsprechenden Angaben vorliegen. Von einer Beeinflussung der Angaben im EORTC-QLQ-C30 durch die teilnehmenden Rehabilitanden aufgrund der zusätzlich aufgenommenen Fragebögen ist bei denjenigen, die auch zum MZP 2 die Unterlagen an die Klinik zurückgeschickt haben, eher nicht auszugehen. Kritisch bleibt zu hinterfragen, ob durch den Einsatz zusätzlicher Instrumente zum MZP 2 eine Selektion der elegiblen Patienten stattgefunden hat. Die Bearbeitung der Zusatzinstrumente kann sich motivationshemmend auf die Teilnahme an der Befragung zum MZP 2 ausgewirkt haben, so dass an dieser Erhebung eher die motivierteren Patienten teilnahmen. Ob ein Selektionsprozess aufgetreten ist, der die Beantwortung der Fragestellungen im weiteren Verlauf beeinflusst haben könnte, kann aufgrund der Auswertungen nicht ausgeschlossen werden. Aufgrund der Drop-out-Analysen zum MZP 2 (Fragestellung 1) sowie der relativ hohen Beteiligung (236 von 328 PCa-Patienten, Rücklauf von 72,0%) an der Post-Messung ist ein bedeutsamer Selektionsbias jedoch eher unwahrscheinlich.

Zur Stichprobengewinnung ist folgendes kritisch anzumerken. Die Klinik behandelt vorzugsweise Patientinnen mit Brust- sowie Patienten mit Prostatakrebs. Diese beiden Personengruppen stellten auch die elegiblen Patientinnen und Patienten im vorgesehenen Erhebungszeitraum dar. Zu beiden Diagnosegruppen liegen zahlreiche Untersuchungen zur Erfassung der Lebensqualität vor (bspw. Fritzsche et al. 2008 für PCa, McGregor & Antoni 2009 für MammaCa). Somit wurden zwei für die Behandlung von Krebserkrankungen sehr häufig und gut untersuchte Diagnosegruppen ausgewählt. Aufgrund der Annahmen zum Response-Shift

nach Sprangers und Schwartz (1999) ist jedoch davon auszugehen, dass dieses Phänomen vor allem in zeitlicher Nähe zu einem kritischen Lebensereignis besonders ausgeprägt ist. Für die vorliegende Stichprobe bedeutet das, dass ein potenzieller Response-Shift umso eher zu erwarten ist, je näher die Messungen am Zeitpunkt des Auftretens des kritischen Ereignisses (Diagnosemitteilung beziehungsweise Behandlungsbeginn) liegen. Aus diesem Grund wurde die Gesamtzahl der in dem Erhebungszeitraum in der Klinik befragten Patientinnen und Patienten auf diejenigen reduziert, die eine Rehabilitation als Anschlussheilbehandlung (AHB), das heißt in kurzem zeitlichen Abstand zur Primärversorgung, durchführten.

Neben dem bereits genannten response-shift-relevanten Grund zur Auswahl der AHB-Patientinnen und -Patienten resultierte aus der Einschränkung der Stichprobe auch eine Beschränkung auf die Patienten mit PCa. Die Entscheidung wurde aufgrund der ungleichen Häufigkeiten der beiden Behandlungsgruppen getroffen, wodurch die Beeinflussung potenzieller Response-Shift-Prozesse aufgrund verschiedener Grunderkrankungen sowie damit zusammenhängender Akut- sowie Nachsorgebehandlungen reduziert werden sollte. Aufgrund der Anzahl von 328 PCa-Patienten mit AHB zum MZP 1 wurde davon ausgegangen, dass trotz der Beschränkung auf die größte und homogene Subgruppe der gesamten Klientel der Klinik in dem Erfassungszeitraum zum MZP 2 eine ausreichend große Stichprobe zur Bearbeitung der Fragestellungen hervorgeht. Bei einem potenziellen Dropout zum MZP 2 von 40% wäre ein Stichprobenumfang von 197 Personen erreicht worden. Der tatsächliche Rücklauf von 72% lag über dem erwarteten. Mit einem  $n > 200$  ist eine in der Literatur häufig genannte Stichprobengröße als Voraussetzung für die Anwendung des statistischen Verfahrens der Konfirmatorischen Faktorenanalyse (KFA) zur Beantwortung der Hauptfragestellung erfüllt (Kline 2005). Die weiteren Auswertungen zu Verteilungen der Indikatorvariablen bestätigten die Anwendung des Maximum-Likelihood-Verfahrens zur Kalkulation der Schätzer, was ebenfalls für die Verwendung der KFA in der vorliegenden Stichprobe und die Vergleichbarkeit mit bereits publizierten Studien zum Thema spricht.

Kritisch ist anzumerken, dass es für die Klinik nicht möglich war, Angaben über die Tumor-Klassifikation, mögliche Zusatzdiagnosen der Patienten sowie die Zeit seit Diagnosestellung beziehungsweise Behandlungsbeginn in der Akutversorgung zu machen. Somit kann die zentrale Auswertungsstichprobe zwar anhand soziodemographischer Merkmale sowie der Ausprägungen auf den einzelnen Indikatoren der GLQ beschrieben werden, eine erkrankungsbezogene Beschreibung der Stichprobe erfolgt jedoch nicht. Die Ergebnisse der vorlie-

genden Arbeit können daher nicht in den Kontext konkreter Erkrankungsprozesse oder weiterer gesundheits- beziehungsweise erkrankungsbezogener Faktoren gestellt werden. Da ein Response-Shift in großem Maße von der Art, dem Ausmaß und der Dauer der Erkrankung abhängig gemacht wird, tragen Angaben darüber zu einer umfassenden Diskussion potenzieller Einflussfaktoren auf einen Response-Shift bei. Auf eine Differenzierung der aufgedeckten Response-Shift-Prozesse im Kontext der Grunderkrankung sowie einer möglichen Kobeziehungsweise Multimorbidität muss verzichtet werden.

Das gewählte Vorgehen zur Modellableitung ist ein weiterer Aspekt, auf den kritisch eingegangen werden soll. Als Ausgangsmodell zur Erfassung der GLQ bei Krebspatienten dient das Modell nach Böhmer und Luszczynska (2006). Die Autorinnen berechneten die Modellpassung bei einer Stichprobe von Krebspatienten mit unterschiedlichen Grunderkrankungen vor dem Hintergrund einer Unterteilung der Skalen beziehungsweise Symptomitems in Indikatoren und Kausalvariablen der GLQ. Das Modell wies eine gute Passung auf. In der vorliegenden Arbeit wurde das Messmodell der Indikatorvariablen übernommen. Auf die Integration der Kausalvariablen wurde verzichtet. Die Änderung des Modells wurde vorab an einer Stichprobe bestehend aus 185 Patientinnen und Patienten untersucht, die in der Klinik Bad Oexen behandelt wurden. Die Diagnosen Prostata- beziehungsweise Mammacarcinom verteilen sich etwa zu je 50% auf diese Gruppe. Eine homogene Stichprobe der PCa-Patienten wurde für eine methodisch angemessene Modellableitung als zu gering angesehen. Aus diesem Grund wurde das GLQ-Modell der vorliegenden Arbeit in einem ersten Schritt an einer Stichprobe mit unterschiedlichen Krebsdiagnosen überprüft. Der EORTC-QLQ-C30 ist ein krebsspezifisches, aber einzelne Tumorlokalisationen übergreifendes Instrument zur Erfassung der GLQ bei verschiedenen Krebserkrankungen, so dass die Heterogenität für die Modellableitung als unkritisch angesehen wurde. Die Modellprüfung wurde als eine Art ‚Vorvalidierung‘ angesehen. Der Vorteil dieses Vorgehens liegt darin, dass die Güte des Modells an einer Stichprobe überprüft wird, die nicht als zentrale Auswertungsstichprobe fungiert (Barclay-Goddard et al. 2009). Auch wenn nur marginale Änderungen im Vergleich zum Modell von Böhmer und Luszczynska (2006) vorgenommen wurden, indem die Kausalvariablen weggelassen wurden, stellt das Modell an sich ein neues Modell dar. Die Modellableitung sollte an einer separaten Stichprobe erfolgen, ohne dass an den Daten dieser Stichprobe auch die Prüfung auf Response-Shift-Prozesse durchgeführt wurde. Außerdem wurden Modifikationen des Modells vorgenommen, indem Messfehlerkorrelationen aufgenommen wurden, um die Passung des Modells zu verbessern. Das derart modifizierte Modell wurde auf die zentrale

Auswertungsstichprobe übertragen und auf Passung überprüft. Auch wenn die Auswahl der Personen in den beiden Stichproben nicht zufällig erfolgte, ist in diesem Vorgehen eine zusätzliche Sicherung des getesteten Modells zu sehen.

Auch auf die Messfehlerkorrelationen soll eingegangen werden. Die Annahmen der Klassischen Testtheorie gehen davon aus, dass Messfehler unabhängig voneinander und unkorreliert sind. Diese Grundannahme der KTT ist in dem vorliegenden Modell verletzt. Messfehlerkorrelationen sind ein Hinweis auf zusätzliche Einflussfaktoren, die auf die manifesten Variablen wirken, die jedoch im Modell nicht definiert, also auch unabhängig von der GLQ als latenten Konstrukt sind. Somit ist für das vorliegende Modell davon auszugehen, dass es einen Einflussfaktor gibt, der die Physische und Rollenfunktionsfähigkeit sowie einen weiteren zusätzlichen Einflussfaktor, der die Rollenfunktionsfähigkeit und die Soziale Funktionsfähigkeit beeinflusst. Welche Einflussfaktoren das sind, ist aufgrund der angewandten Instrumente nicht näher erklärbar. Denkbar wäre beispielsweise, dass die Physische und Rollenfunktionsfähigkeit durch bestimmte Aspekte der räumlichen Umwelt (beziehungsweise deren Wahrnehmung durch die Person) sowie die Rollenfunktionsfähigkeit und die Soziale Funktionsfähigkeit durch Aspekte sozialer Umweltfaktoren beeinflusst werden. Es ist auch denkbar, dass diese drei Skalen bei der Beantwortung im Rahmen des räumlich-sozialen Kontextes der Klinik vor dem Hintergrund der dort erlebten Aufgaben und Anforderungen und Beziehungsmöglichkeiten beantwortet wurden. Gegen diese Annahme spricht jedoch, dass auch zum MZP 2 die Messfehlerkorrelationen zu einer guten Passung des Messmodells beitragen und diese Befragung in der häuslichen Umgebung der Teilnehmer stattgefunden hat. Vor dem Hintergrund des Linear Latent Variable Model (LLVM, Oort 2004) unterscheidet sich die Definition des Messfehlers von der KTT. Im Messfehler sind nach Oort (2004) nicht nur zufällige Einflussfaktoren enthalten, sondern insbesondere auch spezifische Faktoren, die für die jeweilige Indikatorvariable mitbestimmend und vom latenten Konstrukt der GLQ unabhängig sind. Diese Definition des Messfehlers bestätigt die Verwendung von Messfehlerkorrelationen und verweist auf weitere Einflussfaktoren, die für beide betroffenen Skalen von Bedeutung sind.

Das Messmodell (und das Basismodell als Kombination der Messmodelle zu zwei MZPen) der GLQ weist aufgrund der Werte der globalen Fitmaße eine gute Passung auf. Diese Angaben wurden als entscheidende Beurteilungsgrundlage für die Verwendung des Modells zur Beantwortung der Fragestellungen angesehen. Daneben können Modelle auch

anhand der Maße der lokalen Anpassungsgüte bewertet werden (Angaben zur lokalen Anpassungsgüte der Messmodelle befinden sich im Anhang A5). Diese Kriterien überprüfen einzelne Modellbestandteile. Ein solches Kriterium ist die Höhe der quadrierten Faktorladung als Ausmaß der aufgeklärten Varianz eines Indikators durch das latente Konstrukt. Ein latentes Konstrukt soll mindestens 40% der Varianz eines Indikators aufklären (Homburg & Giering 1996). Dies ist im vorliegenden Modell in der Ableitungsstichprobe hinsichtlich der Indikatoren Emotionale, Soziale und Kognitive Funktionsfähigkeit der Fall. In den Modelltestungen auf Skalen- und auf Itemebene überschreiten die Werte der aufgeklärten Varianz der genannten Indikatoren durch die GLQ das beschriebene Kriterium. Die Skalen Physische und Rollenfunktionsfähigkeit weisen jedoch Faktorladungen auf, die diesem Kriterium nicht gerecht werden. In beiden Modellen (Ableitungsstichprobe und Auswertungsstichprobe) liegen die Werte der quadrierten Faktorladungen unter der 40%-Marke. In den Modellen, die anhand der zentralen Auswertungsstichprobe berechnet wurden, weist die Skala Kognitive Funktionsfähigkeit ebenfalls einen geringeren Wert für die aufgeklärte Varianz des Indikators durch die GLQ auf. Aufgrund der Hauptfragestellung, nämlich der Erfassung von Veränderungen in einzelnen Modellparametern im Zeitverlauf, sowie aufgrund der Ergebnisse der Überprüfung des Gesamtmodells wird das Kriterium der Höhe der Faktorladung als Maß der lokalen Anpassungsgüte in der vorliegenden Arbeit jedoch vernachlässigt. Das Ziel der vorliegenden Arbeit war es, Veränderungen response-shift-relevanter Parameter zwischen zwei MZPen zu erfassen. Dazu zählen auch die Faktorladungen als Hinweise auf eine mögliche Repriorisierung beziehungsweise eine Neukonzeptualisierung des Konstrukts. Bei Passung des Gesamtmodells, die in der vorliegenden Arbeit nachgewiesen wurde, waren die Veränderungen einzelner Parameter entscheidender als der tatsächliche Wert des Parameters als Hinweis auf die Modellbildung. Weiterhin sei angemerkt, dass in der publizierten Literatur, die mittels KFA Response-Shift-Einflüsse auf die Erfassung der GLQ untersucht, ebenfalls die Gütmaße der Gesamtmodelle als entscheidende Kriterien herangezogen wurden (Ahmed et al. 2005c, Oort et al. 2005). Somit entspricht das beschriebene Vorgehen der üblichen Praxis zur Response-Shift-Erfassung mittels KFA. Dies ist insbesondere deshalb nachvollziehbar, da unter der Voraussetzung, dass die aufgeklärte Varianz eines Indikators durch das latente Konstrukt in jedem Fall größer 40% sein soll, dies einer Faktorladung  $\lambda > .634$  entsprechen würde. Eine solche Faktorladung ist für die Entwicklung neuer Modelle ein wünschenswertes Orientierungskriterium. Für die Praxis der Response-Shift-Erfassung scheint dies jedoch eher vernachlässigbar zu sein, da explizit *Veränderungen* auf diesen Parametern untersucht werden sollen und der Rahmen dafür durch diese Vorgabe stark eingeschränkt würde.

## 9. Perspektiven

Nachdem im vorangegangenen Kapitel die Ergebnisse der vorliegenden Arbeit kritisch diskutiert wurden, werden abschließend Themen vorgestellt, die über die konkreten Fragestellungen der vorgestellten Arbeit hinausgehen und in einem weiteren Bezug zum Response-Shift stehen. Daraus lassen sich Anknüpfungspunkte für weitere Forschungsfragen zum Response-Shift ableiten. Das Kapitel ist in zwei Abschnitte unterteilt. Im ersten Abschnitt wird Response-Shift als Teil eines Anpassungsprozesses thematisiert, der allgemein bei Auseinandersetzungen mit kritischen Ereignissen stattfindet und deshalb auch für die onkologische Rehabilitation von Bedeutung ist. Der zweite Abschnitt bezieht sich auf Response-Shift als methodisches Phänomen, das für die Veränderungsmessung relevant ist.

### 9.1 Response-Shift im Rahmen von Anpassungsprozessen

Sprangers und Schwartz (1999) beschreiben ein Modell der Veränderung des Bewertungshintergrunds für die subjektive Lebensqualität im Zusammenhang mit der Auseinandersetzung einer Person mit einer massiven gesundheitlichen Belastung. Diese Veränderung bezeichnen die Autorinnen als Response-Shift, der in Form unterschiedlicher Prozesse auftreten kann: einer Rekalibrierung, einer Repriorisierung und einer Neukonzeptualisierung. Diese Response-Shift-Prozesse sind eingebettet in ein System aus bedingenden Faktoren und führen im Zusammenspiel mit diesen zur individuellen Einschätzung der GLQ.

Das Modell ist einerseits allgemein formuliert und bietet gute Möglichkeiten zur Integration potenziell einflussnehmender Faktoren. Andererseits kann aufgrund der weiten und zum Teil unscharfen Definitionen der einzelnen Modellbestandteile beziehungsweise deren Differenzierung auch Kritik geübt werden. Dem Modell zufolge findet ein katalysierendes Ereignis statt, das über eine Interaktion mit Vorbedingungen und Mechanismus-Variablen zu einem Response-Shift führt. Unklar ist jedoch, was ein Katalysator ist beziehungsweise sein kann, welche Faktoren Vorbedingungen sind und welche Mechanismus-Variablen. In den Anfängen der Response-Shift-Diskussion wurden unter Vorbedingungen von Sprangers und Schwartz (1999) beispielsweise eher stabile Faktoren zusammengefasst. Hierzu zählten unter anderem soziodemographische Aspekte (Alter, Geschlecht, Schulbildung beziehungsweise berufliche Qualifikation etc.), Persönlichkeitsaspekte oder allgemeine Konstrukte, die den persönlichen Hintergrund einer Person beschreiben können (bspw. Spiritualität). In späteren Modellbe-

schreibungen werden hingegen Variablen wie körperliche Beeinträchtigung, soziale und physische Umweltaspekte und ‚assozierte Symptome‘ unter den Vorbedingungen zusammengefasst (Schwartz et al. 2007). Der Bezug zu krankheitsbezogenen Faktoren macht die Abgrenzung zum Katalysator/Auslöser eines Response-Shifts schwierig. Denn auch die Frage, was tatsächlich als Katalysator wirkt, ist sehr weit zu fassen. Im Modell von Response-Shift und GLQ nach Sprangers und Schwartz (1999) lässt sich beinahe jedes stattfindende Ereignis als Katalysator beschreiben, das ein gewisses Maß an Einfluss auf die Verarbeitung des Krankheitsgeschehens hat. Direkt nachvollziehbar erscheinen beispielsweise die Diagnosemitteilung oder der Behandlungsbeginn.

Hinsichtlich der Mechanismus-Variablen, die eher als konkrete Bewältigungsprozesse zu werten sind, bestehen ebenfalls Unklarheiten. Neben allgemeinen Coping-Prozessen werden hier Faktoren wie soziale Vergleichsprozesse, Zielveränderungen oder Erwartungsveränderungen benannt. Kritisch ist diesbezüglich die Abgrenzung zu den eigentlichen Response-Shift-Prozessen. Klar definiert wird den Autorinnen zufolge der Response-Shift als Veränderung der internen Metrik, Werte und Konzepte, also als Ergebnis eines Verarbeitungsprozesses. In den Mechanismusvariablen sind diese zum Teil auch beschrieben, jedoch eher prozesshaft zu verstehen. Daher ist es fraglich, ob sich Ziel- beziehungsweise Erwartungsveränderungen als reine Mechanismusvariablen bezeichnen lassen, da sie in ihrem jeweiligen Zustand Ausdruck eines Response-Shifts sind. Diese unklaren Abgrenzungen der Definitionen einzelner Faktoren im Modell nach Sprangers und Schwartz bieten Ansatzpunkte für die weitere inhaltliche Definition von Response-Shift.

Die Ergebnisse der vorliegenden Arbeit liefern Hinweise für die Bedeutsamkeit von Persönlichkeitsmerkmalen als Vorbedingungen wie positiver Affektivität. Weiterhin wurde der Einfluss von Benefit Finding und Selbstwirksamkeitserwartung als Mechanismus-Variablen untersucht. Auch für diese Variablen lassen sich Hinweise finden, dass sie einen Einfluss auf einen Response-Shift haben können. Jedoch ist aufgrund der geringen Unterschiede zwischen den Mittelwerten der verschiedenen Subgruppen eine klare Aussage über den Einfluss der Moderatorvariablen nicht möglich. Für die Modellbildung wären hier weitere Untersuchungen wünschenswert.

Bisher wenig untersucht ist der Rückkopplungsprozess zwischen der durch einen Response-Shift beeinflussten wahrgenommenen Lebensqualität und dem Neubeginn des gesamt-

ten Verarbeitungsprozesses. Hierzu sind vor allem längsschnittliche Studien geeignet, mit denen die Beziehungen zwischen der Veränderung der GLQ und Response-Shift untersucht werden kann. Aussagen über den Zusammenhang zwischen der Veränderung der Lebensqualität und Response-Shift in Form der Rekalibrierung liefern bisher nur Jansen und Kollegen (2001) sowie Lepore und Eton (2000). Während Lepore und Eton (2000) anhand regressionsanalytischer Berechnungen einen Zusammenhang zwischen dem Ausmaß der Veränderung der Lebensqualität und Response-Shift nachweisen konnten, gehen Jansen und Kollegen (2001) der Frage nach, ob die Richtung der Veränderung der Lebensqualität eine Auswirkung auf einen Response Shift hat. Die Autoren der Studien konnten diesen Zusammenhang bestätigen. In der vorliegenden Arbeit konnte ebenfalls ein deutlicher Hinweis auf den Zusammenhang zwischen dem Ausmaß der Veränderung der GLQ und dem Response-Shift festgestellt werden. Neu an der vorliegenden Arbeit ist, dass der Einfluss der Veränderung der Globalen Lebensqualität auf den Response-Shift mittels KFA untersucht wurde und die einzelnen Response-Shift-Prozesse differenziert erfasst wurden. Es zeigte sich, dass in der Gruppe mit deutlichen Veränderungen der Globalen Lebensqualität auch eine Repriorisierung stattfand. Dieses Teilergebnis wirft die Frage nach der Vergleichbarkeit von Zustandsmessungen bei Gruppen von Patienten mit großen Prä-Post-Differenzen auf. Demnach ist insbesondere bei großen Effekten (nach Cohen 1988, nach der üblichen Effektkalkulation) möglicherweise die Vergleichbarkeit der zugrundeliegenden Zustandsmessungen besonders kritisch zu sehen. Ein solcher Effekt ist möglicherweise stärker response-shift-beeinflusst als ein kleiner Effekt. Das Ergebnis ist von Relevanz für die Diskussion über die Vor- und Nachteile verschiedener Strategien der Ergebnismessung in der Rehabilitation. Vor allem Auswertungsstrategien nach Klassen von Veränderungsverläufen beziehungsweise Mischverteilungsmodellen (Personengruppen mit vergleichbaren Veränderungen) unterliegen möglicherweise in unterschiedlichem Ausmaß einem Response-Shift, da die Veränderungen in den Gruppen mit den größten Effekten (positive und negative Veränderungen) möglicherweise stärker response-shift-behaftet sind als die Gruppen mit geringen Veränderungen. Die Beziehung zwischen dem Ausmaß der Veränderung der Zielvariable (GLQ) und Response-Shift sollte in Zukunft weiter untersucht werden, indem weitere Strategien der Evaluation auf Response-Shift getestet werden. Neben dem Auswertungsvorgehen nach Klassen von Veränderungen oder Mischverteilungsmodellen (Wirtz 2006) ist auch die Zielorientierte Ergebnismessung (ZOE, Gerdes 1998) kritisch zu hinterfragen. Ein Response-Shift kann auch bei diesen Vorgehensweisen die Auswertungen beeinflussen, wenn sich die Bedeutung bestimmter Ziele während der Reha verändert (Repriorisierung/Neukonzeptualisierung) und auch eine Rekalibrierung bezüglich dieser



Zielkriterien nicht zwangsläufig auszuschließen ist. Nach Zwingmann (2003) ist sogar zu erwarten, dass die Rehabilitation den Prozess der Veränderung eines Bewertungshintergrundes für bestimmte persönliche Ziele explizit beabsichtigt.

Aus dem Modell nach Sprangers und Schwartz (1999) lassen sich weitere Fragestellungen ableiten, die nicht direkt einer der Kategorien Katalysator, Vorbedingung oder Mechanismus zuzuordnen sind: Einer Klärung bedarf beispielsweise die Frage nach der Bedeutung von Krankheits- und Behandlungsfaktoren (Krankheitsverlauf, Schwere der Erkrankung, Symptombelastung, Intensität bzw. Invasivität der Behandlung). Neben der Untersuchung dieser potenziellen Einflussfaktoren ist auch der Verlauf eines Response-Shifts noch weitgehend ungeklärt. Wann ein Response-Shift genau einsetzt, ob es eine bestimmte Abfolge von Response-Shift-Prozessen gibt oder wann Response-Shift am massivsten auftritt, sind Fragen, deren Beantwortung Aufschluss über das Phänomen liefern. Bezüglich der Abfolge der Prozesse ist die Beziehung zwischen der Repriorisierung und der Neukonzeptualisierung inhaltlich und statistisch eindeutig, da letztere die Extremform der ersteren darstellt. Anders ist es bei der Beziehung zwischen einer Rekalibrierung und einer Repriorisierung. Es existieren jedoch bisher keine Studien, die diesen Zusammenhang empirisch überprüft haben.

Wenn Response-Shift nicht nur als methodisches Phänomen zu bezeichnen ist, sondern auch als Teil eines inhaltlichen Auseinandersetzungsprozesses, dann stellt sich die Frage, wann ein Response Shift positiv, im Sinne von ‚gut‘ für eine Person ist, und wann er dies nicht ist. In diesem Zusammenhang bieten sich Möglichkeiten der Response-Shift-Untersuchungen, die insbesondere auch für Praktiker von Relevanz sind, da ausgehend von mehr Informationen über diese Prozesse auch Interventionen ableitbar wären, die zu einer besseren Verarbeitung der Erkrankung beziehungsweise Behandlung führen und einen negativen Response-Shift – im Sinne eines maladaptiven Response-Shift – möglichst vermeiden helfen (van Rijn 2009).

Barclay-Goddard und Kollegen (2009) weisen zu Recht darauf hin, dass Response-Shift je nach Untersuchungsbereich ein Zielkriterium einer Intervention sein kann. Wie in der vorliegenden Arbeit bereits beschrieben, ist Response-Shift im Zusammenhang mit der onkologischen Rehabilitation ein zu erwartender, je nach subjektiv erlebter und objektiv vorhandener gesundheitlicher Situation sogar ein erwünschter Effekt. Dieser ist inhaltlich vergleichbar mit einer kognitiven Umstrukturierung im Falle einer Auseinandersetzung mit dem Ziel einer Ak-

komodation an kritische Zustände (Sharpe & Curran 2006). Selbst die AWMF (2002) formuliert in ihren Leitlinien der onkologischen Rehabilitation Zielkriterien, die auf einen Response-Shift hinauslaufen. Methodisch betrachtet ist Response-Shift nach Schwartz und Sprangers (2000) als Confounder zu werten. Inhaltlich kann Response-Shift ein ‚Outcome‘ im Rahmen von Interventionen sein. Vor dem Hintergrund dieser Sichtweise sollten die Möglichkeiten der Erfassung von Response-Shift auch für die Anwendung in der Praxis weiter entwickelt werden, um Response-Shift als eigenständiges Outcome-Kriterium – je nach Interventionsbereich – routinemäßig erfassen zu können. Die Integration von Response-Shift als Outcome-Kriterium in die Routineversorgung bei spezifischen Zielgruppen und geeigneten Outcome-Parametern ließe als weitere Konsequenz erwarten, dass auch die Planung und Durchführung von Interventionen so zu gestalten wären, dass ein Response-Shift evaluiert und gegebenenfalls auch gefördert werden kann.

In diesem Zusammenhang erscheint es beinahe verwunderlich, dass Response-Shift in Evaluationsstudien zur Wirksamkeit von Psychotherapieverfahren bisher kaum explizit Berücksichtigung findet. Zwar werden verschiedene Formen der Veränderungsmessung zur Erfolgsbeurteilung von Psychotherapien oder psychotherapeutischen Maßnahmen eingesetzt (Michalak et al. 2003). Dabei handelt es sich jedoch in der Regel entweder um die indirekte Veränderungsmessung mittels herkömmlicher Effektkalkulationen oder um Verfahren der direkten Veränderungsmessung, für die auch entsprechende Instrumente entwickelt wurden (bspw. Kieler Änderungs-Sensitive Symptomliste (KASSL) von Zielke (1979) oder der Veränderungsfragebogen des Erlebens und Verhaltens (VEV) von Zielke und Kopf-Mehnert (1978)). Auch werden vermehrt Modellierungen von Verlaufs- und Prozessinformationen mit Wachstumsmodellen anhand hierarchischer linearer Modelle (HLM) eingesetzt (Lutz & Grawe 2007), jedoch wird dabei bisher die Response-Shift-Analyse nicht explizit berücksichtigt. Es gibt aktuell kaum Studien, die einen Response-Shift im Zusammenhang mit Psychotherapie untersuchen. Stieglitz (1990) verglich den Thentest zur Response-Shift-Kontrolle mit der direkten und indirekten Veränderungsmessung und kam zu dem Ergebnis, dass die quasi-indirekte Veränderungsmessung (Thentest) und die indirekte Veränderungsmessung vergleichbare Ergebnisse zeigen, jedoch die quasi-indirekte Variante tendenziell höher korreliert mit Einschätzungen von Veränderungen durch Dritte sowie der direkten Veränderungsmessung. Kritisch ist zu diesen Ergebnissen anzumerken, dass Response-Shift auf Symptomskalen erfasst wurde, die derart gestaltet waren, dass nach dem Auftreten von verschiedenen konkreten Beschwerden gefragt wurde. Bei dieser Art von Items ist nach Schwartz und Rapkin

(2004) jedoch davon auszugehen, dass kein Response-Shift auftritt, da es sich nicht um eine subjektive Bewertung evaluationsbezogener Items handelt. Vielmehr standen im Falle der quasi-indirekten Veränderungsmessung in erster Linie die Erinnerung an die Wahrnehmung von körperlichen und psychischen Beschwerden beziehungsweise konkret wahrnehmbare Erlebnisqualitäten im Vordergrund. Ein Unterschied in den Angaben der Patienten ist den Autorinnen zu Folge eher auf Wahrnehmungsfehler im Sinne eines Messfehlers nach der KTT als auf einen Response-Shift zurückzuführen. Von daher ist die Übereinstimmung zwischen der quasi-indirekten und der indirekten Veränderungsmessung nicht überraschend. Dennoch bleibt festzuhalten, dass die Psychotherapieforschung ein Anwendungsfeld für Response-Shift-Untersuchungen ist, auch um die inhaltliche Diskussion des Phänomens zu präzisieren. Psychotherapeutische Methoden wie die kognitive Umstrukturierung können – je nach Inhalt der bearbeiteten Kognitionen und deren Zusammenhang zu Outcomeparametern – als praktische Ansätze zur Herstellung von Response-Shift angesehen werden. Die Untersuchung von Response-Shift und Veränderungen durch Psychotherapie ist auch theoretisch ableitbar: Nach dem Phasenmodell therapeutischer Veränderungen nach Howard und Kollegen (1993) wird eine spezielle Folge von unterschiedlichen Veränderungsbereichen oder Phasen im Genesungsprozess durch Psychotherapien angenommen. Dabei handelt es sich als erstes um eine Remoralisierungsphase, die durch positive Erwartungen und Hoffnungen des Patienten gekennzeichnet ist, da vom Patienten durch den Beginn der Therapie eine Besserung der Beschwerden erwartet wird. In der zweiten Phase, der Remediationsphase, steht eine Linderung der Symptomatik oder Lösung aktueller Probleme im Alltag im Vordergrund, die ursächlich für den Beginn der Psychotherapie gewesen sind. Bei der dritten Phase, der Rehabilitationsphase, geht es um allgemeinere Veränderungen wie zum Beispiel die Besserung der allgemeinen Funktionsfähigkeit, das Entwickeln positiver kognitiver Schemata und Ausprobieren neuer Verhaltensweisen. Insbesondere für diese dritte Phase scheint ein Response-Shift von Bedeutung zu sein, da es hier auch um Änderungen der Grundhaltung des Patienten und dessen Bewertungen von persönlich relevanten Themen und Konstrukten geht. Aus verhaltenstherapeutischer Perspektive würde sich ein Response-Shift dadurch ausdrücken, dass bestimmte Grundüberzeugungen im Verlauf der Rehabilitationsphase an Bedeutung verlieren beziehungsweise relativiert würden zugunsten angemessenerer Überzeugungen, wodurch die Befindlichkeit und Lebensqualität der Person wiederum positiver beeinflusst würde.

Fraglich ist, ob sich komplizierte statistische Verfahren für die Praxis zur Response-Shift-Erfassung eignen. Erfolgversprechend erscheinen in diesem Zusammenhang einfache

Fragebogenverfahren, wie beispielsweise der *heiQ perspective* (Osborne et al. 2006), der konzipiert wurde, um Response-Shift im Zusammenhang mit Maßnahmen zur Patientenschulung zu erfassen. Aktuell wird an einer deutschsprachigen Version gearbeitet (Kirchhof et al. 2009). Wie ein Instrument zur Erfassung eines Response-Shift gestaltet sein muss, stellt eine interessante Fragestellung für die Evaluationsforschung dar.

## 9.2 Response Shift als methodisches Konzept

Schwartz und Rapkin (2004) haben psychometrische Aspekte bei der Erfassung von GLQ unter Berücksichtigung von Response-Shift diskutiert. Ein Aspekt, auf den die Autorinnen eingehen, ist die Art der Variablen, die eingesetzt werden, um GLQ zu erfassen. Dabei wird davon ausgegangen, dass Response-Shift vor allem bei Items auftritt, die eine subjektive Bewertung eines Zustands erfassen (evaluation-based). Bei Fragen nach direkt wahrnehmbaren Inhalten (perception-based) oder objektiv bestimmbaren Leistungen (performance-based) sollte ein Response-Shift nicht auftreten (siehe auch Schwartz et al. 2007). Die Annahme bezüglich der evaluationsbasierten Items kann durch die Ergebnisse der vorliegenden Arbeit bestätigt werden. Vor allem auf den Skalen Physische und Rollenfunktionsfähigkeit werden die Patienten um Auskunft über deren selbsteingeschätzte *Schwierigkeiten* in Zusammenhang mit der Ausführung bestimmter Aktivitäten gebeten. Ähnliches gilt für die beiden Items der Skala Kognitive Funktionsfähigkeit. Auf diesen drei Skalen zeigen sich die ausgeprägteren Response-Shift-Prozesse bei den untersuchten 212 PCa-Patienten. Ausgehend von der Frage, welche Itemformulierung für einen Response-Shift besonders anfällig ist, sollte auch untersucht werden, welche gängigen Lebensqualitätsfragebogen aufgrund der bewertungsorientierten Frageformulierung besonders durch Response-Shift beeinflussbar sind. Die bisherigen Studien zur Response-Shift-Erfassung konzentrieren sich auf die Entwicklung von Methoden zur Erfassung von Response-Shift. Nachdem mittlerweile eine Vielfalt an Verfahren zur Verfügung stehen, sollten diese zur Analyse häufig verwendeter Lebensqualitätsinstrumente eingesetzt werden. Neben der allgemeinen Bewertung von Lebensqualitätsinstrumenten ließen sich so weitere Erkenntnisse über die Art der Itemformulierungen gewinnen.

In diesem Zusammenhang sei auf die Möglichkeit hingewiesen, die der KFA-Ansatz bietet, um die Änderungssensitivität von Lebensqualitätsinstrumenten zu untersuchen. Die Differenzierung eines beobachteten Effekts auf einer Indikatorvariable der GLQ in einen Effekt durch die GLQ (true score) sowie durch weitere spezifische Einflussfaktoren auf diesen Indikator lässt sich in Übereinstimmung bringen mit Ansätzen der Untersuchung zur Ände-

rungssensitivität (Terwee et al. 2003). Wenn ein Instrument konzipiert wurde, um die GLQ als latentes Konstrukt zu erfassen, dann sollten die Indikatorskalen des Fragebogens auch dieses Konstrukt abbilden. Der KFA-Ansatz nach Oort (2005a) bietet somit eine geeignete Möglichkeit, die Eignung und Sensitivität von Instrumenten zur Erfassung von Veränderungen durch die GLQ zu prüfen. Außerdem würden sich mit diesem statistischen Vorgehen auch Einflussfaktoren auf die Änderungssensitivität von Instrumenten untersuchen lassen. Wie im vorangegangenen Abschnitt beschrieben, hat möglicherweise die Itemformulierung einen Einfluss auf einen Response-Shift (Schwartz & Rapkin 2004). Würde durch entsprechende Itemformulierungen der Einfluss spezifischer Faktoren auf einen beobachtbaren Effekt, gemessen mit herkömmlichen Strategien zur Effektkalkulation, reduziert, wäre die entsprechende Skala sensibler für Veränderungen bedingt durch das latente Konstrukt. Die Anwendung des KFA-Ansatzes und die Zerlegung der beobachteten Effekte stellen somit ein geeignetes Vorgehen dar, um die Änderungssensitivität von Instrumenten zu prüfen und weiter zu entwickeln.

In der vorliegenden Arbeit wurde das statistische Verfahren der Konfirmatorischen Faktorenanalyse eingesetzt. Hier soll nochmals die Bedeutung und Entwicklung dieses statistischen Ansatzes für die weitere Forschung zum Thema diskutiert werden. Das von Oort (2004) entwickelte Vorgehen bietet den wesentlichen Vorteil, dass es im Rahmen herkömmlicher Prä-Post-Messungen ohne zusätzliche Messungen beziehungsweise den weiteren Einsatz von Instrumenten alle drei Response-Shift-Prozesse nach Sprangers und Schwartz (1999) abbilden kann. Darüber hinaus sind zusätzlich Differenzierungen innerhalb einzelner Response-Shift-Prozesse möglich (bspw. die uniforme und non-uniforme Rekalibrierung, aber auch Reprioritisierungen verschiedener Ordnungen z. B. auf Indikatorebene, auf Messfehlerebene oder bei Verwendung eines Modells mit mehreren latenten Variablen auch auf Ebene der latenten Variablen). Der wesentliche Nachteil der Methode liegt darin, dass es sich um ein Verfahren mit Auswertungen auf Gruppenebene handelt. Eine Aussage über individuelle Response-Shift-Prozesse ist somit anhand der KFA-Methode nach Oort nicht möglich. Dieser Ansatz kann somit nicht angewendet werden, wenn die Betrachtung individueller Behandlungsverläufe und Response-Shift untersucht werden soll. Grundsätzlich gilt für die statistischen Verfahren der Residualanalyse, der Hierarchischen Latenten Modelle sowie der Wachstumsmodelle, dass diese eine Rekalibrierung nicht erfassen. Im Fokus der genannten Verfahren stehen Veränderungen von Regressionskoeffizienten als Operationalisierung der Reprioritisierung beziehungsweise Neukonzeptualisierung. Eine Rekalibrierung ist nur über eine zusätzliche Datenerhebung beziehungsweise den Einsatz einer weiteren Strategie möglich. Allgemein ge-

sprochen gilt das auch für alle anderen Response-Shift-Erfassungsstrategien mit Ausnahme des Thentests (Howard et al. 1979), der jedoch wiederum keine Repriorisierung beziehungsweise Neukonzeptualisierung erfasst, sowie des qualitativen Ansatzes nach Rapkin und Schwartz (2004). Der qualitative Ansatz ist jedoch bereits eine Kombination aus Interviewverfahren und Thentest und hat den Vorteil, dass er von der Auswahl individueller, persönlich relevanter Werte und Ziele (Repriorisierung) ausgeht und eine retrospektive Einschätzung dieser subjektiv gewählten Werte und Ziele vornimmt. Der Nachteil des Ansatzes ist der massive Aufwand im Rahmen der Datenerhebung und -auswertung. Eine Untersuchung großer Stichproben ist mit diesem Ansatz nicht zu leisten. Für wissenschaftliche Untersuchungen besitzt dieser Zugang jedoch einen hohen Stellenwert, insbesondere für explorative Untersuchungen zum Response-Shift. Für die Praxis ist dieser Zugang ebenfalls eher ungeeignet. Hier wird vor allem der Thentest als Methode der Wahl angesehen. Jedoch sieht sich auch dieser Designansatz erheblicher Kritik ausgesetzt, insbesondere aufgrund von Erinnerungsverzerrungen und impliziten Veränderungseinschätzungen (Norman 2003). Insgesamt scheint der statistische Ansatz der Oort-Methode für wissenschaftliche Fragestellungen, die auf Gruppenebene beantwortet werden können (bspw. Evaluation von Behandlungsmaßnahmen), am vielversprechendsten zu sein, da er Response-Shift-Prozesse genau erfassen kann und in seiner Anwendung ökonomisch ist. Viele verschiedene Fragestellungen lassen sich mit diesem Ansatz sehr gut und auch ohne eine aufwändige Datenerhebung untersuchen. Für wissenschaftliche Untersuchungen beispielsweise im Bereich der rehabilitationswissenschaftlichen Forschung lassen sich vorhandene Datensätze hinsichtlich eines Response-Shifts reanalysieren, wenn die Daten die Analyse über ein Messmodell unter Verwendung der KFA ermöglichen. Denkbar sind beispielsweise Reanalysen zur Klärung von Fragestellungen über Einflussfaktoren wie Art der Erkrankung, Behandlungsform, Verlauf des Response-Shift (je nach Zeitintervall zwischen den Messungen beziehungsweise bei mehr als 2 MZPen).

Nach Ahmed und Kollegen (2009) sowie Millsap und Hartog (1988) ist im Response-Shift auch ein potenzieller Einflussfaktor zu sehen, der die interne Validität selbst von randomisierten kontrollierten Studien (randomized control trials, RCTs) bedroht. Bisher ist davon auszugehen, dass potenzielle Confounder bei Prä-Post-Evaluationsstudien durch den Einbezug einer Kontrollgruppe sowie die Randomisierung der Studienteilnehmer auf Interventions- und Kontrollgruppe optimal kontrolliert werden. Jedoch ist noch nicht klar, welche Bedeutung ein Response-Shift für die Auswertung von RCTs hat. Wie zur Evaluation von Schulungen zu Beginn der Response-Shift-Studien von Howard und Dailey (1979) ist grundsätzlich

bei Interventionen trotz Randomisierung und Kontrollgruppendesign davon auszugehen, dass Response-Shift die Veränderungsmessung beeinflussen kann. Response-Shift kann grundsätzlich auch in RCTs zu einer Unter- oder Überschätzung von Effekten führen (siehe auch Gollwitzer und Jäger 2007, die Response-Shift in diesem Kontext als Messwiederholungsartefakt bezeichnen). Dies ist insbesondere dann der Fall, wenn Response Shift entweder nur in der Interventionsgruppe oder in Interventions- und Kontrollgruppe in unterschiedlichem Maße auftritt (z.B. in Folge der Intervention). Response-Shift ist jedoch anhand der in der vorliegenden Arbeit eingesetzten Erfassungsweise kontrollierbar. Unter Verwendung der KFA-Methode nach Oort lassen sich Interventions- und Kontrollgruppe auf Response-Shift testen, so dass auch hier eine genaue Ableitung eines Response-Shift-Einflusses auf die beobachteten Effekte kalkulierbar wäre. Response-Shift-Erfassung würde demnach zu einer Verbesserung der internen Validität von RCTs beitragen.

Unklar ist zum gegenwärtigen Zeitpunkt außerdem, welche Auswirkung die Passung zwischen Modell und empirischer Datenmatrix auf die Anwendung der KFA hat. Grundsätzlich ist von folgender Annahme auszugehen: Je besser die Güte des Modells, desto genauer werden Response-Shift-Prozesse abbildbar. Da jedoch die KFA-Methode bei jeder Parameterfreisetzung die Varianz/Kovarianzstruktur der Daten verändert, stellen Studien zur Bedeutung der unterschiedlichen Güte von Modellen auf die Response-Shift-Identifikation eine weitere offene Frage dar. So konnten Ahmed und Kollegen (2004) keine Übereinstimmung zwischen der Oort-Methode und dem Thentest feststellen, wobei die Passung mancher Modelle in der durchgeführten Studie als grenzwertig inakzeptabel bewertet werden können. Zur weiteren Untersuchung dieses Themas sind Simulationsstudien denkbar, in denen die Passung zwischen theoretischen Modellen und empirischer Datenbasis variiert und entsprechend kontrolliert wird, um die Auswirkungen auf Response-Shift zu untersuchen.

Weiterhin sind mehr Studien mit Vergleichen unterschiedlicher Response-Shift-Erfassungen wünschenswert, um die Aussagekraft der bisher eher einzeln eingesetzten Verfahren genauer zu untersuchen. Dies gilt insbesondere für die Verwendung statistischer Verfahren. Obwohl mittlerweile einige Studien zum Vergleich von Response-Shift-Erfassungsmethoden vorliegen (Ahmed et al. 2005b, Ahmed et al. 2009, Nolte et al. 2009, Visser et al. 2005), ist eher von einer ‚dürftigen‘ Datenbasis zu sprechen. Die wenigen Studien zum Thema zeigen einerseits positive Befunde, die auf eine Übereinstimmung der Erfassung von Response-Shift mittels verschiedener Methoden hinweisen (Visser et al. 2005). Andererseits sind die Ergeb-

nisse auch widersprüchlich (Ahmed et al. 2005b, Ahmed et al. 2009, Nolte 2009). Aufgrund der Vielzahl bestehender Methoden zur Response-Shift-Erfassung bieten sich unterschiedliche Kombinationen von Strategien an. Wesentlich für eine valide Aussage über einzelne Erfassungsmethoden ist die Verwendung eines Verfahrens als Maßstab zur Bewertung. Die überwiegende Zahl bisher publizierter Studien zum Response-Shift verwendet einen Thentest. Von daher verwundert es nicht, wenn der Thentest in Vergleichsstudien auch als Maßstab herangezogen wird (Ahmed et al. 2005b). Es ist jedoch aufgrund der berechtigterweise bestehenden Kritik am Thentest davon auszugehen, dass dieser nicht als ‚Goldstandard‘ verwendet werden sollte. Aufgrund der statistischen und inhaltlichen Gestaltung der KFA-Methode nach Oort ist diese eher als best-practice Methode zum gegenwärtigen Stand der Response-Shift-Erfassungen anzusehen. Von daher ist vor dem Hintergrund der Erkenntnisse über bisher durchgeführte und geprüfte Strategien der Response-Shift-Erfassung dieser Ansatz als Bewertungsgrundlage für Vergleichsstudien einzusetzen.

Unter der Annahme, dass die Oort-Methode gegenwärtig als Methode der Wahl zur Erfassung von Response-Shift anzusehen ist, bietet sich folgende Strategie zur Überprüfung des Thentests an: Es sollte eine Invarianztestung hinsichtlich der Prä-Post-Messung mittels Oort-Methode durchgeführt werden, um zu prüfen, ob in der vorliegenden Stichprobe ein Response-Shift auftritt. Wenn entsprechende Ergebnisse ermittelt werden, ist eine Invarianztestung zwischen Post- und retrospektiver Prämessung nach der Oort-Methode durchzuführen. Unter der Annahme, dass diese beiden Beurteilungen quasi zeitgleich vor demselben Bewertungshintergrund erfolgen, sollte hier eine volle Invarianz der Parameter vorliegen. Die Passung des Nullmodells sollte sich durch entsprechende Parameterfreisetzungen nicht mehr statistisch signifikant verbessern. Inhaltlich ist dann vom selben Bewertungshintergrund auszugehen. Weiterhin sollte die Invarianzmessung zwischen der tatsächlichen und der retrospektiven Prämessung entsprechend dem Prä-Post-Vergleich Parameterfreisetzungen erbringen. Dieses Ergebnis würde für die Verwendung des Thentests als Verfahren zur Erfassung der Rekalibrierung sprechen. Somit könnte der Thentest anhand der Oort-Methode validiert werden. Nolte und Kollegen (2009) führten eine Invarianztestung zwischen Post- und Thentestmessung durch und kamen zu dem Ergebnis, dass ein Response-Shift zwischen Post- und Thentestmessung auftrat. Outcome-Kriterium war jedoch nicht GLQ, sondern Empowerment, Selbstmanagement und Compliance. Jelitte und Kollegen (2010) kamen zu einem vergleichbaren Ergebnis in Bezug auf die Überprüfung des Thentests mit der KFA-Methode bei einer Stichprobe von 96 PCa-Patienten in der AHB für das Outcome-Kriterium GLQ. Jedoch ist die



Anzahl der Personen in der Stichprobe zu gering, um valide Aussagen über den Response-Shift-Einfluss auf den Thentest bei PCa-Patienten abzuleiten. Die Ergebnisse dieser Pilotstudie sowie die Ergebnisse von Nolte und Kollegen (2009) stellen die Verwendung des Thentest jedoch in Frage. Weitere Untersuchungen im Rahmen der Erfassung von GLQ in der onkologischen Rehabilitation erscheinen wünschenswert.

Zukünftig sollten Response-Shift-Studien auch die Stichprobenabhängigkeit der Ergebnisse in der Form berücksichtigen, dass sie Response-Shift-Analysen an unabhängigen Stichproben vergleichend durchführen. Neben dem Vergleich von Parameterfreisetzungen bei unabhängigen Stichproben wäre es beispielsweise denkbar, auch die Reihenfolge der Parameterfreisetzungen genauer zu untersuchen. Im Sinne einer Validierung an unabhängigen Stichproben können Response-Shift-Prozesse in Art und Reihenfolge auf ihr tatsächliches Auftreten in der Grundgesamtheit getestet werden. Diesbezüglich konnten Jelitte und Kollegen (2010) im Rahmen einer Überprüfung der in dieser Studie identifizierten Response-Shift-Prozesse eine hohe Übereinstimmung mit Response-Shift-Prozessen in einer von der Patientengruppe vergleichbaren Klientel feststellen (PCa-Patienten, AHB). Jedoch lag bei der gerade genannten Untersuchung zwischen der Prä- und Posteinschätzung ein Zeitraum von vier bis fünf Monaten ab Rehabeginn. Ob die marginalen Differenzen hinsichtlich der identifizierten Response-Shift-Prozesse auf den Unterschied in der zeitlichen Differenz der Erfassungen zurückzuführen sind, sollte in weiteren Studien untersucht werden. Insgesamt sprechen die Ergebnisse jedoch für die Bestätigung der in der vorliegenden Arbeit identifizierten Response-Shift-Prozesse.

Eine weitere Strategie zur Response-Shift-Validierung ist in Studien zu sehen, die eine Kontrollgruppe einbeziehen, in der davon auszugehen ist, dass kein Response-Shift stattfindet. Da jedoch auch Zeit an sich einen Einfluss auf die Vergleichbarkeit von Messungen haben kann (Oort et al. 2009), ist auch für diesen Fall eine möglichst genaue Spezifizierung des erwarteten Response-Shifts in der Interventionsgruppe erforderlich.

In der vorliegenden Arbeit konnte ein Response-Shift bei PCa-Patienten in der onkologischen Rehabilitation nachgewiesen werden, der die Beurteilung von Veränderungen aufgrund herkömmlich berechneter Effekte auf der Ebene manifester Variablen (Skalenmittelwerte) in Frage stellt. Somit wird das Response-Shift-Phänomen als ein wichtiges Thema für die Beurteilung und Weiterentwicklung der Veränderungsmessung angesehen. Abschließend ist zu

konstatieren, dass die Response-Shift-Erfassung mittels KFA vor dem Hintergrund der Ergebnisse der vorliegenden Arbeit einerseits, als auch aufgrund der bisher publizierten Studienergebnisse zu eingesetzten Verfahren und Strategien zur Response-Shift-Erfassung andererseits, einen gut operationalisierten Zugang zur Erfassung des Response-Shift sowie der Differenzierung verschiedener Response-Shift-Prozesse bietet. Zum gegenwärtigen Zeitpunkt scheint kein anderes Instrument geeignet zu sein, alle Response-Shift-Prozesse durch eine einfache Prä-Post-Messung zu untersuchen.

## Literaturverzeichnis

- Aaronson, N.K. (1991). Methodologic issues in assessing the quality of life in cancer patients. *Cancer*, 67, 844-850.
- Aaronson, N.K., Acquadro, C., Alonso, J. & Apolone, G. (1992). International Quality of Life Assessment (IQOLA) Project. *Quality of Life Research*, 1, 349-351.
- Aaronson, N.K., Ahmedzai, S., Bergmen, B., Bullinger, M., Cull, A., Duez, N.J., Filiberti, A., Flechtner, H., Fleishman, S.B., de Haes, J.C.J.M., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P.B., Schraub, S., Sneeuw, K., Sullivan, M., Takeda, F. for the European Organization for Research and Treatment of Cancer Study Group on Quality of Life (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, 85, 5, 365-376.
- Aaronson, N.K., van Dam, F.S.A.M., Polak, C.E. & Zittoun, R. (1986). Prospects and problems in european psychosocial oncology: A survey of the EORTC study group on quality of life. *Journal of Psychosocial Oncology*, 4, 43-53.
- Adang, E.M.M., Kootstra, G., Baeten, C.G. & Engel, G.L. (1997). Quality of life ratings in patients with chronic illnesses. *Journal of the American Medical Association*, 277, 1038.
- Adang, E.M.M., Kootstra, G., Engel, G.L., van Hooff, J.P. & Merckelbach, H.L.G.L. (1998). Do retrospective and prospective quality of life assessments differ for pancreas-kidney transplant patients. *Transplantation International*, 11, 11-15.
- Affleck, G. & Tennen, H. (1996). Construing benefits from adversity: Adaptional significance and dispositional underpinnings. *Journal of Personality*, 64, 900-922.
- Ahmed, S., Bourbeau, J., Maltais, F. & Mansour, A. (2009). The Oort structural equation modeling approach detected a response shift after a COPD self-management program not detected by the Schmitt technique. *Journal of Clinical Epidemiology*, 62, 1165-1172.
- Ahmed, S. & Mayo, N.E. (2005). Response to donaldson's commentary. *Quality of Life Research*, 14, 2357-2358.
- Ahmed, S., Mayo, N.E., Wood-Dauphinee, S., Hanley, J.A. & Cohen, S.R. (2004). Response-Shift influenced estimates of change in health-related quality of life poststroke. *Journal of Clinical Epidemiology*, 57, 561-570.
- Ahmed, S., Mayo, N.E., Wood-Dauphinee, S., Hanley, J. & Cohen, R. (2005a). Using the patient generated index to evaluate response-shift post-stroke. *Quality of Life Research*, 14, 2247-2257.
- Ahmed, S., Mayo, N.E., Wood-Dauphinee, S., Hanley, J. & Cohen, R. (2005b). The structural equation modeling technique did not show a response-shift, contrary to the results of the then test and the individualized approaches. *Journal of Clinical Epidemiology*, 58, 1125-1133.
- Ahmed, S., Mayo, N.E., Corbiere, M., Wood-Dauphinee, S., Hanley, J. & Cohen, R. (2005c). Change in quality of life of people with stroke over time: True change or response-shift. *Quality of Life Research*, 14, 611-627.
- Ahamavaara, Y. (1954). Transformation analysis of factorial data. *Annales Academiae Scientiarum Fennicae*, B88, 2-2.
- Aiken, L.S. & West, S.G. (1990). Invalidity of true experiments: Self-report pretest biases. *Evaluation Review*, 14, 374-390.
- Akaike, H. (1974). A new look at statistical model identification. *IEEE transactions on Automatic Control*, 19, 716-723.

- Albertsen, P.C., Aaronson, N.K., Muller, M.J., Keller, S.D. & Ware, J.E., Jr. (1997). Health related quality of life among patients with metastatic prostate cancer. *Urology*, 49, 207-217.
- Albrecht, G.L. & Devlieger, P.J. (1999). The disability paradoxon: High quality of life against all odds. *Social Science and Medicine*, 48, 977-988.
- Amelang, M. & Zielinski, W. (1997). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Andresen, E.M., Vahle, V. J. & Lollar, D. (2001). Proxy reliability: health-related quality of life (HRQoL) measures for people with disability. *Quality of Life Research*, 10, 609-619.
- Antoni, M.H., Lehman, J.M., Kilbourn, K.M., Boyers, A.E., Culver, J.L., Alferi, S.M., Yount, S.E., McGregor, B.A., Arena, P.L., Harris, S.D., Price, A.A. & Carver, C.S. (2001). Cognitive-behavioral stress management intervention decreases the prevalence of depression and enhances benefit finding among women under treatment for early-stage breast cancer. *Health Psychology*, 20, 20-32.
- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) (2002). Interdisziplinäre Leitlinien der Deutschen Krebsgesellschaft. Federführung: Arbeitsgemeinschaft Rehabilitation, Nachsorge und Sozialmedizin (ARNS). *Prinzipien der onkologischen Rehabilitation*. <http://www.uni-duesseldorf.de/AWMF/II/032-005.htm>
- Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modelling* (pp. 243-277). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Arbuckle, J.L. (2005). *AMOS 17.0 User's Guide*. Chicago: SmallWaters Corporation.
- Armenakis, A.A. (1988). A review of research on the change typology. *Research in Organizational Change and Development*, 2, 163-194.
- Armenakis, A.A. & Zmud, R.W. (1979). Interpreting the measurement of change in organizational research. *Personal Psychology*, 32, 709-723.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Bagozzi, R.P. & Edwards, J.R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, 1, 45-87.
- Bandura, A. (1997). *Self-efficacy: The exercise of Control*. New York: Freeman.
- Barclay-Goddard, R., Lix, L.M., Tate, R. Weinberg, L. & Mayo, N.E. (2009). Response shift was identified over multiple occasions with a structural equation modeling framework. *Journal of Clinical Epidemiology*, 62, 1181-1188.
- Bar-On, D. & Amir, M. (1993). Re-examining quality of life of hypertensives: A new self-structured measure. *American Journal of Hypertension*, 6, 62S-66S.
- Bar-On, D., Lazar, A. & Amir, M. (2000). Quantitative assessment of Response-Shift in qol research. *Social Indicator Research*, 49, 37-49.
- Bedian, A.G. & Armenakis, A.A. & Randolph, W.A. (1988). The significance of congruence coefficients: a comment and statistical test. *Journal of Management*, 14, 559-566.
- Bellizzi, K.M. & Blank, T.O. (2007). Cancer-related identity and positive affect in survivors of prostate cancer. *Journal of Cancer Survivorship*, 1, 44-48.
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bernhard, J. Hürny, R., Maibach, R., Herrmann, R. & Laffer, U. (1999). Quality of life as subjective experience: Reframing of perception in patients with colon cancer undergo-

- ing radical resection with or without adjuvant chemotherapy. *Annals of Oncology*, 10, 775-782.
- Bernhard, J., Lowy, A., Maibach, R. & Hürny, C. (2001). Response-shift in the perception of health for utility evaluation: an explorative investigation. *European Journal of Cancer*, 37, 1729-1735.
- Bernhard, J., Lowy, A., Mathys, N., Herrmann, R. & Hürny, C. (2004). Health related quality of life: A changing construct? *Quality of Life Research*, 13, 1187-1197.
- Bestmann, B., Rohde, V., Siebmann, J.-U., Galalae, R., Weidner, W. & Kuchler, T. (2006). Validation of the German prostate-specific module. *World Journal of Urology*, 24, 94-100.
- Blank, T.O. & Bellizzi, K.M. (2006). After prostate cancer: Predictors of well-being among long-term prostate cancer survivors. *Cancer*, 2128-2135.
- Blessmann, A., Kohlmann, T. & Raspe, H. (2001). Übereinstimmung zwischen indirekter und direkter Veränderungsmessung an kardiologischen und orthopädischen Rehabilitationspatienten. *Gesundheitswesen*, 63, A48.
- Boccardo, F., Rubagotti, A., Barichello, M., Battaglia, M., Carmignani, G., Comeri et al. (1999). Bicalutamide monotherapy versus flutamide plus goserelin in prostate cancer patients: results of an Italian prostate Cancer Project study. *Journal of Clinical Oncology*, 17, 2027-2038.
- Boehmer, S. & Luszczynska, A. (2006). Two kinds of items in quality of life instruments: 'Indicator and causal variables' in the EORTC-QLQ-C30. *Quality of Life Research*, 15, 131-141.
- Boehmer, S., Luszczynska, A. & Schwarzer, R. (2007). Coping and quality of life after tumor surgery: Personal and social resources promote different domains of quality of life. *Anxiety, Stress, and Coping*, 20, 61-75.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: Wiley & Sons.
- Boosma, A. & Hoogland, J.J. (2001). The robustness of LISREL modelling revisited. In R. Cudeck, S. du Toit & D. Sörbom (Eds.), *Structural equation models: Present and future. A Festschrift in honor of Karl Joreskog* (pp. 139-168). Chicago: Scientific Software International.
- Borghede, G., Karlsson, J. & Sullivan, M. (1997). Quality of life in patients with prostate cancer: results from a Swedish population study. *Journal of Urology*, 158, 1477-1485.
- Borghede, G. & Sullivan, M. (1996). Measurement of quality of life in patients in localized prostatic cancer patients treated with radiotherapy. Development of a prostate cancer-specific module supplementing the EORTC QLQ-C30. *Quality of Life Research*, 5, 212-222.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44, s.176-S181.
- Borsboom, G.J., Korfage, I.J., Essink-Brot, M.-L. & Duivenvoorden, H.J. (2007). The structural equation model technique did not show a Response-Shift, contrary to the results of the then test and the individualized approaches, Letter to the editors. *Journal of Clinical Epidemiology*, 60, 426-427.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. Aufl.). Berlin: Springer.
- Brandeis, J.M., Litwin, M.S., Burnison, C.M. & Reiter, R.E. (2000). Quality of life outcomes after brachytherapy for early stage prostate cancer. *Journal of Urology*, 163, 851-857.
- Brandtstädter, J. & Greve, W. (1994). The aging self: stabilizing and protective processes. *Developmental Review*, 14, 52-80.
- Brandtstädter, J. & Renner, G. (1990). Tenacious goal pursuit and flexible goal adjustment: explication and age-related analysis of assimilative and accommodative strategies of coping. *Psychology and Aging*, 5, 58-67.

- Bray, J.H., Maxwell, S.E. & Howard, G.S. (1984). Methods of analysis with response-shift bias. *Educational and Psychological Measurement*, 44, 781-804.
- Breetvelt, I.S. & van Dam, F.S.A.M. (1991). Underreporting by cancer patients: the case of response-shift. *Social Science & Medicine*, 32, 981-987.
- Broberger, E., Sprangers, M.A.G. & Tishelman, C. (2006). Do internal standards of quality of life change in lung cancer patients? *Nursing Research*, 55, 274-282.
- Brossart, D.F., Clay, D.L. & Willson, V.L. (2002). Methodological and statistical considerations for threats to internal validity in pediatric outcome data: Response-shift in self-report outcomes. *Journal of Pediatric Psychology*, 27, 97-107.
- Brown, C.H. (1983). Asymptotical comparison of missing data procedures for estimating factor loadings. *Psychometrika*, 48, 269-291.
- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways in assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Bullinger, M. (1994). Lebensqualität: Grundlagen und Anwendungen. In: F. Petermann & K.-C. Bergmann (Hrsg.), *Lebensqualität und Asthma* (S. 17-28). München: Quintessenz.
- Bullinger, M. (2000). Lebensqualität – Aktueller Stand und neuere Entwicklung der internationalen Lebensqualitätsforschung. In U. Ravens-Sieberer & A. Cieza (Hrsg.), *Lebensqualität und Gesundheitsökonomie in der Medizin* (S. 13-24). Landsberg: ecomed.
- Byrne, B.M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, applications and programming*. Mahwah, NJ: Erlbaum.
- Byrne, B.M., Shavelson, R.J. & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- Calhoun, L.G. & Tedeschi, R.G. (1999). *Facilitating Posttraumatic growth: a clinicians's guide*. Mahwah, NJ: Lawrence Erlbaum Association, Inc.
- Calhoun, L.G. & Tedeschi, R.G. (2001). Posttraumatic growth: the positive lessons of loss. In R.A. Neimeyer (Ed.): *Meaning reconstruction and the experience of loss* (pp. 157-172), Washington D.C.: American Psychological Association.
- Campbell, D.T. & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Carboon, I., Anderson, V.A., Pollard, A., Szer, J. & Seymour, J.F. (2005). Posttraumatic growth following a cancer diagnosis: Do world assumptions contribute? *Traumatology*, 11, 269-283.
- Carver, C.S. & Scheier, M.F. (1981). *Attention and self-regulation: a control-theory approach to human behavior*. New York: Springer.
- Carver, C.S. & Scheier, M.F. (1982). Control theory: A usefull conceptual framework for personality-social, clinical, and health psychology. *Psychological Bulletin*, 92, 111-135.
- Carver, C.S. & Scheier, M.F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97, 19-35.
- Carver, C.S. & Scheier, M.F. (1999). Themes and issues in the self-regulation of behavior. In: R.S. Wyer Jr. (Ed.): *Advances in Social Cognition*, 12, S. 1-105, Mahwah: Erlbaum.
- Carver, C.S. & Scheier, M.F. (2000). Scaling back goals and recalibration of the affect system are processes in normal adaptive self-regulation: understanding ‚response-shift’ phenomena. *Social Science and Medicine*, 50, 1715-1722.
- Cella, D., Hahn, E.A. & Dineen, K. (2002). Meaningful change in cancer-specific quality of life scores: Differences between improvement and worsening. *Quality of Life Research*, 11, 207-211.

- Cella, D.F., Tulsky, D.S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., Silberman, M., Yellen, S.B., Winicour, P., Brannon, J., Eckberg, K., Lloyd, S., Purl, S., Blendowski, C., Goodman, M., Barnicle, M., Stewart, I., McHale, M., Bonomi, P., Kaplan, E., Taylor, S., Thomas, C.R. & Harris, J. (1993). The Functional Assessment of Cancer Therapy Scale: Development and validation of the general measure. *Journal of Clinical Oncology*, 11, 570-579.
- Chin, K., Fukuhara, S., Takahashi, K., Sumi, K., Nakamura, T., Matsumoto, H., Niimi, A., Hattori, N., Mishima, M. & Nakamura, T. (2004). Response-shift in perception of sleepiness in obstructive sleep apnea-hypopnea syndrome before and after treatment with nasal CPAP. *Sleep*, 27, 490-493.
- Clark, J.A., Inui, T.S., Silliman, R.A., Bokhour, B.G., Krasnow, S.H., Robinson, R.A., Spaulding, M. & Talcott, J.A. (2003). Patients' perceptions of quality of life after treatment for early prostate cancer. *Journal of Clinical Oncology*, 20, 3777-3784.
- Clark, J.A., Wray, N.P. & Ashton, C.M. (2001). Living with treatment decisions: regrets and quality of life among men treated for metastatic prostate cancer. *Journal of Clinical Oncology*, 19, 72-80.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd. Ed.). Hillsdale, NJ: Erlbaum.
- Cole, B.F., Gelber, R.D., Gelber, S. & Mukhopadhyay, P. (2004). A quality-adjusted survival (Q-TWIST) model for evaluating treatments for advanced stage cancer. *Journal of Biopharmaceutical Statistics*, 14, 111-124.
- Cooper, J. & Axson, D. (1982). Effort justification in psychotherapy. In: G. Weary & H.L. Mirels (Eds.), *Integrations of Clinical and Social Psychology* (pp. 214-230). New York: Oxford University Press.
- Cronbach, L.J. & Furby, L. (1970). How we should measure „change“ – or should we? *Psychological Bulletin*, 74, 68-80.
- Crosby, R.D., Kolotkin, R. & Williams, G.P. (2003). Defining clinical meaningful change in health-related quality of life. *Journal of Clinical Epidemiology*, 56, 395-407.
- Cunningham, A.J., Lockwood, G.A. & Cunningham, J.A. (1991). A relationship between self-efficacy and quality of life in cancer patients. *Patient Education and Counseling*, 17, 71-78.
- Curran, D., Fossa, S., Aaronson, N., Kiebert, G., Keuppens, F., Hall, R. & Genito-Urinary Tract Cancer Cooperative Group (GUT-CCG) (1997). Baseline quality of life of patients with advanced prostate cancer, *European Journal of Cancer*, 33, 1809-1814.
- Daig, I. & Lehmann, A. (2007). Verfahren zur Messung der Lebensqualität. *Zeitschrift für Medizinische Psychologie*, 16, 5-23.
- Davis, J.W., Kuban, D.A., Lynch, D.F. & Schellhammer, P.F. (2001). Quality of life after treatment for localized prostate cancer: differences based on treatment modality. *The Journal of Urology*, 166, 947-952.
- Deck, R. & Röckelein, E. (1999). Zur Erhebung soziodemographischer und sozialmedizinischer Indikatoren in den rehabilitationswissenschaftlichen Forschungsverbänden. In Verband Deutscher Rentenversicherungsträger (VDR) (Hrsg.), *Förderschwerpunkt Rehabwissenschaften. Empfehlungen der Arbeitsgruppe „Generische Methoden“, „Routinedaten“ und „Reha-Ökonomie“* (S. 85-102). Frankfurt a.M.: Postverlag.
- Delbrück, H., Schmid, L., Bartsch, H. & Kruck, P. (2000). Zur Ergebnisqualität in der onkologischen Rehabilitation. *Rehabilitation*, 39, 359-362.
- Delbrück, H. (2006). Onkologische Rehabilitation. *Der Onkologe*, 12, 399-400.
- Deutsche Krebshilfe e.V. (2006). *Prostatakrebs. Der blaue Ratgeber 17*. Bonn: Deutsche Krebshilfe.

- Deutsche Rentenversicherung Bund (DRV) (Hrsg.) (2005). *Rentenversicherung in Zeitreihen*. DRV-Schriften Band 22.
- DeVellis, R.F. (2006). Classical test theory. *Medical Care*, 44, S50-S59.
- Dirmaier J., Zaun S., Koch U., Schulz H. (2004). Psychometric properties of the EORTC Quality of Life Questionnaire (EORTC QLQ-C30) in patients of inpatient cancer rehabilitation. *Palliative and Supportive Care*, 2, 115-124.
- Donaldson, G.W. (2005). Structural equation models for quality of life response-shifts: promises and pitfalls. *Quality of Life Research*, 14, 2345-2352.
- Echteld, M.A., Deliens, L., Ooms, M.E., Ribbe, M.W. & van der Wal, G. (2005). Quality of life change and response-shift in patients admitted to palliative care units: a pilot study. *Palliative Medicine*, 19, 381-388.
- Echteld, M.A., van Zuylen, L., Bannink, M., Witkamp, E. & Van der Rijt, C. (2007). Changes in and correlates of individual quality of life in advanced cancer patients admitted to an academic unit for palliative care. *Palliative Medicine*, 21, 199-205.
- Edwards, A.L. (1957). *The technique of attitude scale construction*. New York: Appleton-Century-Crofts.
- Efficace, F., Bottomley, A. & van Andel, G. (2003). Health related quality of life in prostate carcinoma patients. A systematic review of randomized control trials. *Cancer*, 15, 377-388.
- Efron, B. & Tibshiran, (1993). *An Introduction to the bootstrap*. New York: Chapman & Hall.
- Ekman, P. (1982). *Emotion in the human face*. Cambridge, UK: Cambridge University Press.
- Enders, C.K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, 8, 128-141.
- Enders, C.K. & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8, 430-457.
- Enders, C.K. & Peugh, J.L. (2004). Using EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inference. *Structural Equation Modeling*, 11, 1-19.
- Epstein, R.S. (2000). Responsiveness in quality-of-life assessment. Nomenclature, determinants, and clinical applications. *Medical Care*, 38, SII 91-94.
- Eton, D.T. & Lepore, S.J. (2002). Prostate cancer and health-related quality of life: A review of the literature. *Psychooncology*, 11, 307-326.
- Eton, D.T., Lepore, S.J. & Helgeson, V.S. (2001). Early quality of life in patients with localized prostate carcinoma. *Cancer*, 1451-1459.
- Erickson, P. (2000) Assessment of the evaluative properties of health status instruments. *Medical Care*, SII 95-99.
- Erickson, P. (2005). Assessing health status and quality of life of cancer patients: the use of general instruments. . In J. Lipscomb, C.C. Gotay & C. Snyder (Eds.), *Outcomes Assessment in Cancer* (pp. 31-68). Cambridge University Press.
- Esper, P., Mo, F., Chodak, G., Sinner, M., Cella, D. & Pienta, K.J. (1997). Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. *Adult Urology*, 50, 920-928.
- EuroQol Group. (1990). EuroQol – a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy*, 16, 199–208.
- Faller, H. (1998). *Krankheitsverarbeitung bei Krebskranken*. Göttingen: Verlag für Angewandte Psychologie.
- Faller, H. (2001). Krankheitsbewältigung und Überlebenszeit bei Krebskranken. Literaturübersicht und Ergebnisse einer Untersuchung mit Lungenkrebspatienten. *Psychotherapeut*, 46, 20-35.



- Faller, H. (2004). Beeinflussen psychologische Faktoren den Verlauf einer Krebserkrankung? Ergebnisse, Methoden, Mechanismen. *Zeitschrift für Medizinische Psychologie*, 13, 99-108.
- Faller, H., Vogel, H. & Bosch, B. (2000). Erwartungen von Rehabilitanden hinsichtlich der Methoden und Ergebnisse ihrer Rehabilitation - Eine kontrollierte Studie mit Rückenschmerz- und onkologischen Patienten. *Rehabilitation*, 39, 205-214.
- Fayers, P.M. (2004). Quality of life measurements in clinical trials – the impact of causal variables. *Journal of Biopharmaceutical Statistics*, 14, 155-176.
- Fayers, P.M., Aaronson, N., Bjordal, K. & Sullivan, M. (1995). *EORTC QLQ-C30 Scoring Manual*. Brussels: EORTC Study Group on Quality of Life.
- Fayers, P.M. & Hand, D.J. (1997). Factor analysis, causal indicators and quality of life. *Quality of Life Research*, 6, 139-150.
- Fayers, P.M. & Hand, D.J. (2002). Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society*, 165, 233-261.
- Fayers, P.M., Hand, D.J., Bjordal, K. & Groenvold, M. (1997). Causal indicators in quality of life research. *Quality of Life Research*, 6, 393–406. (2002).
- Fayers, P.M. & King, M.T. (2009). How to guarantee finding a statistical significant difference: the use and abuse of subgroup analysis. *Quality of Life Research*, 18, 527-530.
- Fayers, P., Langston, A. & Robertson, C. (2007). Implicit selfcomparisons against others can bias quality of life assessments. *Journal of Clinical Epidemiology*, 60, 1034-1039.
- Ferrans, C. E. (2005). Definition and conceptual models of quality of life. In J. Lipscomb, C.C. Gotay & C. Snyder (Eds.), *Outcomes Assessment in Cancer* (pp. 14-30). Cambridge University Press.
- Fischer, D., Stewart, A.L., Bloch, D.A., Lorig, K., Laurent, D. & Holman, H. (1999). Capturing the patient's view of change as a clinical outcome measure. *Journal of the American Medical Association*, 282, 1157-1162.
- Folkman, S. (1997) Positive psychological states and coping with severe stress. *Social Science and Medicine*, 45, 1207– 1221.
- Folkman, S. & Greer, S. (2000). Promoting Psychological Well-being in the face of serious illness: When theory, research and practice inform each other. *Psychooncology*, 9, 11-19.
- Fossa, S.D., Aaronson, N.K., Newling, D., van Cangh, P.J., Denis, L., Kurth, K.H. & de Pauw, M. (1990). Quality of life and treatment of hormone resistant metastatic prostate cancer. The EORTC Genito-Urinary Group. *European Journal of Cancer*, 26, 1133-1136.
- Fowler, F.J., Barry, M.J., Lu-Yao, G., Wasson, J., Roman, A. & Wenneberg, J. (1995). Effect of radical prostatectomy for prostate cancer on patients quality of life: results from a Medicare survey. *Urology*, 45, 1007-1013.
- Fransson, P. & Widmark, A. (1999). Late side effects unchanged 4-8 years after radiotherapy for prostate carcinoma: a comparison of age-matched controls. *Cancer*, 85, 678-688.
- French, B.F. & Finch, W.H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural equation Modeling*, 13, 378-402.
- Fritzsche, K., Diedrich, D. & Schultze-Seemann, W. (2008). Psychoonkologie des Prostatakarzinoms. Psychosoziale belastungen und Behandlungsansätze: eine systematische Literaturübersicht. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 54, 329-353.
- Galbraith, M.E., Ramirez, J.M. & Pedro, L.W. (2001). Quality of life, health outcomes, and identity for patients with prostate cancer in five different treatment groups. *Oncological Nursing Forum*, 28, 551-560.

- Garratt, A.M., Ruta, D.A., Abdalla, M.I. & Russell, I.T. (1996). Responsiveness of the SF-36 and a condition-specific measure of health for patients with varicose veins. *Quality of Life Research*, 5, 223–234.
- Garson, G. (2008). *Structural Equation Modeling*. Zugriff, 26.04.2010, <http://faculty.chass.ncsu.edu/garson/PA765/structur.htm#modelfit>
- Gerdes, N. (1998). Rehabilitationseffekte bei “zielorientierter Ergebnismessung“. Ergebnisse der IRES-ZOE-Studie 1996/97. *Deutsche Rentenversicherung*, 3-4, S. 217-237.
- Giannakopoulos, X., Charalabopoulos, K., Charalabopoulos, A., Golias, C., Peschos, D. & Sofikitis, N. (2005). Quality of life survey in patients with advanced prostate cancer. *Experimental Oncology*, 27, 13-17.
- Glatzer, W. & Zapf, W. (1984). *Lebensqualität in der Bundesrepublik Deutschland*. Frankfurt: Campus.
- Golembiewski, R.T., Billingsley, K. & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*, 12, 133-157.
- Golembiewski, R.T. & Billingsley, K. (1980). Measuring change in OD panel designs: A response to critics. *Academy of Management Review*, 5, 97-103.
- Gollwitzer, M. & Jäger, R.S. (2007). *Evaluation. Workbook*. Weinheim: Beltz Verlag.
- Gomela, L.G., James, J. & Trabulsi, E.J (2009). Current prostate cancer treatments: Effect on quality of life. *Journal of Urology*, 73, S5a, 28-35.
- Graves, K.D. (2003). Social cognitive theory and cancer patient’s quality of life: A meta-analysis of psychosocial intervention components. *Health Psychology*, 22, 210-219.
- Gregorich, S.E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44, S78-S94.
- Griner-Hill, L. & Betz, D.L. (2005). Revisiting the retrospective Pretest. *American Journal of Evaluation*. 26, 501-517.
- Güthlin, C. (2004). Response-Shift: alte Probleme der Veränderungsmessung, neu angewendet auf gesundheitsbezogene Lebensqualität. *Zeitschrift für Medizinische Psychologie*, 13, 165-174.
- Guyatt, G.H., Norman, G.R., Juniper, E.F. & Griffith, L.E. (2002). A critical look at transition ratings. *Journal of Clinical Epidemiology*, 55, 900-908.
- Guaytt, G.H., Kirshner, B. & Jaeschke, R. (1992). Measuring health status: What are the necessary measurement properties? *Journal of Clinical Epidemiology*, 45, 1341-1345.
- Haaf, H.-G. (2005). Ergebnisse zur Wirksamkeit der Rehabilitation. *Rehabilitation*, 44,e1-e20.
- Hagedoorn, M., Sneeuw, K.C.A. & Aaronson, N.K. (2002). Changes in physical functioning and quality of life in patients with cancer – response-shift and relative evaluation of one’s condition. *Journal of Clinical Epidemiology*, 55, 176-183.
- Haes, de J.C.J.M., Van Knippenberg, F.C.E. & Neijt, J.P. (1990). Measuring psychological and physical distress in cancer patients: structure and application of the Rotterdam Symptom Checklist. *British Journal of Cancer*, 62, 1034-38.
- Haes, de J.C.J.M., de Ruiter, J.H., Tempelaar, R. & Pennink, B.J.W. (1992). The distinction between affect and cognition in the quality of life of cancer patients – sensitivity and stability. *Quality of Life Research*, 1, 315-322.
- Haes, de J.C.J.M. & Olschewski, M. (1998). Quality of life assessment in cross-cultural context: Use of the Rotterdam Symptom Checklist in a multinational randomised trial comparing CMF and Zoladex (Goserlin) treatment in early breast cancer. *Annals of Oncology*, 9, 745-750.
- Hanlon, A.L., Watkins Bruner, D., Peter, R. & Hanks, G.E. (2001). Quality of life study in prostate cancer patients treated with three-dimensional conformal radiation therapy:

- Comparing late bowel and bladder quality of life symptoms to that of normal population. *International Journal of Radiation in Oncology*, 49, 51-59.
- Harman, H. H. (1967). *Modern Factor Analysis* (2nd ed.). Chicago: University of Chicago Press.
- Hays, R.D. & Hadorn, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research*, 1, 73-75.
- Heckhausen, J. & Schulz, R. (1995). A life-span theory of control. *Psychological Review*, 102, 284-304.
- Heim, M. & Schwerte, U. (2006). Rehabilitation bei Patienten mit Prostatakarzinom. *Onkologie*, 12, 434-443.
- Helgeson, V.S., Reynolds, K.A. & Tomich, P.L. (2006). A meta-analytic review of benefit finding and growth. *Journal of Consulting and Clinical Psychology*, 74, 797-816.
- Hergert, A., Koch, U. & Bergelt, C. (2009). Rehabilitationsrelevante Interventionen bei Prostatakarzinompatienten – Ergebnisse einer systematischen Literaturanalyse. *Abstract-Sammlung Forschungstag der Medizinischen Fakultät der Universität Hamburg*, 87-88.
- Herr, H.W. & O'Sullivan, M. (2000). Quality of life of asymptomatic men with nonmetastatic prostate cancer on androgen deprivation therapy. *Journal of Urology*, 163, 1743-1746.
- Herschbach, P. (2002). Das „Zufriedenheitsparadoxon“ in der Lebensqualitätsforschung. Wovon hängt unser Wohlbefinden ab? *Psychotherapie, Psychosomatik und Medizinische Psychologie*, 52, 141-150.
- Herschbach, P. & Keller, M. (1997). Onkologische Erkrankungen. In F. Petermann (Hrsg.), *Rehabilitation: Ein Lehrbuch der Verhaltensmedizin* (S. 217-244). Göttingen: Hogrefe.
- Herschbach, P., Marten-Mittag, B. & Henrich, G. (2003). Revision und psychometrische Prüfung des Fragebogens zur Belastung von Krebskranken (FBK-R23). *Zeitschrift für Medizinische Psychologie*, 12, 69-76.
- Homburg, C. & Giering, A. (1996). Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung. *Marketing ZFP*, 18, 5-24.
- Hoogland, J.J. & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26, 329-367
- Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *Journal of Experimental Education*, 50, 200-204.
- Hoogstraten, J. (1985). Influence of objective measures on self-reports in a retrospective pretest-posttest design. *Journal of Experimental Education*, 53, 207-210.
- Horn, J.L. & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- Howard, G.S. & Dailey, P.R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64, 144-150.
- Howard, G.S., Ralph, K.M., Gulanick, N.A., Maxwell, S.E. Nance, D.W. & Gerber, S.K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement*, 3, 1-23.
- Howard, G.S., Millham, J., Slaten, S. & O'Donnell, L. (1981). Influence of subject response style effects on retrospective measures. *Applied Psychological Measurement*, 5, 89-100.
- Howard, K.I., Lueger, R.J., Maling, M.S. & Martinowich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology*, 61, 678-685.
- Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hu, J.C., Elkin, E.P., Pasta, D.J., Lubeck, D.P., Kattan, M.W., Carroll, P.R. & Litwin, M. (2004). Predicting quality of life after radical prostatectomy: Results from CaPSURE. *The Journal of Urology*, 703-708.

- Hunt, S.M., McEwen, J., McKenna, S.P., Williams, J. & Papp, E. (1981). The Nottingham Health Profile: Subjective health status and medical consultation. *Social Science & Medicine*, 15A, 221-229.
- Igl, W., Zwingmann, C. & Faller, H. (2005). Änderungssensitivität. *Die Rehabilitation*, 44, 100-106.
- Igl, W., Zwingmann, C. & Faller, H. (2006). Änderungssensitivität von Fragebogen zur Erfassung der subjektiven gesundheit – Ergebnisse einer prospektiven vergleichenden Studie. *Die Rehabilitation*, 45, 232-242.
- Izard, C.E. (1991). *The psychology of emotions*. New York: Plenum.
- Jaeschke, R., Singer, J. & Giyatt, G. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, 10, 410-415.
- Jansen, S.J.T., Stiggelbout, A.M., Nooij, M.A., Noordijk, E.M. & Kievit, J. (2001). Response-Shift in quality of life measurement in early-stage breast cancer patients undergoing radiotherapy. *Quality of Life Research*, 9, 603-615.
- Jansen, S.J.T., Stiggelbout, A.M., Wakker, P.P., Nooij, M.A., Noordijk, E.M. & Kievit, J. (2000). Unstable preferences: A shift in valuation or an effect of the elicitation procedure. *Medical Decision Making*, 20, 62-71.
- Jelitte, M., Faller, H., Schulte, T. (2010). Retrospektive Veränderungsmessung und Response-Shift. In Verband Deutscher Rentenversicherungsträger (Hrsg.), 19. Rehabilitationswissenschaftliches Kolloquium: Qualität in der Rehabilitation – Management, Praxis, Forschung vom 8. bis 10. März 2010 in Leipzig (S. 103-104). Frankfurt am Main: VDR-Selbstverlag.
- Jelitte, M., Schulte, T., Reusch, A., Igl, W. & Faller, H. (2006). Response-Shift als Einflussfaktor auf Veränderungsmessungen der gesundheitsbezogenen Lebensqualität in der onkologischen Rehabilitation. In Verband Deutscher Rentenversicherungsträger (Hrsg.), 15. Rehabilitationswissenschaftliches Kolloquium: Rehabilitation und Arbeitswelt – Herausforderungen und Strategien – vom 13. bis 15. März 2006 in Bayreuth (S. 103-104). Frankfurt am Main: VDR-Selbstverlag.
- Jobe, J.B. (2003). Cognitive Psychology and self-reports: Models and methods. *Quality of Life Research*, 12, 219-227.
- Jöreskog, K.G. & Sörbom, D. (1986). *LISREL-VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*. Mooresville: Scientific Software, Inc.
- Johnstone, P.A., Gray, C & Powell, C.R. (2000). Quality of life in T1-3N0 prostate cancer patients treated with radiation therapy with minimum 10year follow-up. *International Journal of Radiation Oncology, Biology, Physics*, 46, 833-838.
- Joore, M.A., Potjewijd, J., Timmerman, A.A. & Anteunis, L.J.C. (2002). Response-shift in the measurement of quality of life in hearing impaired adults after hearing aid fitting. *Quality of Life Research*, 11, 299-307.
- Karakiewicz, P.I., Kattan, M.W., Tanguay, S., Elhilali, M.M., Bazinet, M., Scardino, P.T. & Aprikian, A. (2003). Cross-cultural validation of the UCLA prostate cancer index. *Adult Urology*, 61, 302-307.
- Keller, M. & Mussell, M. (2005). Psychoonkologische Basisdokumentation (PO-Bado). *Forum DKG*, 20, 58-61.
- Kinsinger, D.P., Penedo, F.J., Antoni, M.H., Dahn, J.R., Lechner, S. & Schneiderman, N. (2006). Psychosocial and sociodemographic correlates of Benefit Finding in men treated for localized prostata cancer. *Psychooncology*, 15, 954-961.
- Kirchhof, R., Schuler, M., Musekamp, G., Nolte, S., Osborne, R. H., Ehlebracht-König, I., Faller, H., Gutenbrunner, C., & Schwarze, M. (2009). Erfassung und Bewertung der Wirksamkeit von Patientenschulungen bei chronischen Erkrankungen: Übersetzung und Evaluation des Health Education Impact Questionnaire (heiQ) in Deutschland. *Physika-*

- Kirschner-Hermanns, R. & Jaske, G. (2002). Quality of life following radical prostatectomy. *Critical Reviews in Oncology/Hematology*, 43, 141-151.
- Kirshner, B. & Gyatt, G. (1985). A methodological framework for assessing health indices. *Journal of Chronical Disorders*, 38, 27-36.
- Kline, R.B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd Ed.), New York: Guilford Press.
- Koch, U. (2000). Aktuelle Entwicklungen in der onkologischen Rehabilitation. *Die Rehabilitation*, 39, 315-316.
- Kohlmann, T. & Raspe, H. (1998). Zur Messung patientennaher Erfolgskriterien in der medizinischen Rehabilitation: Wie gut stimmen „indirekte“ und „direkte“ Methoden der Veränderungsmessung überein? *Die Rehabilitation*, 37, Suppl.1, S30-S37.
- Korfage, I.J., de Koning, H.J. & Essink-Bot, M-L. (2007). Response-shift due to diagnosis and primary treatment of localized prostate cancer: a then-test and a vignette study. *Quality of Life Research*, 16, 1627-2634.
- Korfage, I.J., Essink-Bot, M.-L., Borsboom, G.J.J.M., Madalinska, J.B., Kirkel, W.J., Habbema, J.D.F., Schröder, F.H. & de Koning, H.J. (2005). Five-year follow-up of health-related quality of life after primary treatment of localized prostate cancer. *International Journal of Cancer*, 291-296.
- Krause, N.M. & Jay, G. M. (1994). What do global self-rated health items measure? *Medical Care*, 32, 930-942.
- Kressin, N.R., Spiro, A., Skinner, K.M. (2000). Negative affectivity and health-related quality of life. *Medical Care*, 38, 858-867.
- Krischke, N.R. (1995). *Lebensqualität und Krebs*. München: Quintessenz.
- Krohne, H.W., Egloff, B., Kohlmann, C.-W. & Tausch, A. (1996). Untersuchung mit einer deutschen Version der „Positive and Negative Affect Schedule“ (PANAS). *Diagnostica*, 42, 139-156.
- Lachman, M.E. & Burack, O.R. (1993). Planning and control processes across the life course: an overview. *International Journal of Behavioral Development*, 16, 131-145.
- Lam, T.C.M. & Bengo, P. (2003). A Comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24, 65-80.
- Lazarus, R.S. & Folkman, S. (1990). *Stress, appraisal, and coping*. New York: Springer.
- Lepore, S.J. (2001). A social-cognitive processing model of emotional adjustment to cancer. In: A. Baum & B.L. Andersen (Eds.), *Psychosocial Interventions for Cancer* (pp. 99-118). APA: Washington, DC.
- Lepore, S.J. & Eton, D.T. (2000). Response-shift in prostate cancer patients: An evaluation of suppressor and buffer models. In C.E. Schwartz & M.A.G. Sprangers (Eds.), *Adaption to Changing Health. Response-Shift in Quality-of-Life Research* (pp. 37-51). Washington: American Psychological Association.
- Lev, E.L. (1997). Bandura's theory of self-efficacy: Application to oncology. *Scholarly Inquiry for Nursing Practice*, 11, 21-37.
- Liang, M.H. (2000). Longitudinal construct validity. Establishment of clinical meaning in patient evaluative instruments. *Medical care*, 38, SII 84-90.
- Lim, A.J., Brandon, A.H., Fiedler, J., Brickman, A.L., Boyer, C.I., Raub, W.A., Jr. & Soloway, M.S. (1995). Quality of life: Radical prostatectomy versus radiation therapy for prostate cancer. *Journal of Urology*, 154, 1420-1425.
- Lindell, M.K. & Drexler, J.A. (1979). Issues in using survey methods for measuring organizational change. *Academy of Management Review*, 4, 13-19.

- Lindell, M.K. & Drexler, J.A. (1980). Equivocality of factor incongruence as an indicator of type of change in OD interventions. *Academy of Management Review*, 5, 105-107.
- Linley, P.A. & Joseph, S. (2004). Positive change following trauma and adversity: A review. *Journal of Traumatic Stress*, 17, 11-21.
- Little, T.D. (1997). Mean and covariance structures (MACS) analysis of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32, 53-76.
- Little, R.J.A. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: John Wiley and Sons.
- Litwin, M.S., Lubeck, D.P., Henning, J.M. & Carrol, P.R. (1998a). Difference in urologist and patient assessments of health related quality of life in men with prostate cancer: Results of the CaPSURE database. *Journal of Urology*, 159, 1988-1992.
- Litwin, M.S. & Shpall, A.I., Dorey, F. & Nguyen, T.H. (1998b). Quality of life outcomes in long-term survivors of advanced prostate cancer. *American Journal of Clinical Oncology*, 21, 327-332.
- Litwin, M.S., McGuigan, K.A., Shipall, A.I. & Dhanani, N. (1999a). Recovery of health related quality of life in the year after radical prostatectomy: Early experience. *The Journal of Urology*, 161, 515-519.
- Litwin, M.S., Flanders, S.C., Pasta, D.J., Stoddard, M.L., Lubeck, D.P. & Henning, J.M. (1999b). Sexual function and bother after radical prostatectomy or radiation for prostate cancer: Multivariate quality-of-life analysis from CaPSURE. *Urology*, 54, 503-508.
- Litwin, M.S. & Talcott, J.A. (2005). Measuring quality of life in prostate cancer: Progress and challenges. In J. Lipscomb, C.C. Gotay & C. Snyder (Eds.), *Outcomes Assessment in Cancer* (pp. 126-215). Cambridge: University Press.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lowy, A. & Bernhard, J. (2004). Quantitative assessment of change in patients' constructs of quality of life: An application of multilevel methods. *Quality of Life Research*, 13, 177-1185.
- Lubeck, B.P., Litwin, M.S., Henning, J.M., Stoddard, M.L., Flanders, S.C. & Carroll, P.R. (1999). Changes in health-related quality of life in the first year after treatment for prostate cancer: Results from CaPSURE. *Urology*, 53, 180-186.
- Lübbe, A.S. (1998). Onkologische Rehabilitation. Entwicklungen und Aufgaben. *Onkologie*, 4, 260-267.
- Luszczynska, A., Scholz, U. & Schwarzer, R. (2005). The general self-efficacy-scale: Multicultural validation studies. *Journal of Psychology*, 5, 439-457.
- Luszczynska, A., Gutiérrez-Dona, B. & Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: Evidence from five countries. *International Journal of Psychology*, 40, 80-89.
- Lutz, W. & Grawe, K. (2007). Psychotherapieforschung: Grundlagen, Konzepte und neue Trends. In B. Strauß, F. Hohagen & F. Caspar (Hrsg.), *Lehrbuch Psychotherapie* (S. 727-768). Göttingen: Hogrefe.
- Lydick, E. & Epstein, R.S. (1993). Interpretation of quality of life changes. *Quality of Life Research*, 2, 221-226.
- MacCallum, R.C. & Austin, J.T. (2000). Application of structural equation modelling in psychological research. *Annual Review of Psychology*, 51, 201-226.
- MacNair, D.M., Lorr, M., Droppleman, I.F. (1992). *Profile of Mood States*. San Diego: Educational and Industrial Testing Service.
- Madalinska, J.B., Essing-Bot, M.L., de Koning, H.J., Kirkels, W.J., van der Maas, P.J. & Schroder, F.H. (2001). Health-related quality of life effects of radical prostatectomy and primary radiotherapy for screen-detected or clinical diagnosed localized prostate cancer. *Journal of Clinical Oncology*, 19, 1619-1628.

- Maercker, A. & Zoellner, T. (2004). The janus face of self-perceived growth: Toward a two-component model of posttraumatic growth. *Psychological Inquiry*, 15, 41-48.
- Manne, S., Ostroff, J., Winkel, G., Goldsetin, L., Fox, K. & Grana, G. (2004). Posttraumatic growth after breast cancer: Patient, partner, and couple perspective. *Psychosomatic Medicine*, 66, 442-454.
- Marsh, H.W., Balla, J.R. & Hau, K.F. (1996). An evaluation of incremental Fit Indices: A clarification of mathematical and empirical properties. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced Structural Equation Modeling. Issues and Techniques* (pp. 315-355). Mahwah: Lawrence Erlbaum.
- Marsh, H. W. & Hau, K. T. (1999). Confirmatory factor analysis: strategies for small sample sizes. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 251-284). Newbury Park, CA : Sage.
- Maurischat, C., Morfeld, M., Kohlmann, Th. & Bullinger, M. (Hrsg.) (2004). *Lebensqualität. Nützlichkeit und Psychometrie des Health Survey SF-36/SF-12 in der medizinischen Rehabilitation*. Lengerich: Papst.
- Mayo, N.E., Scott, S.C., Dendukui, N., Ahmed, S. & Wood-Dauphinee, S. (2008). Identifying response shift statistically at the individual level. *Quality of Life Research*, 17, 627-639.
- McCammon, K.A., Kolm, P., Main, B. & Schellhammer, P.F. (1999). Comparative quality of life analysis after radical prostatectomy or external beam radiation for localized prostate cancer. *Urology*, 54, 509-516.
- McGregor, B.A. & Antoni, M.H. (2009). Psychological intervention and health outcomes among women treated for breast cancer: A review of stress pathways and biological markers. *Brain, Behavior, and Immunity*, 23, 159-166.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Meredith, W. & Teresi, J.A. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44, S69-S77.
- Michalak, J., Kosfelder, J., Meyer, F. & Schulte, D. (2003). Messung des Therapieerfolgs: Veränderungsmaße oder retrospektive Erfolgsbeurteilung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 32, 94-103.
- Miller, D.C., Sanda, M.G., Dunn, R.L., Montie, J.E., Pimentel, H., Sandler, H.M., McLaughlin, W.P. & Wei, J.T. (2005). Long-term outcomes among localized prostate cancer survivors: health-related quality-of-life changes after radical prostatectomy, external radiation, and brachytherapy. *Journal of Clinical Oncology*, 23, 2772-2780.
- Millsap, R.E. & Everson, H.T. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research*, 26, 479-497.
- Millsap, R.E. & Everson, H.T. (1993). Methodology review statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17, 297-334.
- Millsap, R.E. & Hartog, S.B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, 73, 574-584.
- Moinpour, C.M., Savage, M.J., Troxel, A., Lovato, L.C., Eisenberger, M., Veith, R.W. Higgins, B., Skeel, R., Yee, M., Blumenstein, B.A., Crawford, E.D. & Meyeskens, F.L. (1998). Quality of life in advanced prostate cancer: results of a randomized therapeutic trial. *Journal of the National Cancer Institute*, 90, 1537-1544.
- Mohamed, N.E. & Böhmer, S. (2004). Die deutsche Version der Benefit Finding Skala: Ihre psychometrische Eigenschaften bei Tumorpatienten. *Zeitschrift für Medizinische Psychologie*, 2, 85-91.
- Mols, F., van de Poll-Franse, L.V., Vingerhoets, A.J.J.M., Hendriks, A., Aaronson, N.K., Houterman, S., Coebergh, J.W.W. & Essink-Bot, M.L. (2006). Long-term quality of life among dutch prostate cancer survivor. *Cancer*, 2186-2196.

- Niebrügge, S. (1999). *Die Lebensqualität von Krebspatienten in der stationären onkologischen Rehabilitation*. Lengerich: Pabst.
- Nieuwkerk, P.T., Tollenaar, M.S., Oort, F.J., Sprangers, M.A.G. (2007). Are retrospective measures of change in quality of life more valid than prospective measures? *Medical Care*, 45, 199-205.
- Nolte, S., Elsworth, G.R., Sinclair, A.J. & Osborne, R.H. (2009). Tests of measurement invariance failed to support the application of the “then-test”. *Journal of Epidemiology*, 62, 1173-1180.
- Norman, G.R., Sloan, J.A. & Wyrwich, K.W. (2003). Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation. *Medical Care*, 5, 582-592.
- Norman, G. (2003) Hi! How are you? Response-shift, implicit theory and differing epistemologies. *Quality of Life Research*, 12, 239-249.
- Norman, G.R., Stratford, P. & Regehr, G. (1997). Methodological problems in retrospective computation of responsiveness to change: The lessons of Cronbach. *Journal of Clinical Epidemiology*, 50, 869-879.
- Norman, G.R., Sridhar, F.G., Guyatt, G.H. & Walter, S.D. (2001). Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Medical Care*, 39, 1039-1047.
- Northouse, L. L., Mood, D., Kershaw, T., Schafenacker, A., Mellon, S., Walker, J., Galvin, L. & Decker, V. (2002). Quality of life of women with recurrent breast cancer and their family members. *Journal of Clinical Oncology*, 20, 4050-4064.
- Nübling, R., Steffanowski, A., Wittmann, W.W. & Schmidt, J. (2004). Strategien der Ergebnismessung am Beispiel der psychosomatischen Rehabilitation. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 65, 35-44.
- O’Boyle, C.A., McGee, H.K. & Browne, J.P. (2000). Measuring response-shift using the schedule for evaluation of individual quality of life. In C.E. Schwartz & M.A.G. Sprangers (Eds.), *Adaption to Changing Health. Response-Shift in Quality-of-Life Research* (pp. 123-136). Washington: American Psychological Association.
- O’Leary, V.E., Alday, C.S. & Ickovics, J.R. (1998). Models of lifechange and posttraumatic growth. In R.G. Tedeschi, C.L. Parks & R.G. Calhoun (Eds.), *Posttraumatic growth: Positive change in the aftermath of crisis* (pp. 127-151). Mahwah, NJ: Erlbaum.
- Oort, F.J. (2004). Types of change in self-report data: definition, interpretation, and operationalization. In C. van Dijkum, J. Blasius, H. Kleijer & B. van Hilten (Eds.), *Recent developments and applications in social research methodology* (pp. 1-21). Proceedings of the Rc33 Sixth International Conference on Social Science Methodology: Amsterdam.
- Oort, F.J. (2005a). Using structural equation modeling to detect response-shift and true change. *Quality of Life Research*, 14, 587-598.
- Oort, F.J. (2005b). Towards a formal definition of response-shift (In reply to G.W. Donaldson). *Quality of Life Research*, 12, 2353-2355.
- Oort, F.J., Visser, M.R.M. & Sprangers, M.A.G. (2005). An application of structural modeling to detect response-shift and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research*, 14, 599-609.
- Oort, F.J., Visser, M.R.M. & Sprangers, M.A.G. (2009). Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, 62, 1126-1137.
- Osborne, R.H., Hawkins, M. & Sprangers, M.A.G. (2006). Change of perspective: A measurable and desired outcome of chronic disease self-management intervention programs that violates the premise of preintervention/postintervention assessment. *Arthritis Care and Research*, 55, 458-465.



- Osoba, D., Zee, B., Pater, J., Warr, D., Kaizer, L. & Latreille, J. (1994). Psychometric properties and responsiveness of the EORTC quality of life questionnaire (QLQ-C30) in patients with breast, ovarian and lung cancer. *Quality of Life Research*, 3, 353-364.
- Osoba, D., Tannock, I.F., Ernst, D.S. & Neville, A.J. (1999). Health-related quality of life in men with metastatic prostate cancer treated with prednisone alone or mitoxandron and prednisone. *Journal of Clinical Oncology*, 17, 1654.
- Park, C.L., Cohen, L.H. & Murch, R.L. (1996). Assessment and prediction of stress related growth. *Journal of Personality*, 64, 71-105.
- Park, C.L. & Folkman, S. (1997). Meaning in the context of stress and coping. *Review of General Psychology*, 1, 115-144.
- Penson, D.F., Litwin, M.S. & Aaronson, N.K. (2003) Health related quality of life in men with prostate cancer. *Journal of Urology*, 169, 1653-1661.
- Ployhart, R.E. & Oswald, F.I. (2004). Application of mean and covariance structure analysis: integrating correlational and experimental approaches. *Organizational Research Methods*, 7, 27-65.
- Pohl, N.F. (1982). Using retrospective pre-ratings to counteract response-shift confounding. *Journal of Experimental Education*, 50, 211-214.
- Postular, D., Eddy, M.S. & Adang, M.M. (2000). Response-shift and adaption in chronic ill patients. *Health Economics*, 20, 186-193.
- Raju, N.S., Laffitte, L.J. & Byrne, B.M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87, 517-529.
- Rapkin, B.D. (2000). Personal goals and response-shifts: Understanding the impact of illness and events on the quality of life of people living with AIDS. In C.E. Schwartz & M.A.G. Sprangers (Eds.), *Adaption to Changing Health. Response-Shift in Quality-of-Life Research* (pp. 53-71). Washington: American Psychological Association.
- Rapkin, B.D. & Fisher, K. (1992). Personal goals of older adults: Issues in assessment and prediction. *Psychology and Aging*, 7, 127-137.
- Rapkin, B.D. & Schwartz, C.E. (2004). Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of Response-Shift. *Health and Quality of Life Outcomes*, 24, 1-12.
- Randolph, W.A. & Elloy, D.F. (1989). How can OD consultants and researchers assess gamma change? A comparison of two analytical procedures. *Journal of Management*, 15, 633-648.
- Razmjou, H., Schwartz, C.E., Yee, A. & Finkelstein, J.A. (2009). Traditional assessment of health outcome following total knee arthroplasty was confounded by Response-Shift phenomenon. *Journal of Clinical Epidemiology*, 62, 91-96.
- Razmjou, H., Yee, A. Ford, M. & Finkelstein, J.A. (2006). Response-shift in outcome assessment in patients undergoing total knee arthroplasty. *Journal of Bone and Joint Surgery*, 2590-2595.
- Ravens-Sieberer, U. & Cieza, A. (2000). Lebensqualitätsforschung in Deutschland – Forschungsstand, Methoden, Anwendungsbeispiele und Implikationen. In: U. Ravens-Sieberer & A. Cieza (Hrsg.), *Lebensqualität und Gesundheitsökonomie in der Medizin. Konzepte, Methoden, Anwendung* (S. 25-50). München: Ecomed.
- Rees, J., Clarke, M.G., Waldron, D., O'Boyle, C., Ewings, P. & MacDonagh, R.P. (2005). The measurement of Response-Shift in patients with advanced prostate cancer and their partners. *Health and Quality of Life Outcomes*, 3, 21-29.
- Rees, J., Waldron, D., O'Boyle, C., Ewings, P. & MacDonagh, R. (2003). Prospective vs retrospective assessment of lower urinary tract symptoms in patients with advanced prostate cancer: the effect of 'Response-Shift'. *British Journal of Urology International*, 92, 703-706.

- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. Oldenbourg: München.
- Reise, S.P., Widaman, K.F. & Pugh, R.H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.
- Reise, S.P. & Widaman, K.F. (1999). Assessing the fit of measurement models at the individual level: a comparison of item response theory and covariance structure approaches. *Psychological Methods*, 4, 3-21.
- Revicki, D.A., Feeney, D., Hunt, T.L. & Cole, B.F. (2006). Analyzing oncology clinical trials data using the Q-TWIST method: Clinical importance and sources for health state preference data. *Quality of Life Research*, 15, 411-423.
- Richards, T.A. & Folkman, S. (2000). Response-shift: A coping perspective. In C.E. Schwartz & M.A.G. Sprangers (Eds.), *Adaption to Changing Health. Response-Shift in Quality-of-Life Research* (pp. 25-36). Washington: American Psychological Association.
- Ring, L., Höfer, S., Heuston, F., Harris, D. & O'Boyle, C.A. (2005). Response-shift masks the treatment impact on patient reported outcomes (PROs): the example of individual quality of life in edentulous patients. *Health and Quality of Life Outcomes*, 3, 55.
- Ringdal, G.I. & Ringdal, K. (1993). Testing the EORTC quality of life questionnaire on cancer patients with heterogeneous diagnoses. *Quality of Life Research*, 2, 129-140.
- Roberts, K.J., Lepore, S.J. & Helgeson, V. (2006). Social-cognitive correlates of adjustment to prostate cancer. *Psychooncology*, 15, 183-192.
- Rogosa, D. & Willett, J.B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, 50, 203-228.
- Rose, M. (2003). *Messung der Lebensqualität bei chronischen Erkrankungen*. Habilitationsschrift. Berlin: Humboldt-Universität.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Göttingen: Verlag Hans Huber.
- Roesch, S.C., Adams, L., Hines, A., Palmores, A., Vyas, P., Tran, C., Pekin, S. & Vaughn, A.A. (2005). Coping with prostate cancer: A meta-analytic review. *Journal of Behavioral Medicine*, 28, 281-293.
- Ruta, D.A., Garrat, A.M., Leng, M., Russell, I.T. & MacDonald, L.M. (1994). A new approach to the measurement of quality of life: the patient-generated index. *Medical Care*, 32, 1109-1126.
- Ruta, D.A., Garrat, A.M., & Russell, I.T. (1999). Patient centered assessment of quality of life for patients with four common conditions. *Quality in Health Care*, 8, 22-29.
- Sanchez-Ortiz, R.F., Broderick, G.A., Rovner, E.S., Wein, A.J., Whittington, R. & Malkowicz, S.B. (2000). Erectile function and quality of life after interstitial radiation therapy for prostate cancer. *International Journal for Impotence Research*, Suppl. 12, S18-S24.
- Schag, C.A.C. & Heinrich, R.L. (1990). Development of a comprehensive quality of life measurement tool: CARES, *Oncology*, 4, 29-50.
- Schapira, M.M., Lawrence, W.F., Katz, D.A., McAuliffe, T.L. & Nattinger, A.B. (2001). Effect of treatment on quality of life among men with clinically localized prostate cancer. *Medical Care*, 39, 243-253.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74.
- Schmitt, N. (1982). The use of analysis of covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*, 17, 343-358.
- Schmitt, N., Pulakos, E. & Lieblein, A. (1984). Comparison of three techniques to assess group-level beta and gamma change. *Applied Psychological Measurement*, 8, 249-260.

- Schover, L.R., Fouladi, R.T., Warnecke, C.L., Neese, L., Klein, E.A., Zippe, C. & Kupelian, P.A. (2002). Defining sexual outcomes after treatment for localized prostate cancer. *Cancer*, 95, 1173-1185.
- Schuck, P. (2000). Designs und Kennziffern zur Ermittlung der Änderungssensitivität von Fragebogen in der gesundheitsbezogenen Lebensqualitätsforschung. *Zeitschrift für Medizinische Psychologie*, 3, 125-130.
- Schulte, D. (1993). Wie soll Therapieerfolg gemessen werden? *Zeitschrift für Klinische Psychologie*, 22, 374-393.
- Schulz, U. & Mohamed, N.E. (2004). Turning the tide: Benefit finding after cancer surgery. *Social Science and Medicine*, 59, 653-662.
- Schwarz, R. (2005). Psychotherapie bei Krebs. In H. Faller (Hrsg). *Psychotherapie bei somatisch Erkrankten* (S 168 – 183). Stuttgart: Thieme.
- Schwarz, R., Bernhard, J., Flechtner, H., Küchler, Th. & Hürny, Ch. (1995). *Lebensqualität in der Onkologie*. München: Zuckschwerdt.
- Schwarz, R. & Hinz, A. (2001). Reference data for the quality of life questionnaire EORTC-QLQ-C30 in the general German population. *European Journal of Cancer*, 37, 1345-1351.
- Schwarzer, R. (1992). *Self-Efficacy: Thought control of action*. Washington, DC: Hemisphere.
- Schwarzer, R. & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, Berks, UK: NferNelson.
- Schwarzer, R. & Jerusalem, M. (Hrsg.) (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin.
- Schwarzer, R., Luszczynska, A., Boehmer, S., Taubert, S. & Knoll, N. (2006). Changes in finding benefit after surgery and the prediction of well-being one year later. *Social Sciences and Medicine*, 63, 1614-1624.
- Schwarzer, R. & Knoll, N. (2003). Positive coping: Mastering demands and searching for meaning. In S.J. Lopez & C.R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures* (pp. 393-409). Washington, DC: APA.
- Schwartz, C.E. (1996). *Preference value assessment: Card sort approach*. Chestnut Hill: Frontier Science and Technology Foundation.
- Schwartz, C.E. (1999). Teaching coping skills enhances quality of life more than peer support: Results of a randomized trial with multiple sclerosis patients. *Health Psychology*, 18, 211-220.
- Schwartz, C.E., Andresen, E.M., Nosek, M.A., Krahn, G.L. & RRTC Expert Panel on Health Status Measurement (2007). Response-Shift theory: Important implications for measuring quality of life in people with disability. *Archives of Physical Medicine and Rehabilitation*, 88, 529-536.
- Schwartz, C.E., Bode, R., Repucci, N., Becker, J., Sprangers, M.A.G. & Fayers, P.M. (2006). The clinical significance of adaption to changing health: A meta-analysis of response-shift. *Quality of Life Research*, 15, 1533-1550.
- Schwartz, C.E., Cole, B.F., Vickrey, B.G. & Gelber, R.D. (1995). The Q-TWIST approach to assessing health-related quality of life in epilepsy. *Quality of Life Research*, 4, 135-141.
- Schwartz, C.E. & Rapkin, B.D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response-shift and appraisal. *Health and Quality of Life Outcomes*, 2, 1-11.

- Schwartz, C.E. & Sprangers, M.A.G. (1999). Methodological approaches for assessing response-shift in longitudinal health-related quality-of-life research. *Social Science & Medicine*, 48, 1531-1548.
- Schwartz, C.E. & Sprangers, M.A.G. (2000). Methodological approaches for assessing response-shift in longitudinal health-related quality-of-life research. In C.E. Schwartz & M.A.G. Sprangers (Eds.), *Adaption to Changing Health. Response-Shift in Quality-of-Life Research* (pp. 81-107). Washington: American Psychological Association.
- Schwartz, C.E. & Sprangers, M.A.G. (2002). An Introduction of quality of life assessment in oncology: the value of measuring patient-reported outcomes. *The American Journal of Managed Care*, 8, S550-559.
- Schwartz, C.E., Sprangers, M.A.G., Carey, A. & Reed, G. (2004). Exploring response-shift in longitudinal data. *Psychology and Health*, 19, 1611-1618.
- Schwartz, C.E. & Sprangers, M.A.G. (2010). Guidelines for improving the stringency of response shift research using the thentest. *Quality of Life Research*, 19, 455-464.
- Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and quasi-experimental Designs for generalized causal inference*. New York: Houghton Mifflin Company.
- Sharpe, L., Butow, P., Smith, C., McConnell, D. & Clarke, S. (2005). Changes in quality of life in patients with advanced cancer – evidence of Response-Shift and response restriction. *Journal of Psychosomatic Research*, 58, 497-504.
- Sharpe, L. & Curran, L. (2006). Understanding the process of adjustment to illness. *Social Science & Medicine*, 62, 1153-1166.
- Shrader-Bogen, C.L., Kjellberg, J.L., McPherson, C.P. & Murray, C.L. (1997). Quality of life and treatment outcomes. *Cancer*, 79, 1977-1986.
- Skeff, K.M., Stratos, G.A. & Bergen, M.R. (1992). Evaluation of medical faculty development program: A comparison of traditional pre/post and retrospective pre/post self-assessment ratings. *Evaluation & the Health Professions*, 15, 350-366.
- Slevin, M.L., Plant, H. & Lynch, D. (1988). Who should measure quality of life, the doctor or the patient. *British Journal of Cancer*, 57, 107-112.
- Sommers, S.D. & Ramsey, S.D. (1999). A Review of Quality-of-Life Evaluations in Prostate Cancer. *Pharmacoeconomics*, 16, 127-140.
- Sonn, G.A., Sadetsky, N., Presti, J.C. & Litwin, M.S. (2009). Differing perceptions of quality of life in patients with prostate cancer and their doctors. *The Journal of Urology*, 182, 2296-2302.
- Spilker, B. (1996). Introduction. In B. Spilker (Ed.): *Quality of Life and Pharmacoeconomics in Clinical Trials* (2<sup>nd</sup> Edition) (pp 1-23). Philadelphia: Lippincott-Raven Publisher.
- Spitzer, W.O., Dobson, A.J., Hall, J., Chesterman, E., Levi, J., Shepherd, R., Battista, R.N. & Catchlove, B.C. (1981). Measuring the quality of life of cancer patients. A concise QL-Index for use by physician. *Journal of Chronic Disease*, 34, 585-597.
- Sprangers, M.A.G. (1996). Response-shift-bias: A challenge to the assessment of patients' quality of life in cancer trials. *Cancer Treatment Review*, 22, S A55-62.
- Sprangers, M.A.G. (2002). Quality of life assessment in oncology. Achievements and challenges. *Acta Oncologica*, 41, 229-237.
- Sprangers, M.A.G. & Aaronson, N.K. (1992). The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. *Journal of Clinical Epidemiology*, 45, 743-760.
- Sprangers, M.A.G. & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest-post-test designs. *Journal of Applied Psychology*, 74, 265-272.
- Sprangers, M.A.G. & Schwartz, C.E. (1999). Integrating response shift into health-related quality-of-life research: A theoretical model. *Social Science & Medicine*, 48, 1507-1515.

- Sprangers, M.A.G. & Schwartz, C.E. (2000). Integrating response-shift into health-related quality-of-life research: A theoretical model. In C.E. Schwartz & M.A.G. Sprangers (Eds.), *Adaption to Changing Health. Response-Shift in Quality-of-Life Research* (pp. 11-23). Washington: American Psychological Association.
- Stamatiadis-Smidt, H., Hausen, H. & Wiestler, O.D. (2006). *Thema Krebs*. Heidelberg: Springer.
- Stanton, A.L., Bower, J.E. & Low, C.A. (2006). Posttraumatic growth after cancer. In L.G. Calhoun & R.G. Tedeschi (Eds.), *Handbook of posttraumatic growth: Research and practice* (pp. 138-175). Mahwah, NJ: Erlbaum.
- Statistisches Bundesamt Deutschland (2005). [www.destatis.de/daten1/stba/html/basis/d/gesu/gesutab19.php](http://www.destatis.de/daten1/stba/html/basis/d/gesu/gesutab19.php)
- Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Steiger, J.H., Shapiro, A. & Browne, M.W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253-264.
- Steenkamp, J.E.M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Steyer, R. & Eid, M. (1993). *Messen und Testen*. Berlin: Springer.
- Steyer, R., Hannover, W. & Tesler, C. (1997). Zur Evaluation intraindividuellere Veränderung. *Zeitschrift für Klinische Psychologie*, 26, 291-299.
- Stieglitz, R.-D. (1990). Validitätsstudien zum retrospektiven Vortest in der Therapiefor-schung. *Zeitschrift für Klinische Psychologie*, 19, 144-150.
- Streiner, D.L. & Norman, G.R. (1989). *Health measurement scales. A practical guide to their development and use*. New York: Oxford University Press, 1989.
- Sugarbaker, P.H., Barofsky, I., Rosenberg, S.A. & Gianola, F.J. (1981). Quality of life assessment of patients in extremity sarcoma clinical trails. *Surgery*, 9, 17-23.
- Talcott, J.A., Rieker, P., Propert, K.J., Clark, J.A., Wishnow, K.I., Loughlin, K.R., Richie, J.P. & Kantoff, P.W. (1997). Patient-reported impotence and incontinence after nerve-sparing radical prostatectomy. *Journal of the National Cancer Institution*, 89, 1117-1123.
- Talcott, J.A. & Clark, J.A. (2005). Quality of life in prostate cancer. *European Journal of Cancer*, 41, 922-931.
- Tedeschi, L.G., Park, C.L., Calhoun, R.G. (1998). Posttraumatic growth: conceptual issues. In L.G. Tedeschi, C.L. Park & R.G. Calhoun (Eds.), *Posttraumatic growth: Positive change in the aftermath of crisis* (pp. 1-22). Mahwah, NJ: Erlbaum.
- Tedeschi, L.G. & Calhoun, R.G. (2004). Posttraumatic growth: Conceptual foundations and empirical evidence. *Psychological Inquiry*, 15, 1-18.
- Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A.H. Tuma & J.D. Maser (Eds.), *Personality and disease* (pp. 178-200). New York: Wiley.
- Terborg, J.R. & Davis, G.A. (1982). Evaluation of a new method for assessing change to planned job redesign as applied to Hackman and Oldham' job characteristic model. *Organizational Behavior and Human Performance*, 29, 112-128.
- Terborg, J.R., Howard, G.S. & Maxwell, S.E. (1980). Evaluation planned organizational change: a method for assessing alpha, beta, and gamma change. *Academy of Management Review*, 5, 109-121.
- Terwee, C.B., Dekker, F.W., Wiersinga, W.M., Prummel, M.F. & Bossuyt, P.M.M. (2003). On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Quality of Life Research*, 12, 349-362.
- Testa, M.A. (2000). Interpretation of quality-of-life outcomes. Issues that affect magnitude and meaning. *Medical Care*, SII 166-174.

- Thompson, B. (2004). *Exploratory and confirmatory factor analysis. Understanding concepts and applications*. Washington: American Psychological Association.
- Thompson, R.C. & Hunt, J.G. (1996). Inside the black box of alpha, beta, and gamma change: Using a cognitive-processing model to assess attitude structure. *Academy of Management Review*, 21, 655-690.
- Thornton, A.A. (2002). Perceiving benefits in the cancer experience. *Journal of Clinical Psychology in Medical Settings*, 9, 153-165.
- Thornton, A.A. & Perez, M.A. (2006). Posttraumatic growth in prostata cancer survivors and their partners. *Psychooncology*, 15, 285-296.
- Thunedborg, K. Allerup, P., Bech, P. & Joyce, C.R.B. (1993). Development of the repertory grid for measurement of individual quality of life in clinical trails. *International Journal of Methods in Psychiatric Research*, 3, 45-56.
- Timmerman, A.A., Anteunis, L.J. & Meesters, C.M. (2003). Response-shift bias and parent-reported quality of life in children with otitis media. *Archives of Otolaryngology – Head and Neck Surgery*, 129, 987-991.
- Torrance, G.W. (1986). Measurement of health state utilities for economic appraisal: A review. *Journal of health Economics*, 5, 1-30.
- Tschuschke, V. (2006). *Psychoonkologie. Psychologische Aspekte der Entstehung und Bewältigung von Krebs*. Stuttgart: Schattauer.
- Tsevat, J., Cook, E.F., Green, M.L., Matchar, D.B., Dawson, N.V., Broste, S.K., Wu, A.W., Phillips, R.S., Oye, R.K. & Goldman, L. (1995). Health values of the seriously ill. *Annals of Internal Medicine*, 122, 514-520.
- Ubel, P.A., Loewenstein, G., Hershey, J., Baron, J., Mohr, T., Asch, D.A. & Jepson, C. (2001). Do nonpatients underestimate the quality of life associated with chronic health conditions because of focusing illusions? *Medical Decision Making*, 21, 190-199.
- Upshaw, H.S. (1962). Own attitude as an anchor in equal appearing intervals. *Journal of Abnormal and Social Psychology*, 64, 85-96.
- Vandenberg, R.J. & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practice, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- Van Rijn, T. (2009). A physiatrist's view of response shift. *Journal of Clinical Epidemiology*, 62, 1191-1195.
- Verband Deutscher Rentenversicherung (VDR) (2005). *VDR Statistik Rehabilitation. Leistungen zur medizinische Rehabilitation, sonstige Leistungen zur Teilhabe und Leistungen zur Teilhabe am Arbeitsleben der gesetzlichen Rentenversicherung im Jahre 2004*. Band 154. Marktheidenfeld: Schleunungsdruck GmbH.
- Verband Deutscher Rentenversicherung (VDR) (1996). *VDR Statistik Rehabilitation des Jahres 1995*. Würzburg: VDR.
- Visser, M.R.M., Smets, E.M.A., Sprangers, M.A.G. & de Haes, H.J.C.J.M. (2000). How response-shift may affect the measurement of change in fatigue. *Journal of Pain and Symptom Management*, 20, 12-18.
- Visser, M.R.M., Oort, F.J. & Sprangers M.A.G. (2005). Methods to detect response-shift in quality of life data: A convergent validity study. *Quality of Life Research*, 14, 629-639.
- Vitinius, F. & Rohde, D. (2006). Uropsychoonkologie. *Onkologie*, 12, 64-68.
- Voogt, E., van der Heide, A., van Leeuwen, A.F., Visser, A.P., Cleiren, M.P.H.D., Passchier, J. & van der Maas, P.J. (2005). Positive and negative affect after diagnosis of advanced cancer. *Psychooncology*, 14, 262-273.
- Wagner, J.A. (2005). Response-shift and glycemic control in children with diabetes. *Health and Quality of Life Outcomes*, 14, 38.

- Waldron, D., O'Boyle, C.A., Kearney, M., Mariaty, M. & carney, D. (1999). Quality-of-life measurement in advanced cancer: Assessing the individual. *Journal of Clinical Oncology*, 17, 3603-3611.
- Ware, J.E. (1984). Methodology in behavioral and psychosocial cancer research. Conceptualizing disease impact and treatment outcomes, *Cancer*, 53 (10 Suppl.), 2316-23.
- Ware, J.E. & Sherbourne, C.D. (1992). The MOS 36-item short-form health status survey (SF-36): 1. Conceptual framework and item selection. *Medical Care*, 30, 473-483.
- Watson, D., Clark, L.A. & Tellegen, A. (1988a). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Watson, D., Clark, L.A. & Carey, G. (1988b). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*, 97, 346-353.
- Watson, D. & Pennebaker, J.W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review*, 96, 234-254.
- Watson, D. & Telegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219-235.
- Weber, B. A., Roberts, B. L., Resnick, M., Deimling, G., Zauszniewski, J. A., Musil, C., et al. (2004). The effect of dyadic intervention on self-efficacy, social support, and depression for men with prostate cancer. *Psychooncology*, 13, 47-60.
- Wei, J.T., Dunn, R.L., Sandler, H.M., McLaughlin, P.W., Montie, J.E., Litwin, M.S., Nyquist, L. & Sanda, M.G. (2002). Comprehensive comparison of health-related quality of life after contemporary therapies for localized prostate cancer. *Journal of Clinical Oncology*, 20, 557-566.
- West, S. G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In: R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75), Thousand Oaks: SAGE Publications.
- Westerman, M.J., The, A.-M., Sprangers, M.A.G., Groen, H.J.M., van der Wal, G. & Hak, T. (2006). Small-cell lung cancer patients are just 'a little bit' tired: Response-shift and self-presentation in the measurement of fatigue. *Quality of Life Research*, 16, 853-861.
- Westerman, M.J., Hak, T., Echteld, M.A., Groen, H.J.M. & van der Wal, G. (2007). Change in what matters to palliative patients: Eliciting information about adaption with SEIQoL-DW. *Palliative Medicine*, 21, 581-586.
- WHOQOL Group (1993). Study protocol for the World Health Organization to develop a quality of life assessment instrument (WHOQOL). *Quality of Life Research*, 2, 153-159.
- Widows, M.R., Jacobsen, P.B., Booth-Jones, M. & Fields, K.K. (2005). Predictors of post-traumatic growth following bone marrow transplantation for cancer. *Health Psychology*, 24, 266-273.
- Wilson, I.B. (1999). Clinical understanding and clinical implication of response-shift. *Social Science & Medicine*, 48, 1577-1588.
- Wirtz, M. (2006). Identifikation typischer Veränderungsverläufe durch Mischverteilungsmodelle. In Verband Deutscher Rentenversicherungsträger (Hrsg.), 15. Rehabilitationswissenschaftliches Kolloquium: Rehabilitation und Arbeitswelt – Herausforderungen und Strategien – vom 13. bis 15. März 2006 in Bayreuth (S. 117-118). Frankfurt am Main: VDR-Selbstverlag.
- Wrosch, C., Heckhausen, J. & Lachman, M.E. (2000). Primary and secondary strategies for managing health and financial stress across adulthood. *Psychology and Aging*, 15, 387-399.
- Wrosch, C., Schulz, R. & Heckhausen, J. (2002). Health stresses and depressive symptomatology in the elderly: The importance of health engagement control strategies. *Health Psychology*, 21, 340-348.

- Wrosch, C. & Scheier, M.F. (2003). Personality and quality of life: The importance of optimism and goal adjustment. *Quality of Life Research*, Suppl. 1, 59, 59-72.
- Wrosch, C., Scheier, M.F., Miller, G.E., Schulz, R. & Carver, C.S. (2003). Adaptive self-regulation of unattainable goals: Goal disengagement, goal reengagement, and subjective well-being. *Personality and Social Psychology Bulletin*, 29, 1494-1508.
- Wrosch, C., Schulz, R., Miller, G.E., Lupien, S. & Dunne, E. (2007a). Physical health problems, depressive mood, and cortisol secretion in old age: Buffer effects of health engagement control strategies. *Health Psychology*, 26, 341-349.
- Wrosch, C., Miller, G.E., Scheier, M.F. & Brun de Pontet, S. (2007b). Giving up on unattainable goals: Benefits for health? *Personality and Social Psychology Bulletin*, 33, 251-265.
- Wyrwich, K.W., Bullinger, M., Aaronson, N., Hays, R.D., Patrick, D.L., Symonds, T. & the Clinical Significance Consensus Meeting Group (2005). Estimating clinically significant differences in quality of life outcomes. *Quality of Life Research*, 14, 285-295.
- Wyrwich, K.W. & Tardino, V.M. (2006). Understanding global transition assessments. *Quality of Life Research*, 15, 995-1004.
- Yalom, I.D. & Lieberman, M.A. (1991). Bereavement and heightend existential awareness. *Psychiatry*, 54, 334-345.
- Yardley, L. & Dibb, B. (2007). Assessing subjective change in chronic illness: An examination of response-shift in health-related and goal-oriented subjective status. *Psychology and Health*, 22, 813-828.
- Zielke, M. & Kopf-Mehnert, C. (1978). *Veränderungsfragebogen des Erlebens und Verhaltens*. Weinheim: Beltz.
- Zielke, M. (1979). *Die Kieler Änderungs-Sensitive Symptomliste (KASSL)*. Weinheim: Beltz.
- Zmud, R.W. & Armenakis, A.A. (1977). Understanding the measurement of change. *Academy of Management Review*, 7, 661-669.
- Zoellner, T. & Maercker, A. (2006). Posttraumatic growth in clinical psychology – a critical review and introduction of a two component model. *Clinical Psychology Review*, 26, 626-653.
- Zwingmann, C. (2003). Zielorientierte Ergebnismessung (ZOE) mit dem IRES-Patientenfragebogen: Eine kritische Zwischenbilanz. *Die Rehabilitation*, 42, 226-235.



# Tabellenverzeichnis

Tabelle 1:	Generische Instrumente zur Erfassung der GLQ bei Krebskranken .....	23
Tabelle 2:	Krebsspezifische Instrumente zur Erfassung der GLQ bei Krebskranken .....	24
Tabelle 3:	Krankheitsspezifische Instrumente zur Erfassung der GLQ bei PCa-Patienten .....	25
Tabelle 4:	Beeinträchtigungen der Gesundheit bei Prostatakrebspatienten .....	41
Tabelle 5:	Rehabilitationsziele bei Prostatakrebspatienten (Heim & Schwerte 2006).....	42
Tabelle 6:	Formen des Response-Shift nach Howard und Kollegen (1979) .....	46
Tabelle 7:	Weitere potenzielle Formen des Response-Shift .....	47
Tabelle 8:	Formen der Veränderung nach Golembiewski und Kollegen (1976).....	48
Tabelle 9:	Formen des Response-Shift nach Sprangers und Schwartz (2000).....	49
Tabelle 10:	Hypothesen der Testung auf Invarianz von Messungen (nach Vandenberg & Lance 2000) .....	84
Tabelle 11:	Hypothesen zur uniformen Rekalibrierung .....	132
Tabelle 12:	Hypothesen zur non-uniformen Rekalibrierung .....	132
Tabelle 13:	Hypothesen zur Repriorisierung .....	133
Tabelle 14:	Hypothesen zur Neukonzeptualisierung .....	133
Tabelle 15:	Hypothesen zur Veränderung der Globalen Lebensqualität und Response-Shift.....	136
Tabelle 16:	Allgemeine Angaben zu den Ausgangsdaten .....	146
Tabelle 17:	Stichprobenszusammensetzung.....	148
Tabelle 18:	Mittelwerte und Standardabweichungen (in Klammern) von Alter und Lebensqualitätsskalen – Modellableitung .....	149
Tabelle 19:	Soziodemographische Angaben zur Gesamtstichprobe (N = 212).....	150
Tabelle 20:	Angaben zu den Skalen des EORTC-QLQ-C30 und verschiedenen Cronbachs Alpha-Werten .....	153
Tabelle 21:	Schiefe und Kurtosis in den zentralen Modellvariablen für die beiden Stichproben .....	157
Tabelle 22:	Cronbachs Alpha für die Indikatoren der GLQ des EORTC-QLQ-C30 .....	158
Tabelle 23:	Cronbachs Alpha für weitere Skalen .....	159
Tabelle 24:	Restriktionen zur Testung von Parameterinvarianz zu zwei Messzeitpunkten .....	161
Tabelle 25:	Angaben zur Modellbewertung mit Wertebereichen für Modellpassungen...	165
Tabelle 26:	Bewertung des Modells zur gesundheitsbezogenen Lebensqualität bei 185 Brustkrebspatientinnen und Prostatakrebspatienten .....	167

Tabelle 27:	Bewertung des Modells zur gesundheitsbezogenen Lebensqualität bei 185 Brustkrebspatientinnen und Prostatakrebspatienten auf Itemebene .....	168
Tabelle 28:	Bewertung des Basismodells zur GLQ bei 212 PCa-Patienten zu beiden MZPen .....	172
Tabelle 29:	Schätzung der Response-Shift-Parameter im Basismodell (ohne Restriktionen).....	172
Tabelle 30:	Restriktionen zur Testung von Parameterinvarianz zwischen zwei Messzeitpunkten .....	173
Tabelle 31:	Passung des ‚Null-Modells‘ .....	174
Tabelle 32:	Schätzung der Response-Shift-Parameter im vollständig restringierten Modell (‘Null-Modell’) .....	174
Tabelle 33:	Passung des ‚Response-Shift-Modells‘ .....	176
Tabelle 34:	Vergleich zweier verschiedener Response-Shift-Modelle .....	176
Tabelle 35:	Vergleich der geschachtelten Modelle mit dem Response-Shift-Modell.....	178
Tabelle 36:	Restriktionen zur Testung auf Parameterinvarianz zwischen zwei Messzeitpunkten .....	178
Tabelle 37:	Schätzung der Response-Shift-Parameter .....	179
Tabelle 38:	Weitere Parameterschätzungen des ‚Response-Shift-Modells‘ .....	181
Tabelle 39:	Vergleich der geschätzten Modelle .....	182
Tabelle 40:	Effektkalkulation und Response-Shift.....	184
Tabelle 41:	Angaben zur Globalen Lebensqualität (N = 212).....	188
Tabelle 42:	Subgruppenbildung und Globale LQ-Differenz .....	189
Tabelle 43:	Soziodemographische Angaben zur Subgruppenanalyse Globale LQ-Differenz.....	190
Tabelle 44:	Subgruppen Globale LQ-Differenz und Indikatoren.....	192
Tabelle 45:	Modellpassung nach Globale LQ-Differenz.....	194
Tabelle 46:	Vergleich der geschachtelten Modelle mit dem Response-Shift-Modell nach Gruppe Globale LQ-Differenz $\leq M_d$ vs. $>M_d$ .....	195
Tabelle 47:	Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Globale LQ-Differenz.....	196
Tabelle 48:	Vergleich der geschachtelten Modelle nach Gruppe Globale LQ-Differenz $\leq M_d$ vs. $>M_d$ .....	197
Tabelle 49:	Schätzung der Response-Shift-Parameter nach Globale LQ-Differenz .....	198
Tabelle 50:	Globale LQ-Differenz als Einflussfaktor auf Response-Shift-Effekte.....	200
Tabelle 51:	Soziodemographische Angaben zu den jeweiligen Substichproben - Moderatoren.....	204

Tabelle 52:	Angaben zum Alter der jeweiligen Substichproben - Moderatoren .....	205
Tabelle 53:	Dropoutanalyse zu den jeweiligen Substichproben - Moderatoren .....	207
Tabelle 54:	Ausprägung der Moderatorvariablen in den Subgruppen.....	208
Tabelle 55:	Mittelwerte und Standardabweichungen der Indikatorvariablen in den Subgruppen .....	210
Tabelle 56:	Passung des Basismodells bei verschiedenen Substichproben.....	211
Tabelle 57:	Passung des Response-Shift-Modells bei verschiedenen Substichproben .....	212
Tabelle 58:	Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß des Benefit Finding .....	214
Tabelle 59:	Schätzung der Response-Shift-Parameter – Benefit Finding .....	215
Tabelle 60:	Benefit Finding als Einflussfaktor auf Response-Shift .....	217
Tabelle 61:	Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß der Selbstwirksamkeit .....	218
Tabelle 62:	Schätzung der Response-Shift-Parameter - Selbstwirksamkeit.....	219
Tabelle 63:	Selbstwirksamkeit als Einflussfaktor auf Response-Shift .....	220
Tabelle 64:	Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß der positiven Affektivität .....	222
Tabelle 65:	Schätzung der Response-Shift-Parameter – Positive Affektivität .....	223
Tabelle 66:	Positive Affektivität als Einflussfaktor auf Response-Shift .....	224
Tabelle 67:	Gruppenspezifische Response-Shift-Ableitung und Modellpassung nach Ausmaß der negativen Affektivität.....	226
Tabelle 68:	Schätzung der Response-Shift-Parameter – Negative Affektivität .....	227
Tabelle 69:	Negative Affektivität als Einflussfaktor auf Response-Shift .....	228
Tabelle A1:	Angaben zur Modellbewertung mit Wertebereichen für Modellpassungen...	316
Tabelle A2:	Response-Shift-Modelle mit und ohne GLQ-Varianzrestriktion .....	318
Tabelle A3:	Korrelation nach Pearson zwischen Indikatorausprägung zum MZP 1 und der Prä-Post-Differenz der Indikatorvariablen / Gesamtstichprobe .....	319
Tabelle A4.1:	Korrelation nach Pearson zwischen Indikatorausprägung zum MZP 1 und der Prä-Post-Differenz der Indikatorvariablen / Veränderung der Globalen LQ $\leq$ Md ..	320
Tabelle A4.2:	Korrelation nach Pearson zwischen Indikatorausprägung zum MZP 1 und der Prä-Post-Differenz der Indikatorvariablen / Veränderung der Globalen LQ $>$ Md ..	320
Tabelle A5:	Angaben über lokale Anpassungsgüte der Messmodelle .....	321

## Abbildungsverzeichnis

Abbildung 1:	GLQ nach Boehmer und Luszczynska (2006) gemessen mit dem krebsspezifischen Lebensqualitätsfragebogen EORTC-QLQ-C30 .....	20
Abbildung 2:	Prozessmodell von Response-Shift und Lebensqualität nach Sprangers und Schwartz (1999, Übersetzung von GÜthlin 2004).....	50
Abbildung 3:	Theoretisches Modell der Bewertungs- und Copingprozesse nach Folkman und Greer (2000).....	65
Abbildung 4:	Modell der Anpassung an chronische Krankheiten nach Sharpe und Curran (2006, adaptiert nach Park und Folkman 1997).....	73
Abbildung 5:	Wirkmodell der Veränderung gesundheitsbezogener Lebensqualität und Response-Shift bei PCa-Patienten mit Anschlussheilbehandlung.....	124
Abbildung 6:	Messmodell zur GLQ bei 185 Krebspatientinnen und -patienten .....	166
Abbildung 7:	Messmodell auf Itemebene zur GLQ bei 185 Brustkrebspatientinnen und Prostatakrebspatienten auf Ebene der Einzelitems und latenten Faktoren 1. und 2. Ordnung .....	169
Abbildung 8:	Modellstruktur des Basismodells .....	171
Abbildung 9:	Beziehung zwischen RMSEA-Veränderungen und Parameterfreisetzungen .....	177
Abbildung A1:	Standardisierte Lösung des Zwei-Faktorenmodells der GLQ .....	314
Abbildung A2:	Unstandardisierte Lösung des Zwei-Faktorenmodells der GLQ .....	315
Abbildung A3:	Restriktion der Varianz der GLQ zum MZP 1 und MZP 2 im Response-Shift-Modell.....	317

# Anhangsverzeichnis

## **Anhang A: Ergänzende Modellberechnungen**

- Anhang A1: Zweifaktorenmodell der Gesundheitsbezogenen Lebensqualität mit dem EORTC-QLQ-C30
- Anhang A2: Modelltestung – Restriktion der Varianz der GLQ zum MZP 1 und MZP 2 im Response-Shift-Modell
- Anhang A3: Berechnung des Effekts der statistischen Regression zur Mitte für die Gesamtstichprobe
- Anhang A4: Berechnung des Effekts der statistischen Regression zur Mitte für die Subgruppen ‚Veränderung der Globalen Lebensqualität  $\leq$  Md‘ vs. ‚Veränderung der Globalen Lebensqualität  $\geq$  Md‘
- Anhang A5: Maße der lokalen Anpassungsgüte der Messmodelle

## **Anhang B: Verwendete Fragebögen**

- Anhang B1: European Organization for Research and Treatment of Cancer – Quality of Life Questionnaire – Core 30 / EORTC-QLQ-C30
- Anhang B2: Benfit-Finding Skala
- Anhang B3: Allgemeine Selbstwirksamkeitserwartung
- Anhang B4: Positive and Negative Affect Schedule / PANAS
- Anhang B5: Soziodemographische Daten / Soz-Dat

## **Anhang C: Lebenslauf**

## **Anhang D: Erklärung nach § 8 Abs. 4 der Promotionsordnung der Universität Koblenz-Landau**

## **Anhang A: Ergänzende Berechnungen**

Anhang A1: Zweifaktorenmodell der Gesundheitsbezogenen Lebensqualität mit dem EORTC-QLQ-C30

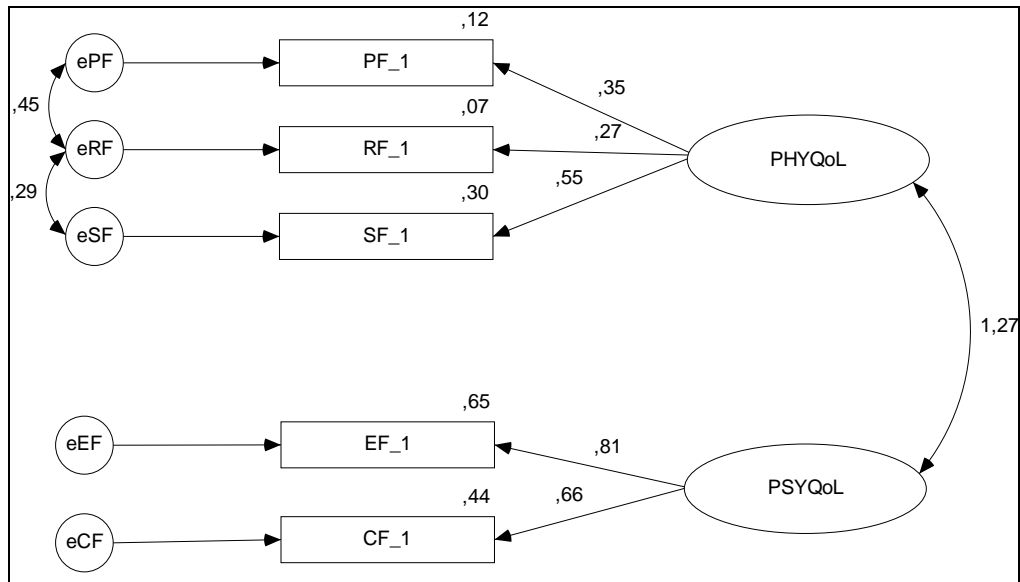


Abbildung A1: Standardisierte Lösung des Zwei-Faktorenmodells der GLQ

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, e = Messfehleranteile der jeweiligen Indikatoren, PHYQoL = Physische Skala der Gesundheitsbezogenen Lebensqualität, PSYQoL = Psychische Skala der Gesundheitsbezogenen Lebensqualität

Anhang A1: Zweifaktorenmodell der Gesundheitsbezogenen Lebensqualität mit dem EORTC-QLQ-C30

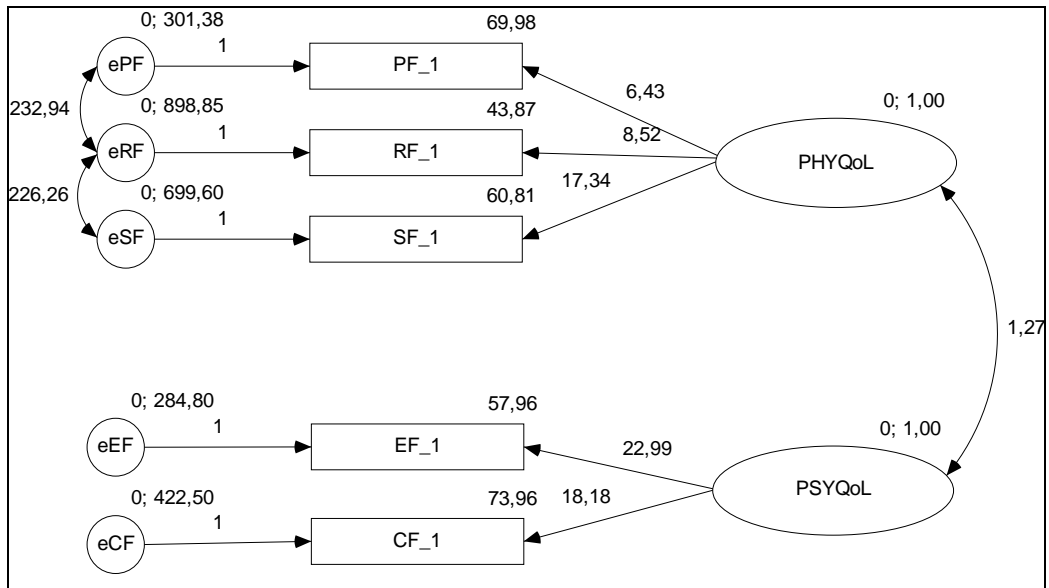


Abbildung A2: Unstandardisierte Lösung des Zwei-Faktorenmodells der GLQ

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, e = Messfehleranteile der jeweiligen Indikatoren, PHYQoL = Physische Skala der Gesundheitsbezogenen Lebensqualität, PSYQoL = Psychische Skala der Gesundheitsbezogenen Lebensqualität



Anhang A1: Zweifaktorenmodell der Gesundheitsbezogenen Lebensqualität mit dem EORTC-QLQ-C30

Tabelle A1: Angaben zur Modellbewertung mit Wertebereichen für Modellpassungen

Gütemaß	Einfaktorenlösung	Zweifaktorenlösung
$\chi^2/\text{df}$	2.14	1.41
RMSEA	0.08	0.05
CFI	0.99	1.00
TLI	0.95	0.98
AIC	40.4	38.8
ECVI	0.22	0.21

Anmerkungen:  $\chi^2/\text{df}$  = Diskrepanzfunktion relativiert an den Freiheitsgraden; RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index, AIC = Akaike Information Criterion, ECVI = Expected Cross Validation Index

Anhang A2: Modelltestung – Restriktion der Varianz der GLQ zum MZP 1 und MZP 2 im Response-Shift-Modell

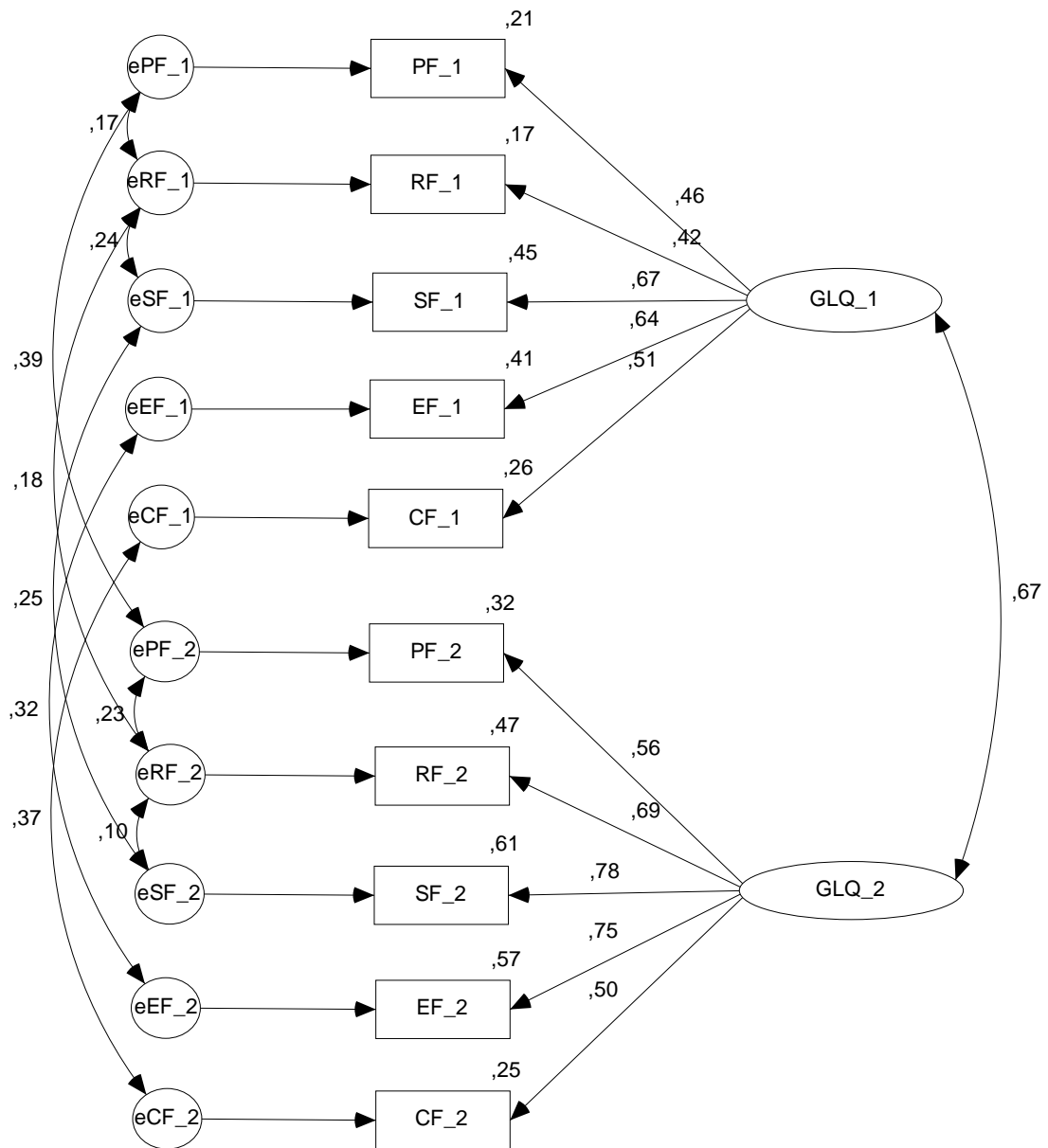


Abbildung A3: Restriktion der Varianz der GLQ zum MZP 1 und MZP 2 im Response-Shift-Modell

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, e = Messfehleranteile der jeweiligen Indikatoren, <sub>1</sub> = MZP 1, <sub>2</sub> = MZP 2, GLQ = Gesundheitsbezogene Lebensqualität

Anhang A2: Modelltestung – Restriktion der Varianz der GLQ zum MZP 1 und MZP 2 im Response-Shift-Modell

Tabelle A2: Response-Shift-Modelle mit und ohne GLQ-Varianzrestriktion

Gütemaß	$\chi^2/df$	RMSEA	CFI	TLI
RS-Modell	1.31	.04	.99	.98
RS-Modell – Restriktion/Varianz_GLQ	1.27	.04	.99	.98

Anmerkungen: RS = Response-Shift,  $\chi^2/df$  = Diskrepanzfunktion relativiert an den Freiheitsgraden, RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index

Anhang A3: Berechnung des Effekts der statistischen Regression zur Mitte für die  
Gesamtstichprobe

Tabelle A3: Korrelation nach Pearson zwischen Indikatorausprägung zum MZP 1 und der  
Prä-Post-Differenz der Indikatorvariablen / Gesamtstichprobe

Indikatoren	Prä-Post-Differenz				
	PF_2 – PF_1	RF_2 – RF_1	EF_2 – EF_1	SF_2 – SF_1	CF_2 – CF_1
PF_1	-0.68**				
RF_1		-0.75**			
EF_1			-0.63**		
SF_1				-0.63**	
CF_1					-0.60**

Anmerkungen: PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, \_1 = MZP 1, \_2 = MZP 2, \*\* p < 0.01

Anhang A4: Berechnung des Effekts der statistischen Regression zur Mitte für die Subgruppen ,Veränderung der Globalen LQ  $\leq$  Md‘ vs. ,Veränderung der Globalen LQ  $\geq$  Md‘

Tabelle A4.1: Korrelation nach Pearson zwischen Indikatorausprägung zum MZP 1 und der Prä-Post-Differenz der Indikatorvariablen / Veränderung der Globalen LQ  $\leq$  Md

Prä-Post-Differenz – Veränderung der Globalen LQ  $\leq$  Md

Indikatoren	PF_2 – PF_1	RF_2 – RF_1	EF_2 – EF_1	SF_2 – SF_1	CF_2 – CF_1
PF_1	-0.60**				
RF_1		-0.71**			
EF_1			-0.56**		
SF_1				-0.51**	
CF_1					-0.51**

Anmerkungen: LQ = Lebensqualität, Md = Median, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, \_1 = MZP 1, \_2 = MZP 2, \*\* p < 0.01

Tabelle A4.2: Korrelation nach Pearson zwischen Indikatorausprägung zum MZP 1 und der Prä-Post-Differenz der Indikatorvariablen / Veränderung der Globalen LQ  $>$  Md

Prä-Post-Differenz – Veränderung der Globalen LQ  $\geq$  Md

Indikatoren	PF_2 – PF_1	RF_2 – RF_1	EF_2 – EF_1	SF_2 – SF_1	CF_2 – CF_1
PF_1	-0.74**				
RF_1		-0.76**			
EF_1			-0.66**		
SF_1				-0.71**	
CF_1					-0.62**

Anmerkungen: LQ = Lebensqualität, Md = Median, PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, \_1 = MZP 1, \_2 = MZP 2, \*\* p < 0.01

## Anhang A5: Maße der lokalen Anpassungsgüte der Messmodelle

Tabelle A5: Angaben über lokale Anpassungsgüte der Messmodelle

MLA	Modelle				
	Modell- ableitung	Basismodell		Response-Shift-Modell	
		MZP 1	MZP 2	MZP 1	MZP 2
DEV	0,36	0,30	0,44	0,31	0,44
Faktorreliabilität <sup>1</sup>	0,72	0,68	0,79	0,68	0,79
quadrierte FL					
PF	0,16	0,21	0,32	0,20	0,32
RF	0,10	0,18	0,47	0,23	0,45
EF	0,69	0,42	0,57	0,42	0,56
SF	0,45	0,45	0,61	0,46	0,61
CF	0,44	0,26	0,25	0,23	0,27

Anmerkungen: MLA = Maß der lokalen Anpassungsgüte, DEV = durchschnittliche erklärte Varianz (Werte > 0,5 gelten als angemessen), <sup>1</sup> = Werte > 0,6 gelten als angemessen, FL = Faktorladung (Werte > 0,4 gelten als angemessen), PF = Physische Funktionsfähigkeit, RF = Rollenfunktionsfähigkeit, EF = Emotionale Funktionsfähigkeit, CF = Kognitive Funktionsfähigkeit, SF = Soziale Funktionsfähigkeit, MZP = Messzeitpunkt

## **Anhang B: Verwendete Fragebögen**

Anhang B1: European Organization for Research and Treatment of Cancer – Quality of Life

Questionnaire – Core 30 (EORTC-QLQ-C30)

Anhang B2: Benefit-Finding Skala

Anhang B3: Allgemeine Selbstwirksamkeitserwartung

Anhang B4: Positive and Negative Affect Schedule (PANAS)

Anhang B5: Soziodemographische Daten (Soz-Dat)

Anhang B1: European Organization for Research and Treatment of Cancer – Quality of Life  
Questionnaire – Core 30 (EORTC-QLQ-C30)



Wir sind an einigen Angaben interessiert, die Sie und Ihre Gesundheit betreffen. Bitte beantworten Sie die folgenden Fragen selbst, indem Sie die Zahl ankreuzen, die am besten auf Sie zutrifft. Es gibt keine „richtigen“ und „falschen“ Antworten. Ihre Angaben werden streng vertraulich behandelt.

	Überhaupt			
	nicht	Wenig	Mäßig	Sehr
1. Bereitet es Ihnen Schwierigkeiten, sich körperlich anzustrengen (z.B. eine schwere Einkaufstasche oder einen Koffer zu tragen)?	1	2	3	4
2. Bereitet es Ihnen Schwierigkeiten, einen längeren Spaziergang zu machen?	1	2	3	4
3. Bereitet es Ihnen Schwierigkeiten, eine kurze Strecke außer Haus zu gehen?	1	2	3	4
4. Müssen Sie tagsüber im Bett liegen oder in einem Sessel sitzen?	1	2	3	4
5. Brauchen Sie Hilfe beim Essen, Anziehen, Waschen, Benutzen der Toilette?	1	2	3	4
<b>Während der letzten Woche:</b>	<b>Überhaupt</b>			
	<b>nicht</b>	<b>Wenig</b>	<b>Mäßig</b>	<b>Sehr</b>
6. Waren Sie bei Ihrer Arbeit oder bei anderen tagtäglichen Beschäftigungen eingeschränkt?	1	2	3	4
7. Waren Sie bei Hobbys oder anderen Freizeitbeschäftigungen eingeschränkt?	1	2	3	4
8. Waren Sie kurzatmig?	1	2	3	4
9. Hatten Sie Schmerzen?	1	2	3	4
10. Mussten Sie sich ausruhen?	1	2	3	4
11. Hatten Sie Schlafstörungen?	1	2	3	4
12. Fühlten Sie sich schwach?	1	2	3	4
13. Hatten Sie Appetitmangel?	1	2	3	4
14. War Ihnen übel?	1	2	3	4
15. Haben Sie erbrochen?	1	2	3	4
16. Hatten Sie Verstopfung?	1	2	3	4
17. Hatten Sie Durchfall?	1	2	3	4

	Überhaupt nicht	Wenig	Mäßig	Sehr
18. Waren Sie müde?	1	2	3	4
19. Fühlten Sie sich durch Schmerzen in Ihrem alltäglichen Leben beeinträchtigt?	1	2	3	4
20. Hatten Sie Schwierigkeiten, sich auf etwas zu konzentrieren, z.B. auf das Zeitung lesen oder das Fernsehen?	1	2	3	4
21. Fühlten Sie sich angespannt?	1	2	3	4
22. Haben Sie sich Sorgen gemacht?	1	2	3	4
23. Waren Sie reizbar?	1	2	3	4
24. Fühlten Sie sich niedergeschlagen?	1	2	3	4
25. Hatten Sie Schwierigkeiten, sich an Dinge zu erinnern?	1	2	3	4
26. Hat Ihr körperlicher Zustand oder Ihre medizinische Behandlung Ihr Familienleben beeinträchtigt?	1	2	3	4
27. Hat Ihr körperlicher Zustand oder Ihre medizinische Behandlung Ihr Zusammensein oder Ihre gemeinsamen Unternehmungen mit anderen Menschen beeinträchtigt?	1	2	3	4
28. Hat Ihr körperlicher Zustand oder Ihre medizinische Behandlung für Sie finanzielle Schwierigkeiten mit sich gebracht?	1	2	3	4

Bitte kreuzen Sie bei den folgenden Fragen die Zahl zwischen 1 (sehr schlecht) und 7 (ausgezeichnet) an, die am besten auf Sie zutrifft!

29. Wie würden Sie insgesamt Ihren Gesundheitszustand während der letzten Woche einschätzen?

1	2	3	4	5	6	7
sehr schlecht						ausgezeichnet

30. Wie würden Sie insgesamt Ihre gesamte Lebensqualität während der letzten Woche einschätzen?

1	2	3	4	5	6	7
sehr schlecht						ausgezeichnet

## Anhang B2: Benefit-Finding Skala

**Patienten gewinnen manchmal den Eindruck, dass die Krankheit nicht nur Probleme erzeugt, sondern auch positive Begleiterscheinungen hat. Bitte geben Sie an, in welchem Maß Sie den folgenden Aussagen zustimmen oder sie ablehnen.**

Meine Erkrankung ...	gar nicht	kaum	mittel- mäßig	ziem- lich	sehr
1. ... hat mich gelehrt, mich an Umstände anzupassen, die ich nicht ändern kann.	1	2	3	4	5
2. ... hat mir geholfen, die Dinge so zu nehmen, wie sie sind.	1	2	3	4	5
3. ... hat meine Familie enger zusammengeführt.	1	2	3	4	5
4. ... hat mich gelehrt, dass jeder im Leben eine Bestimmung hat.	1	2	3	4	5
5. ... hat mir bewusst gemacht, wie wichtig es ist, die Zukunft meiner Familie zu planen.	1	2	3	4	5
6. ... hat mich gelehrt, geduldig zu sein.	1	2	3	4	5
7. ... hat zur Entwicklung und Stärkung meiner Persönlichkeit beigetragen.	1	2	3	4	5
8. ... hat mir geholfen zu erkennen, wer meine wahren Freunde sind.	1	2	3	4	5
9. ... hat dazu beigetragen, meinem Leben einen Sinn zu verleihen und andere Schwerpunkte zu setzen.	1	2	3	4	5
10. ... hat mir geholfen, mich auf das Wesentliche zu konzentrieren und meinem Leben einen tieferen Sinn zu verleihen.	1	2	3	4	5

## Anhang B3: Allgemeine Selbstwirksamkeitserwartung

**Im Folgenden finden Sie Aussagen zum Umgang mit Schwierigkeiten und Problemen.  
Bitte kreuzen Sie an, was für Sie zutrifft.**

	stimmt nicht	stimmt kaum	stimmt eher	stimmt genau
Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen.	1	2	3	4
Die Lösung schwieriger Probleme gelingt mir immer, wenn ich mich darum bemühe.	1	2	3	4
Es bereitet mir keine Schwierigkeiten, meine Absichten und Ziele zu verwirklichen.	1	2	3	4
In unerwarteten Situationen weiß ich immer, wie ich mich verhalten soll.	1	2	3	4
Auch bei überraschenden Ereignissen glaube ich, dass ich gut mit ihnen zurechtkommen kann.	1	2	3	4
Schwierigkeiten sehe ich gelassen entgegen, weil ich meinen Fähigkeiten immer vertrauen kann.	1	2	3	4
Was auch immer passiert, ich werde schon klarkommen.	1	2	3	4
Für jedes Problem kann ich eine Lösung finden.	1	2	3	4
Wenn eine neue Sache auf mich zukommt, weiß ich, wie ich damit umgehen kann.	1	2	3	4
Wenn ein Problem auftaucht, habe ich meist mehrere Ideen, wie ich es lösen kann.	1	2	3	4

Anhang B4: Positive and Negative Affect Schedule (PANAS)

Dieser Fragebogen enthält eine Reihe von Wörtern, die unterschiedliche Gefühle und Empfindungen beschreiben. Lesen Sie jedes Wort und tragen Sie dann in die Skala neben jedem Wort die Intensität ein. Sie haben die Möglichkeit, zwischen fünf Abstufungen zu wählen:

1 = gar nicht; 2 = ein bisschen; 3= einigermaßen; 4 = erheblich; 5 = äußerst

Geben Sie bitte an, wie Sie sich im Allgemeinen fühlen.

	gar nicht	ein bisschen	einigermaßen	erheblich	äußerst
aktiv	1	2	3	4	5
bekümmert	1	2	3	4	5
interessiert	1	2	3	4	5
freudig erregt	1	2	3	4	5
verärgert	1	2	3	4	5
stark	1	2	3	4	5
schuldig	1	2	3	4	5
erschrocken	1	2	3	4	5
feindselig	1	2	3	4	5
angeregt	1	2	3	4	5
stolz	1	2	3	4	5
gereizt	1	2	3	4	5
begeistert	1	2	3	4	5
beschämt	1	2	3	4	5
wach	1	2	3	4	5
nervös	1	2	3	4	5
entschlossen	1	2	3	4	5
aufmerksam	1	2	3	4	5
durcheinander	1	2	3	4	5
ängstlich	1	2	3	4	5



## Anhang B5: Soziodemographische Daten (Soz-Dat)

**Zuerst möchten wir Sie bitten, uns einige Informationen zu Ihrer Person zu geben.**

**Wie ist Ihr Familienstand?** ..... ledig      verheiratet      geschieden      verwitwet

Ja      Nein      Wenn ja, wie viele?

**Haben Sie Kinder?** .....

deutsch      andere      Wenn andere, welche Nationalität?

**Welche Nationalität haben Sie?** .....

Partner      Kinder      allein      Sonstige

**Mit wem leben Sie in Ihrem Haushalt?** .....

**Welchen höchsten Schulabschluss haben Sie?**

- Hauptschule / Volksschule
- Realschule / Mittlere Reife
- Polytechnische Oberschule
- Fachhochschulreife
- Abitur / allgemeine Hochschulreife
- anderen Schulabschluss
- keinen Schulabschluss

**Welche Berufsausbildung haben Sie abgeschlossen?**

- Lehre (beruflich-betriebliche Ausbildung)
- Fachschule (Meister-, Technikerschule, Berufs-Fachakademie)
- Fachhochschule, Ingenieurschule
- Universität, Hochschule
- andere Berufsausbildung
- keine Berufsausbildung

**Sind Sie derzeit erwerbstätig?**

- Ja, ganztags
- Ja, halbtags
- Ja, weniger als halbtags
- Nein, Hausfrau / Hausmann
- Nein, in Ausbildung
- Nein, arbeitslos / erwerbslos
- Nein, Erwerbs-, Berufsunfähigkeitsrente
- Nein, Altersrente
- Nein, anderes

**Ist nach Abschluß der Reha noch eine weitere Therapie gegen die Krebserkrankung erfolgt....**

- Nein
- Operation
- Chemotherapie
- Bestrahlung
- Hormontherapie
- anderweitige Therapie
- Krankenhausaufenthalt nicht wegen Krebs

**Gab es ein erneutes Auftreten eines Tumors nach der Reha?**

- Nein
- Ja
- Verdacht auf
- weiß ich nicht

## Anhang C: Lebenslauf

Name	Jelitte
Vorname	Matthias
Geburtsdatum und -ort	28.02.1973 in Bad Driburg
August 1979 bis Juli 1983	Katholische Grundschule Neuenheerse
August 1983 bis Juni 1992	St. Kaspar Gymnasium der Missionare vom Kostbaren Blut, Neuenheerse
Juli 1992 bis September 1993	Zivildienst
Oktober 1993 bis September 1999	Studium der Psychologie an der Ernst-Moritz-Arndt-Universität Greifswald
Oktober 1999 bis Mai 2000	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Entwicklungspsychologie der Lebensspanne/Pädagogische Psychologie der Ernst-Moritz-Arndt-Universität Greifswald
seit Mai 2000	Wissenschaftlicher Mitarbeiter am Institut für Psychotherapie und Medizinische Psychologie der Bayerischen Julius-Maximilians-Universität Würzburg
seit Oktober 2004	Ausbildung zum Psychologischen Psychotherapeuten (Verhaltenstherapie)
seit Mai 2007	Mitarbeiter des Krankenhaus für Psychiatrie, Psychotherapie und Psychosomatische Medizin (KPPPM) Schloss Werneck

## **Anhang D: Erklärung nach § 8 Abs. 4 der Promotionsordnung der Universität Koblenz-Landau**

Hiermit erkläre ich, Matthias Jelite, geb. am 28.02.1973 in Bad Driburg, gemäß § 8 Abs. 4 der Promotionsordnung des Fachbereichs Psychologie der Universität Koblenz-Landau vom 25.07.2007, dass ich,

1. die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel in der Dissertation angegeben habe,
2. die Dissertation oder Teile hiervon noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe,
3. die gleiche oder eine andere Abhandlung nicht bei einer anderen Hochschule eingereicht habe.

Kürnach, den 27.04.2010

Dipl.-Psych. Matthias Jelite