

II METHODISCHES VORGEHEN

Die vorliegende Forschungsarbeit unterteilt sich in zwei Teilstudien: In Studie 1 wurden qualitative Interviews mit einer Stichprobe multimorbid erkrankter älterer Menschen zu ihrer subjektiven Sicht auf Lebensqualität geführt. Aufbauend auf den Ergebnissen dieser Interviewstudie wurde ein Fragebogen zur Lebensqualität multimorbider älterer Menschen entwickelt und in Studie 2 pilotiert. Nachfolgend wird das methodische Vorgehen getrennt für Studie 1 und Studie 2 dargestellt.

2.1 Grundgesamtheit und Stichproben

2.1.1 Grundgesamtheit

Die Arbeit beschäftigt sich mit der Konstruktion eines Messinstruments für ältere, multimorbide, d. h. körperlich mehrfach erkrankte Menschen. Einschlusskriterien sind somit das kalendarische Alter sowie der Krankheitsstatus.

Die Altersgrenze wurde bei mindestens 65 Jahren angesetzt. Zur genaueren Differenzierung möglicher Alterseffekte wird die Population ferner in „junge Alte“ (65 bis 79) und „Hochbetagte“ (80 Jahre und älter) unterteilt (vgl. Abschnitt 1.2.1; Wahl & Rott, 2002).

Aufgrund des hohen diagnostischen Aufwands, zuverlässige Daten zum objektiven Gesundheitsstatus einer Person zu erhalten, ist es vielfach üblich, den Gesundheitsstatus über einen Selbstbericht zu erheben (de Groot, Beckerman, Lankhorst & Bouter, 2003; Katz, Chang, Sangha, Fossel & Bates, 1996; Steinhagen-Thiessen & Borchelt, 1996). Angesichts des häufigen Nicht-Berichtens (*underreporting*) und der subjektiven Unterschätzung von Schwere und Behandlungsbedürftigkeit von Erkrankungen wurde in der vorliegenden Arbeit eine relativ niedrige Schwelle subjektiv berichteter Erkrankungen gewählt: Eine Person gilt als multimorbide, wenn sie nach eigenem Bekunden unter mindestens vier körperlichen Erkrankungen leidet, die behandelt werden oder deren Behandlung von ihr selbst gewünscht wird.

Eine kognitive Beeinträchtigung der Studienteilnehmer über normale Altersveränderungen hinaus soll ausgeschlossen werden. Im gegebenen Kontext gilt als kognitiv beeinträchtigt, wer 21 Punkte oder darunter in der deutschen Version des Mini-

Mental-Status-Test erzielt (Kessler, Folstein & Denzler, 1990). Ferner sind Personen die sich im Terminalstadium einer lebensbedrohlichen Erkrankung befinden ausgeschlossen.

2.1.2 Studie 1 (Qualitative Interviews): Teilnehmergewinnung und Stichprobe

Über die Pflegedienstleitungen eines ambulanten Pflegedienstes in Berlin wurden Kurzinformationen über die geplante Studie an die Klientinnen und Klienten des Dienstes weitergeleitet (siehe Anhang I). Die jeweiligen Pflegefachkräfte warben zudem bei ihren Klienten für die Teilnahme an der Studie. Um einen schnellen Kontakt zu ermöglichen, fragte die Pflegedienstleitung bei interessierten Klienten nach, ob sie Namen und Telefonnummer zur Terminvereinbarung an den Leiter der Studie weitergeben dürften. Das Alter (1. Einschlusskriterium) und die Anzahl der Erkrankungen (2. Einschlusskriterium) wurden telefonisch erfragt. Am Ende der tatsächlichen Datenerhebung wurde mit den Interviewpartnern zur Abklärung des kognitiven Status (Ausschlusskriterium) die deutsche Version der Mini-Mental-Status-Test durchgeführt (Kessler et al., 1990).

2.1.3 Studie 2 (Pilotuntersuchung mit dem FLQM): Teilnehmergewinnung und Stichprobe

In der Pilotphase der Fragebogenentwicklung wurden über drei unterschiedliche Zugänge Teilnehmerinnen und Teilnehmer für die Studie rekrutiert. Bei allen potenziellen Studienteilnehmern wurden zunächst Alter und Morbiditätsstatus (Einschlusskriterien) telefonisch abgeklärt, die Abschätzung des kognitiven Status (Ausschlusskriterium) erfolgte im Anschluss an die Datenerhebung.

Über die Pflegedienstleitungen zweier ambulanter Berliner Pflegedienste wurden schriftliche Kurzinformationen über die Studie an die Klientinnen und Klienten dieser Dienste weitergeleitet (siehe Anhang II). Die Pflegefachkräfte warben zudem bei ihren Klienten für die Teilnahme an der Studie. Um einen schnellen Kontakt zu ermöglichen fragten die Pflegedienstleitungen bei interessierten Klienten nach, ob sie Namen und Telefonnummer zur Terminvereinbarung an den Leiter der Studie weitergeben dürften. Bei Zustimmung wurden diese telefonisch über die Hintergründe und den geplanten Ablauf der Befragung informiert. Bei weiterhin bestehendem Interesse wurden Termine zur Datenerhebung vereinbart.

In einer Praxis für physikalische Therapie wurden von den behandelnden Physiotherapeuten ebenfalls schriftliche Kurzinformationen an Patientinnen und Patienten

über 65 Jahre weitergegeben. Auch hier fand bei Zustimmung der betreffenden Personen eine telefonische Kontaktaufnahme statt.

Schließlich wurde in der Kundenzeitschrift der AOK „Bleibgesund Plus“ (Anonym, 2006) im Rahmen eines Artikels zu Gesundheit und Krankheit im Alter für die Teilnahme an der Pilotuntersuchung geworben (siehe Anhang III). Interessierte Leserinnen und Leser nahmen telefonisch mit dem Studienleiter Kontakt auf. Nach Erläuterungen zu Hintergrund und Ablauf der Befragung und bei gegebener Bereitschaft zur Teilnahme wurden Termine zur Datenerhebung vereinbart.

2.2 Ablauf und Durchführung der Erhebungen

2.2.1 Studie 1 (Qualitative Interviews)

Alle Erhebungen fanden bei den Studienteilnehmern zu Hause statt. Der Interviewer führte nach einer kurzen Vorstellung der Studienziele und des Forschungsprojektes die einzelnen Teile in folgender Reihenfolge mit den Teilnehmerinnen und Teilnehmern durch (die verwendeten Fragebögen finden sich in Anhang IV und IX):

1. Soziodemografischer Fragebogen
2. Fragebogen zum Gesundheitszustand
3. Interview zu subjektiven Vorstellungen zur Lebensqualität
4. Mini-Mental Status Test

2.2.1.1 Soziodemografischer Fragebogen

Der soziodemografische Fragebogen umfasste Selbstangaben zu Alter, Geschlecht, Schulbildung, Familienstand, Anzahl lebender Kinder, Wohnsituation und Pflegestufe.

2.2.1.2 Fragebogen zum Gesundheitszustand

Der Gesundheitsfragebogen umfasste den Selbstbericht über die Anzahl vorhandener Erkrankungen und den Anteil der behandelten und behandlungsbedürftigen Erkrankungen.

Beide Fragebögen wurden jeweils langsam und deutlich vorgelesen, die Antworten trug der Interviewer in die Bögen ein.

2.2.1.3 Interview zu subjektiven Vorstellungen zur Lebensqualität

Die Instruktion und die Fragen des qualitativen Interviews wurden langsam und deutlich vorgelesen. Das gesamte Interview wurde digital (MP3-Spieler *Zen Micro™*) als MP3-Datei aufgezeichnet. Instruktion und Fragen lauteten wie folgt:

„Ich möchte Ihnen nun gerne einige Fragen dazu stellen, was in Ihrem Leben wichtig ist, was Ihr Leben schön macht, aber auch was Ihnen das Leben schwer macht. Denken Sie in Ruhe über jede Frage nach und erzählen Sie mir dann, was Ihnen dazu einfällt. Sie müssen sich nicht beeilen, im Gegenteil – auch wenn Ihnen später noch etwas zu einer Frage einfällt, sagen Sie es einfach. Es geht ja um Ihr eigenes Leben, und die guten und schlechten Seiten Ihres Lebens. Da wissen natürlich Sie selbst am allerbesten die Antworten. Deshalb gibt es auch keine richtigen oder falschen Antworten. Alles was Sie sagen, alles was Ihnen einfällt, ist für mich interessant. Wenn Sie eine Frage nicht auf Anhieb verstehen, macht das nichts – fragen Sie dann bitte einfach nach. Sagen Sie mir bitte auch, falls Sie auf eine Frage nicht antworten möchten. Können wir anfangen?“

- 1. Was kommt Ihnen in den Sinn, wenn Sie den Begriff „Lebensqualität“ hören? Oder anders: Können Sie mir beschreiben, was für Sie Lebensqualität ausmacht?*
- 2. Wenn Sie über ihr jetziges Leben als Ganzes nachdenken, was macht ihr Leben schön? Was trägt zu Ihrer Lebensqualität bei? Sagen Sie einfach alles was Ihnen einfällt!*
- 3. Was macht Ihnen das Leben schwer? Was wirkt sich negativ auf Ihre Lebensqualität aus? Sagen Sie wieder alles was Ihnen dazu einfällt!*
- 4. Wenn Sie nochmals an all die guten und schlechten Dinge in Ihrem Leben denken, die Sie mir gerade geschildert haben: Gibt es da etwas, das für Sie das Allerwichtigste ist?“*

2.2.1.4 Mini-Mental Status Test, MMST

Der Mini-Mental Status Test ist ein kurzer neuropsychologischer Fragebogen zum Demenz-Screening mit 30 Items (Folstein, Folstein & McHugh, 1975; Tombaugh & McIntyre, 1992). Je mehr Punkte eine Person erreicht, desto besser ihr kognitiver Status. Getestet werden zeitliche und örtliche Orientierung, Merkfähigkeit, Aufmerksamkeit und

Rechenfähigkeit, Sprache und konstruktive Praxis (siehe Anhang IX). Als Cut-off-Wert für den Verdacht auf eine kognitive Beeinträchtigung wurden 21 Punkte festgelegt. Verwendet wurde eine deutsche Version des Mini-Mental-Status-Test (Kessler et al., 1990). Da einige der Teilnehmer infolge schmerzhafter Gelenkerkrankungen oder Hemiparese nicht in der Lage waren, einen Stift zu halten bzw. zu schreiben, traten wiederholt Probleme mit den Items Nr. 25, 26, 27, 29 und 30 auf bzw. konnten diese Items aus *körperlichen* Gründen nicht gelöst werden. Nach einer Methode von Borchelt (2005) wurde der tatsächlich erzielte MMST-Wert hochgerechnet, so dass alle Studienteilnehmer eine Punktzahl von 0 bis 30 Punkten erreichen konnten.

Die Teilnehmer waren angehalten, zu jeder Zeit Fragen zu stellen, wenn Ihnen etwas unklar erschien oder sie den Sinn einer Frage nicht verstanden. Nach Beendigung des MMST wurden Sie nochmals aufgefordert alle Fragen zur Studie und den Forschungszielen zu stellen, die noch offen waren.

2.2.2 Studie 2 (Pilotuntersuchung mit dem FLQM)

Die Erhebungen im Rahmen der Pilotstudie fanden in der Regel bei den Teilnehmern zu Hause statt ($n = 39$). Fünf Teilnehmer wurden auf eigenen Wunsch in den Räumlichkeiten der Poliklinik der Charité – Universitätsmedizin Berlin in einem eigens dafür vorgesehenen Raum befragt. Es fand jeweils eine einzige Erhebung statt.

Neben dem eigentlichen Gegenstand der Studie, dem *Fragebogen zur Lebensqualität multimorbider älterer Menschen, FLQM*, wurden den Teilnehmern folgende standardisierten Fragebögen vorgelegt: Eine Einzelfrage zur allgemeinen Lebenszufriedenheit, eine deutsche Version des *Positive And Negative Affect Schedule, PANAS* (Krohne, Egloff, Kohlmann & Tausch, 1996), eine deutsche Version der *Philadelphia Geriatric Centre Morale Scale, PGCMS* (Smith, Fleeson, Geiselman, Settersten & Kunzmann, 1996) sowie der *Fragebogen zum allgemeinen Gesundheitszustand SF-36* (Bullinger & Kirchberger, 1998). Zusätzlich wurden 21 Teilnehmerinnen und Teilnehmer zufällig ausgewählt und in Form eines kurzen begleitenden Meta-Fragebogens zu Aspekten der Verständlichkeit des FLQM befragt. Alle verwendeten Fragebögen finden sich in den Anhängen V bis IX. Die Reihenfolge der Fragebogenvorgabe war wie folgt:

1. Soziodemografischer Fragebogen
2. Fragebogen zum Gesundheitszustand
- 3.a *Frage zur globalen Lebenszufriedenheit* (n = 22 – Erläuterung siehe Text)
- 3.b *Fragebogen zur Lebensqualität multimorbider älterer Menschen, FLQM*
- 4.a *begleitender Meta-Fragebogen zum FLQM* (n = 21 ST)
- 4.b *Frage zur globalen Lebenszufriedenheit* (n = 22 – Erläuterung siehe Text)
5. *Positive And Negative Affect Schedule, PANAS*
6. *Philadelphia Geriatric Centre Morale Scale, PGCMS*
7. *Short Form 36, SF-36*
8. *Mini-Mental-Status-Test, MMST*

2.2.2.1 *Soziodemografischer Fragebogen*

Mit dem soziodemografischen Fragebogen wurden Alter, Geschlecht, Bildungsstand, Familienstand, Anzahl lebender Kinder, Wohnsituation und Pflegestufe erhoben (siehe Anhang V).

2.2.2.2 *Fragebogen zum Gesundheitszustand*

Der Gesundheitsfragebogen umfasste den Selbstbericht über die Anzahl vorhandener Erkrankungen und den Anteil der behandelten und aus eigener Sicht behandlungsbedürftigen Erkrankungen (siehe Anhang V).

Beide Fragebögen (soziodemografischer und Gesundheitsfragebogen) wurden jeweils langsam und deutlich vorgelesen, die Antworten trug der Interviewer in die Bögen ein.

2.2.2.3 *Frage zur globalen Lebenszufriedenheit*

Die Frage zur allgemeinen Lebenszufriedenheit war folgendermaßen formuliert: „*Alles in allem betrachtet – wie zufrieden sind Sie insgesamt mit Ihrem Leben?*“. Die Antwortmöglichkeiten standen den Teilnehmern als bipolare sechsstufige Likert-Skala zur Verfügung: *1 = es gibt praktisch nichts zu verbessern, 2 = sehr zufrieden, 3 = zufrieden, 4 = unzufrieden, 5 = sehr unzufrieden, 6 = es könnte fast nicht schlimmer sein*. Die Antwortmöglichkeiten lagen den Teilnehmern schriftlich vor und wurden gleichzeitig laut vorgelesen (siehe Anhänge V und VI). Der Hälfte der Teilnehmer (n = 22) wurde die Frage

unmittelbar *vor* dem FLQM vorgelegt, der anderen Hälfte ($n = 22$) *nach* dem FLQM oder dem Meta-Fragebogen zum FLQM (falls erhoben). Dadurch sollte für mögliche Priming-Effekte der globalen oder spezifischen Zufriedenheitsurteile kontrolliert werden. Die Zuteilung erfolgte abwechselnd, beginnend mit der Version *vor* dem FLQM. Beide Versionen der Frage unterscheiden sich lediglich durch die Instruktion voneinander: „Ich möchte Ihnen *zunächst* eine ziemlich allgemeine Frage stellen“ *vor* dem FLQM, „Ich möchte Ihnen *nun* eine ziemlich allgemeine Frage stellen“ *nach* dem FLQM bzw. Meta-Fragebogen.

2.2.2.4 Fragebogen zur Lebensqualität multimorbider älterer Menschen, FLQM

Der Aufbau des „Fragebogens zur Lebensqualität multimorbider älterer Menschen“ (FLQM) wird in Abschnitt 1.4 ausführlich erläutert. Im ersten Schritt sollen die Teilnehmer zunächst Bereiche („Dimensionen“) generieren, die in besonderem Maße ihre Lebensqualität und Zufriedenheit beeinflussen. Die spezifischen Zufriedenheiten mit diesen Bereichen werden im zweiten Schritt erfragt, im letzten Schritt werden die Bereiche individuell gewichtet. Die Instruktionen wurden den Teilnehmern schriftlich vorgelegt, gleichzeitig las der Interviewer den Text langsam und deutlich vor. Abweichungen zwischen den Texten für Teilnehmer und Interviewer bestehen lediglich in Handlungsanweisungen in der Version für den Interviewer sowie in abweichenden Zwischenüberschriften (siehe Anhänge V und VI). Gleichzeitig wurde den Teilnehmern ein tabellarisches Formblatt (siehe Anhang VI) vorgelegt, in das den Anweisungen entsprechend der Reihe nach Lebensbereiche, Zufriedenheitsurteile und Gewichtungen eingetragen werden sollten. Den Teilnehmern wurde die Wahl überlassen, ob sie die Tabelle selbst ausfüllen wollten oder ob der Interviewer dies nach ihren Anweisungen tun sollte. Sämtliche Teilnehmer entschieden sich dafür, dass der Interviewer die Tabelle ausfüllen sollte. Die vollständigen Instruktionen finden sich in den Anhängen V und VI.

Für die Bewertungen der spezifischen Zufriedenheiten mit den genannten Lebensbereichen wurden den Teilnehmern eine bipolare („Zufriedenheit – Unzufriedenheit“) und für die Bewertung der spezifischen Bedeutsamkeiten eine unipolare („Ausmaß der Bedeutsamkeit“) Likert-Skala mit je sechs Stufen vorgelegt. Die sechs Stufen der Skala der Zufriedenheitsurteile waren wie folgt bezeichnet: *1 = es gibt praktisch nichts zu verbessern, 2 = sehr zufrieden, 3 = zufrieden, 4 = unzufrieden, 5 = sehr unzufrieden, 6 = es könnte fast nicht schlimmer sein*. Die Stufen der Skala für die

Bedeutsamkeit der Lebensbereiche waren folgendermaßen bezeichnet: 1 = *eigentlich das Allerwichtigste, was es für mich gibt*, 2 = *sehr wichtig*, 3 = *ziemlich wichtig*, 4 = *wichtig*, 5 = *eher unwichtig*, 6 = *hat im Vergleich mit den anderen keine besondere Bedeutung für mich*. Die sechsfache Unterteilung wurde aus zwei Gründen gewählt: Zunächst werden dadurch Tendenzen zur Mitte vermieden, da die Personen gezwungen sind sich zumindest tendenziell für eine Richtung (eher positive oder eher negative Valenz) zu entscheiden (vgl. Streiner & Norman, 2003). Außerdem erinnern die Bewertungsskalen in der sechsstufigen Form an die Schulnotenskala, die den allermeisten Teilnehmern aus ihrer Jugend bekannt sein müsste. Das Risiko, dass dadurch eine schiefe Skala mit einem wahrgenommenen Mittelpunkt zwischen 4 und 5 forciert wird (vgl. (Streiner & Norman, 2003), wurde zu Gunsten der Verständlichkeit der Bewertungsprozesse in Kauf genommen.

Um zu gewährleisten, dass höhere Werte eine bessere Lebensqualität anzeigen werden die von den Teilnehmern genannten Zahlenwerte (spezifische Zufriedenheiten und Gewichte) zur Berechnung des Index' der Lebensqualität umgepolt: Eine vom Teilnehmer genannte „1“ wird als „6“ in die Berechnung aufgenommen, eine „2“ als „5“ usw. Die Ausprägung der subjektiven Lebensqualität einer Person ergibt sich aus der Summe der Produkte ihrer (umgepolten) Bewertungen und deren (umgepolten) Gewichtungen; die resultierende Produktsumme wird ihrerseits durch die Summe der (umgepolten) personenspezifisch genannten Einzelgewichte geteilt. Der endgültige Index ist somit ein individuell gewichteter Mittelwert. Die möglichen Ausprägungen schwanken zwischen 1 als ungünstigstem Wert und 6 als günstigstem Wert für Lebensqualität.

2.2.2.5 *Begleitender Meta-Fragebogen zum FLQM*

In einem kurzen begleitenden Fragebogen wurden $n = 21$ Teilnehmer unmittelbar im Anschluss an den FLQM zu Verständlichkeit, Struktur und allgemeiner Einschätzung des FLQM befragt (vgl. McColl et al., 2003). Die Auswahl der Personen erfolgte über einen Zufalls-Algorithmus der Statistiksoftware SPSS 13.0. Der Fragebogen diente dazu, über die psychometrischen Merkmale des Fragebogens hinaus Hinweise zur Verbesserung von Struktur und Inhalt zu erhalten. Für die Fragen standen jeweils dreistufige Antwortmöglichkeiten zur Verfügung; drei Fragen eröffneten die Möglichkeit einer kurzen freien Antwort. Der vollständige Meta-Fragebogen findet sich in Anhang VII.

2.2.2.6 *Positive And Negative Affect Schedule, PANAS*

Beim PANAS handelt es sich um ein kurzes Instrument zur Erfassung der unabhängigen Dimensionen positiver Affekt und negativer Affekt (Watson et al., 1988). Er besteht aus einer Liste von 20 Adjektiven, die je zehn positive und negative Gefühlszustände beschreiben. Jedes Adjektiv wird vom Teilnehmer auf einer unipolaren Likert-Skala von 1 bis 5 hinsichtlich des von ihm erlebten Ausprägungsgrads des beschriebenen Gefühls bewertet: 1 = *gar nicht*, 2 = *ein bisschen*, 3 = *einigermaßen*, 4 = *erheblich*, 5 = *äußerst*. Getrennt für positiven und negativen Affekt wird jeweils der Mittelwert der vergebenen Bewertungen berechnet. Die verwendete Version des PANAS erfasste die momentane Ausprägung positiver und negativer Gefühle mittels der Leitfrage „*Wie fühlen Sie sich im Moment?*“ (Krohne et al., 1996); siehe Anhänge V und VI). Den Teilnehmern wurden die Adjektive der Reihe nach vorgelesen und sie sollten diese jeweils unmittelbar bewerten.

Niedrige Werte indizieren (bei einer Spannweite von 1 bis 5) eine schwache Ausprägung von positivem bzw. negativem Affekt, hohe Werte eine starke Ausprägung.

2.2.2.7 *Philadelphia Geriatric Centre Morale Scale, PGCMS*

Die PGCMS ist ein Fragebogen zum allgemeinen Wohlbefinden alter Menschen (Lawton, 1975). Die Originalversion umfasst insgesamt 17 Items in Form kurzer Aussagesätze, die den drei Faktoren *Unaufgeregtheit*, *Zufriedenheit mit dem eigenen Alter* und *Lebenszufriedenheit* zugeordnet werden können. Bei dem hier verwendeten Fragebogen handelt es sich um die auf 15 Items gekürzte deutsche Übersetzung von Smith et al. (1996) aus der Berliner Altersstudie (siehe Anhänge V und VI). Die Teilnehmer sollen auf einer Schulnotenskala von 1 (*sehr gut*) bis 5 (*mangelhaft*) bewerten, wie sehr die jeweiligen Aussagen auf sie selbst zutreffen („*Wie sehr trifft folgende Aussage auf Sie zu?*“). Die Auswertung kann getrennt nach den drei Subskalen oder als Gesamtwert erfolgen. Die Aussagen wurden den Teilnehmern der Reihe nach vorgelesen und sie sollten diese unmittelbar bewerten.

Niedrige Werte (jeweils bei einer Spannweite von 1 bis 5) indizieren eine ungünstige Ausprägung auf der jeweiligen Subskala und dem Gesamtscore, hohe Werte eine günstige Ausprägung.

2.2.2.8 Fragebogen zum allgemeinen Gesundheitszustand Short Form 36, SF-36

Der SF-36 ist ein Fragebogen zur subjektiven Einschätzung des eigenen Gesundheitszustandes, der im Rahmen der Medical Outcome Study entwickelt wurde (dt. Bullinger & Kirchberger, 1998). In der vorliegenden Studie wurde der SF-36 in erster Linie aufgrund seines hohen Verbreitungsgrades als Indikator der subjektiven Gesundheit verwendet. Er umfasst insgesamt 36 Items mit unterschiedlichen Antwortformaten (siehe Anhang VIII). Die Selbsteinschätzung, welche der Bogen liefert, kann als Profil getrennt nach den acht Subskalen (körperliche Funktionsfähigkeit, zehn Items; körperliche Rollenfunktion, vier Items; körperliche Schmerzen, zwei Items; allgemeine Gesundheitswahrnehmung, fünf Items; Vitalität, vier Items; soziale Funktionsfähigkeit, zwei Items; emotionale Rollenfunktion, drei Items; psychisches Wohlbefinden, fünf Items) ausgewertet werden. Außerdem ist die Auswertung nach körperlichen und psychischen Facetten der Gesundheit möglich (körperliche bzw. psychische Summenskala).

Die Rohwerte der Subskalen bzw. der Gesamtskala werden linear auf in Skala von 0 bis 100 transformiert. Hohe Werte indizieren eine günstige Ausprägung, d. h. das Fehlen von Symptomen bezogen auf die jeweiligen Skalen und Subskalen, niedrige Werte eine ungünstige Ausprägung, d. h. das Vorliegen von Symptomen oder Beschwerden.

2.2.2.9 Mini-Mental-Status-Test, MMST

Der Mini-Mental-Status-Test ist ein kurzer neuropsychologischer Fragebogen zum Demenz-Screening mit 30 Items (Folstein et al., 1975; Tombaugh & McIntyre, 1992). Je mehr Punkte eine Person erreicht, desto besser ihr kognitiver Status. Getestet werden zeitliche und örtliche Orientierung, Merkfähigkeit, Aufmerksamkeit und Rechenfähigkeit, Sprache und konstruktive Praxis (siehe Anhang IX). Als Cut-off-Wert für den Verdacht auf eine kognitive Beeinträchtigung wurden 21 Punkte festgelegt. Verwendet wurde eine deutsche Version des Mini-Mental-Status-Test (Kessler et al., 1990). Da einige der Teilnehmer infolge schmerzhafter Gelenkerkrankungen oder Hemiparese nicht in der Lage waren, einen Stift zu halten bzw. zu schreiben, traten wiederholt Probleme mit den Items Nr. 25, 26, 27, 29 und 30 auf bzw. konnten diese Items aus *körperlichen* Gründen nicht gelöst werden. Nach einer Methode von Borchelt (2005) wurde der tatsächlich erzielte MMST-Wert hochgerechnet, so dass alle Studienteilnehmer eine Punktzahl von 0 bis 30 Punkten erreichen konnten.

Die Teilnehmer waren angehalten, zu jeder Zeit Fragen zu stellen, wenn Ihnen etwas unklar erschien oder sie den Sinn einer Frage nicht verstanden. Nach Beendigung des MMST wurden Sie nochmals aufgefordert alle Fragen zur Studie und den Forschungszielen zu stellen, die noch offen waren.

2.3 Auswertung

Sämtliche quantitativen statistischen Auswertungen wurden mit SPSS Version 13.0 für Windows durchgeführt. Qualitative Analysen wurden unter zu Hilfenahme von SPSS Version 13.0 per Hand durchgeführt.

2.3.1 Studie 1 (Qualitative Interviews)

2.3.1.1 Charakteristika der Stichprobe

Der soziodemografische Fragebogen wurde ausschließlich deskriptiv ausgewertet. Der Fragebogen zum Gesundheitszustand sowie der MMST dienten der Feststellung von Ein- und Ausschlusskriterien und wurden in der Auswertung nicht weiter berücksichtigt.

2.3.1.2 Auswertung der Interviewdaten – Induktive Kategorienbildung

Für die Auswertung der Interviews wurde in Anlehnung an Mayring (Mayring, 2003) der Zugang induktiver Kategorienbildung gewählt. Bei diesem Vorgehen werden aus transkribierten Interviewtexten systematisch für die Fragestellung relevante Bedeutungseinheiten extrahiert und in mehreren Schritten zu Kategorien zusammengefasst. In Abhängigkeit von der Fragestellung wird mit dieser Methode die inhaltliche Systematik von Texten direkt aus den Inhalten dieser Texte abgeleitet. Im Allgemeinen führen mindestens zwei unabhängige Kodierer die induktive Kategorisierung durch, um die Reliabilität bzw. Objektivität der abgeleiteten Kategorien zu gewährleisten.

Die Kategorienbildung erfolgt in klar gegliederten Teilschritten (Flick, 2002; Mayring, 2003): Zunächst ist es notwendig, eine Fragestellung bzw. ein Thema zu explizieren. Anschließend werden die in Hinblick auf die Fragestellung bedeutsamen Sätze innerhalb des Textes auf ihre Kernaussage reduziert bzw. bei mehreren Kernaussagen in einem Satz zunächst in Teilsätze aufgeteilt. Es werden dann Stichworte zu jeder Kernaussage gebildet, die anschließend unter inhaltliche Kategorien gruppiert werden. Diese können ggf. weiter zu übergeordneten Hauptkategorien zusammengefasst werden.

Nach Kodierung von etwa 25% der Interviews sollte zwischen den Kodierern ein erstes Abgleichen der Kategorienbezeichnungen stattfinden. Abschließend sollte die Reliabilität des Kategoriensystems über eine Bestimmung der Inter-Coder-Reliabilität überprüft werden, für die verschiedene Methoden zur Verfügung stehen.

Diesen Richtlinien entsprechend wurden die Tonaufzeichnungen der qualitativen Interviews zunächst transkribiert (siehe Anhang X). Wurde über längere Strecken des Interviews eindeutig *ohne* direkten Bezug auf die Fragestellungen gesprochen, so wurden diese Teile in der Transkription ausgespart und die jeweilige Stelle im Transkript kenntlich gemacht. Pausen, Intonation, nonverbale Äußerungen usw. wurden *nicht* in das Transkript aufgenommen, da die Auswertung ausschließlich auf der Ebene inhaltlicher sprachlicher Äußerungen stattfand.

Zwei unabhängige Psychologen (im Folgenden *MH* und *UK*), beide mit der Kategorisierung von Interviewdaten vertraut, werteten die Interviews entlang der oben dargelegten Teilschritte aus. Ziel der Auswertung war es, aus den Äußerungen der Teilnehmer induktiv ein Kategoriensystem abzuleiten, welches aus ihrer Sicht für die eigene Lebensqualität relevante Themen oder Bereiche des Lebens abbildet. Die Fragestellung an der sich die Kodierer bei der Kategorienbildung orientieren sollten lautete: *Welche Themen bzw. Bereiche (z. B. Sachverhalte, Kontexte, Gegenstände oder Personen) besitzen nach eigener Aussage eine positive oder negative Bedeutung für die Lebensqualität der Studienteilnehmer?* Die Kategorien sollten möglichst konkret, jedoch *nicht idiografisch* sein, um auf einer mittleren Abstraktionsebene einen Überblick über die inhaltliche Breite des Konstrukts Lebensqualität aus der Perspektive multimorbider älterer Menschen zu gewinnen. Als Analyseeinheiten galten Sätze, Satzteile und ggf. einzelne Wörter, mit denen der Studienteilnehmer auf eine Bedeutung des genannten Sachverhalts, Gegenstandes, der Personen usw. für die eigene Lebensqualität hinwies.

Nach Abschluss der Kategorisierungen wurde die Inter-Coder-Reliabilität bestimmt. Es wurde überprüft, in wie weit die beiden Kodierer aus den einzelnen Interviews übereinstimmende Kategorien ableiteten, d. h. wie groß die Übereinstimmung der Zuordnung der abgeleiteten Kategorien zu den einzelnen Studienteilnehmern ist. Es wurde *nicht* überprüft ob innerhalb der einzelnen Interviews je Analyseeinheit (Satz, Wort) die gleiche Zuordnung getroffen wurde. Ebenfalls wurde keine Evaluation der Übereinstimmungen der *Häufigkeit der Nennung* von Kategorien je Interview vorgenommen.

Die Inter-Coder-Reliabilität wurde über den M-Koeffizienten, ein Maß für die Ähnlichkeit zwischen Objekten mit binärer Variablenstruktur bestimmt (vgl. Gower, 1998). Dazu wurden die beiden unabhängigen Kodierer als Vergleichsobjekte gesetzt, ihre jeweiligen Urteile bezüglich einer Kategorie als „in den Äußerungen des jeweiligen Teilnehmers vorkommend“ oder „nicht vorkommend“ als binäre Variablenausprägungen. Entsprechend gibt es vier unterschiedliche mögliche Kombinationen von Urteilen je Kategorie: Beide Kodierer urteilen positiv („Kategorie kommt im Interview mit dem Teilnehmer vor“), beide urteilen negativ („Kategorie kommt im Interview mit dem Teilnehmer nicht vor“), Kodierer MH urteilt positiv und Kodierer UK negativ, Kodierer UK urteilt positiv und Kodierer MH negativ. Die Urteile der Kodierer wurden in zwei getrennt Matrizen übertragen; in einer dritten Matrix sind Übereinstimmungen und Unterschiede in den Kategorisierungen gekennzeichnet. Die Matrizen und die Berechnung finden sich in Anhang XI. Der M-Koeffizient berechnet sich nach folgender allgemeinen Formel:

Formel 1 *M-Koeffizient: Ähnlichkeitsmaß zweier Objekte mit binärer Variablenstruktur*

$$M = (a + d)/m$$

mit a = Summe der übereinstimmend positiv gekennzeichneten Kategorien
 d = Summe der übereinstimmend negativ gekennzeichneten Kategorien
 m = Gesamtzahl möglichen Urteile innerhalb des Kategoriensystems

2.3.1.3 Erstellen der Anregungsliste

Um Hinweise auf die externe Validität der induktiv gebildeten Kategorien zu erhalten wurden die Ergebnisse von Studie 1 mit einer Anzahl anderer Studien zu verwandten Fragestellungen „dem Augenschein nach“ verglichen (vgl. Tabelle 4). Abweichungen aufgrund unterschiedlicher Wortwahl bzw. der Übersetzung wurden vernachlässigt. Die Anregungsliste für den FLQM wurde auf Basis des induktiv gebildeten Kategoriensystems und des Literaturvergleichs zusammengestellt. Sie sollte, trotz eines möglichst breiten Spektrums an Bereichen, übersichtlich und relativ kurz sein und die aufgeführten Kategorien verständlich und lebensnah wiedergeben. Sie wurde in einem Expertengespräch mit einem Psychometriker und einem Gerontopsychologen diskutiert und in Abstimmung mit ihnen in ihrer endgültigen Form festgehalten. Es sollten sowohl häufig als auch

weniger häufig genannte Kategorien berücksichtigt werden, um den Studienteilnehmern einen Eindruck von der inhaltlichen Vielfalt des Konstruktes Lebensqualität zu geben und auch Nennungen mit eher persönlichen Inhalten anzuregen.

2.3.2 Studie 2 (Pilotuntersuchung mit dem FLQM)

2.3.2.1 *Charakteristika der Studienstichprobe*

Die Studienstichprobe wird hinsichtlich aller im soziodemografischen und Krankheitsfragebogen erhobenen Variablen beschrieben.

2.3.2.2 *Charakteristika des FLQM*

Merkmale der Verteilung des FLQM-Gesamtscore sowie der Durchführungsdauer und der Anzahl genannter Lebensbereiche werden deskriptiv und grafisch dargestellt. Die Normalverteilung der Messwerte wird mittels eines Kolmogorov-Smirnov-Anpassungstest überprüft.

2.3.2.3 *Charakteristika der übrigen Fragebögen (ALZ, PANAS, PGCMS, SF-36)*

Die Werteverteilungen der verwendeten Fragebögen zu den Variablen „allgemeine Lebenszufriedenheit“, „allgemeines subjektives Wohlbefinden“, „positiver Affekt“, „negativer Affekt“, „subjektive Einschätzung der körperlichen Gesundheit“ und „subjektive Einschätzung der psychischen Gesundheit“ werden deskriptiv und grafisch sowohl hinsichtlich des jeweiligen Gesamtscores als auch ihrer Subskalen dargestellt. Für die körperliche und psychische Summenskala des SF-36 liegen Normdaten der deutschen Bevölkerung vor (Bullinger & Kirchberger, 1998). Abweichungen der Mittelwerte der körperlichen und psychischen Summenskalen des SF-36 von diesen Referenzwerten werden über die zugehörigen z-Werte der Standardnormalverteilung überprüft.

2.3.2.4 *Explorative Validierung des FLQM – externe und interne Validität*

Anhand der neben dem FLQM erhobenen Fragebögen wird eine explorative Konstruktvalidierung des FLQM vorgenommen. Eine Konstruktvalidierung ermöglicht über die Überprüfung von Zusammenhangsannahmen zwischen dem Konstrukt („Lebensqualität“) und externen Variablen eine Abschätzung der Gültigkeit der Messergebnisse (vgl. Lienert & Raatz, 1998; Rost, 2004). Für den FLQM wurden Annahmen über Zusammenhänge mit den Merkmalen „allgemeine Lebenszufriedenheit“,

„allgemeines subjektives Wohlbefinden“, „positiver Affekt“, „negativer Affekt“, „subjektive Einschätzung der körperlichen Gesundheit“ und „subjektive Einschätzung der psychischen Gesundheit“ geprüft.

Über Spearman-Korrelationen werden die Zusammenhänge zwischen dem FLQM-Gesamtscore und der allgemeinen Lebenszufriedenheit (mit möglichen Unterschieden in Abhängigkeit von Erhebungszeitpunkt der ALZ), positivem Affekt (PANAS), negativem Affekt (PANAS), dem PGCMS und seinen Subskalen Unaufgeregtheit, Zufriedenheit mit dem eigenen Alter und Lebenszufriedenheit, der körperlichen Lebensqualität (SF-36), der psychischen Lebensqualität (SF-36) sowie den einzelnen Subskalen des SF-36 überprüft. Korrelationskoeffizienten ab 0,10 werden als schwache Effekte, ab 0,30 als mittlere Effekte und solche ab 0,50 als starke Effekte gewertet (vgl. Bortz & Döring, 2003).

Hinweise auf die interne Validität werden über das Antwortverhalten der Stichprobe oder von Teilstichproben gewonnen (vgl. Rost, 2004). Annahmen über Zusammenhänge zwischen FLQM-Gesamtscore und Alter sowie Anzahl der berichteten Erkrankungen werden anhand von Spearman-Korrelationen getestet. Annahmen über Unterschiede im Gesamtscore nach Altersgruppe, Geschlecht, partnerschaftlicher Situation, Bildungsstand und Vorliegen einer Pflegestufe werden mittels Mann-Whitney-U-Tests geprüft.

2.3.2.5 Inhaltsvalidität des FLQM und querschnittliche Unterschiede subjektiver Konstruktionen von Lebensqualität

Inhaltsvalidität bezeichnet die unter theoretisch-inhaltlichen Gesichtspunkten anzunehmende Gültigkeit eines Fragebogens. Sie wird für den FLQM vornehmlich im Rahmen der Diskussion der Ergebnisse bewertet, da es schwierig ist für diesen bedeutenden Validitätsaspekt Grenzwerte oder messbare Standards festzulegen. Die von den Teilnehmern konkret benannten Lebensbereiche werden mittels unterschiedlicher Ansätze ausgewertet: Zunächst werden im Rahmen einer Häufigkeitsanalyse die *genannten* Bereiche von einem Kodierer qualitativ-inhaltsanalytisch Kategorien zugeordnet, welche im wesentlichen auf dem im Rahmen von Studie 1 entwickelten Kategoriensystem (vgl. Abschnitt 2.3.1.3) basieren, ggf. ergänzt, falls die Entwicklung neuer Kategorien aus den Nennungen der Studienteilnehmer notwendig erschien. Die Häufigkeitsverteilung der genannten Lebensbereiche über das Kategoriensystem wird deskriptiv dargestellt. Auf eine quantitative Auswertung bzw. Verteilungsvergleiche zwischen Subgruppen (z. B.

Geschlechter, Altersgruppen) wurde verzichtet, da die Kategorisierung der Nennungen teilweise nicht trennscharf möglich ist und in den subgruppenspezifischen Häufigkeitsverteilungen mit einer größeren Anzahl sehr kleiner Klassengrößen (< 5) zu rechnen ist. Aus diesen Gründen wird in den meisten Darstellungen nur auf die relativen Häufigkeiten der Bereichsnennungen (*Subgruppengröße n geteilt durch absolute Häufigkeit f der Nennung innerhalb der Subgruppe*) Bezug genommen. Vergleiche finden, wenn überhaupt, nur „dem Augenschein nach“ statt.

Mögliche Zusammenhänge zwischen der Anzahl der individuell genannten Dimensionen und dem Alter sowie der Anzahl der berichteten Erkrankungen werden untersucht (beide Spearman-Korrelation), ebenso Unterschiede in der Anzahl der generierten Lebensbereiche bezüglich Alter, Bildungsstand und Geschlecht (Mann-Whitney-U-Tests).

Des Weiteren soll die Unabhängigkeit der individuellen Bereichsnennungen von der Anregungsliste untersucht werden. In diesem Auswertungsschritt werden die idiografischen Bereichsbezeichnungen der Teilnehmer direkt mit der Anregungsliste verglichen. Kategorielle Nähe wird *bei deutlich anderer Wortwahl bzw. stark idiografisch gefärbten Details* nicht als Übereinstimmung gewertet: „geistige Fähigkeiten ausbilden und entwickeln, d. h. mit interessanten Dingen beschäftigen, Sprache lernen“ (ID 51) würde beispielsweise *nicht* als „eigene geistige Fähigkeiten, z. B. das Gedächtnis“ (Beispiel aus Anregungsliste) gewertet werden. Da die Teilnehmer unterschiedlich viele Bereiche benennen, wird für jeden Teilnehmer anschließend der Quotient aus Übereinstimmungen individueller Nennungen mit der Anregungsliste und Gesamtanzahl der individuellen Nennungen bestimmt. Ein Quotient von 0 steht für völlige Unabhängigkeit von der Anregungsliste (keine Übereinstimmungen), bei einem Wert von 1 besteht völlige Übereinstimmung aller Nennungen des Teilnehmers mit Beispielen aus der Liste. Die Verteilung dieser Verhältniswerte wird für die Gesamtstichprobe sowie getrennt nach Geschlecht, Altersgruppen und Bildungsstand dargestellt. Mögliche Unterschiede der zentralen Tendenz zwischen den Teilstichproben (Geschlecht, Altersgruppe, Bildung) mittels Mann-Whitney-U-Tests überprüft.

Informationen bezüglich der Verständlichkeit, Struktur und allgemeinen Einschätzung des FLQM seitens der Teilnehmer werden anhand des begleitenden Meta-Fragebogens untersucht. Neben Teilstichprobencharakteristika und Antwortverteilungen

werden die Kommentare zu den Fragen 1, 3 und 10 des begleitenden Meta-Fragebogens qualitativ-inhaltlich ausgewertet.

2.3.2.6 *Reproduzierbarkeit der Ergebnisse: Explorativer Längsschnitt*

Eine Reliabilitätsanalyse mit einem *test-retest*-Design ist aufgrund der Fluktuation des gemessenen Merkmals nicht sinnvoll. Eine Bestimmung der Reliabilität mittels einer *split-half*-Methode bzw. eine Überprüfung der inneren Konsistenz scheiden in Anbetracht der großen Heterogenität der (individuell generierten) Items ebenfalls aus. Um dennoch einen ersten Eindruck von der Reproduzierbarkeit der Ergebnisse des FLQM zu erhalten, wurde von einer Teilstichprobe ($n = 10$) der FLQM etwa sechs Monate nach der ersten Erhebung ein zweites Mal bearbeitet. Als externer Index wurde zu Beginn der Sitzung die Frage nach der allgemeinen Lebenszufriedenheit gestellt. Im Anschluss an die Bearbeitung des FLQM wurde danach gefragt, wie gut die Person sich an die erste Sitzung mit dem Fragebogen und die von ihr benannten Lebensbereiche erinnern konnte. Dann wurde eine offene Frage nach Veränderungen und Ereignissen im eigenen Leben oder dem von nahe stehenden Dritten gestellt. Diese Informationen sollen Rückschlüsse auf die Ursachen möglicher Veränderungen in Inhalt oder Ausprägung der individuell benannten Bereiche ermöglichen. Zuletzt wurde nach der Anzahl der aktuell bestehenden behandlungsbedürftigen Erkrankungen gefragt (vgl. Anhänge XII und XIII).

2.3.3 Umgang mit fehlenden Werten

Für die Fragebögen PANAS und PGCMS wurden fehlende Werte auf der Ebene der Subskalen jeweils durch den Mittelwert aller gültigen Werte der jeweiligen Subskala ersetzt. Für den SF-36 wurden fehlende Werte gemäß der Auswertungsvorschriften der Handanweisung berechnet (siehe Bullinger & Kirchberger, 1998, S. 17ff). Der Gesamtscore des MMST wurden bei fehlenden Werten mittels der Methode von Borchelt (2005) geschätzt; dabei wird davon ausgegangen, dass die Person alle fehlenden Items hätte lösen können.