

VII. METHODE

1. Aufbau des Fragebogens

Für die vorliegende Studie wurde ein Fragebogen zusammengestellt, der aus insgesamt drei Bausteinen besteht, auf die im Folgenden detailliert eingegangen wird:

- demographisch-soziologische Datenerhebung
- MOS SF-36 zur Erfassung der gesundheitsbezogenen Lebensqualität
- visuelle Analogskala (VAS) zur Selbsteinschätzung der Lebensqualität

1.1 Demographische Datenerhebung

Vom demographischen Teil eines Fragebogens erwartet man, dass er die gängigen personenbeschreibenden Daten erhebt. Üblicherweise werden Fragen nach Geschlecht, Alter, Familienstand, Staatsangehörigkeit und nach Wochenarbeitsstunden gestellt.

Der demographische Teil des eingesetzten Fragebogens enthält darüber hinaus deutlich detailliertere demographische Items. So wurden beispielsweise die Organisationsform der Praxis (Gemeinschaftspraxis / Einzelpraxis), die Jahre bisheriger ärztlicher Tätigkeit, die Zahl der Patienten und die Scheinzahl, Haustiere, Urlaubstage, Zahl der Freunde, Ernährungs-, Schlaf- und Sportgewohnheiten erfasst. Eine solch facettenreiche soziodemographische Befragung bietet einige Vorteile:

Da die gesundheitsbezogene Lebensqualität von Ärzten bisher nicht sehr breit erforscht wurde, gibt es kaum Literatur zu diesem Thema. Zu thematisch verwandten Inhalten finden sich Arbeiten, sie sind jedoch eher soziologischen oder psychologischen Ursprungs. Wenn sie auch nicht wirklich den Gegenstand dieser Untersuchung treffen, so berühren sie doch verwandte Themenkomplexe wie Burn-out, Stress, Arbeits- oder Lebenszufriedenheit.

Um Gewinn bringende Vergleiche wenigstens mit diesen thematisch nur angenäher-ten Arbeiten ziehen zu können, wurden zahlreiche soziodemographische Variablen in den ersten Baustein des Fragenkataloges mit aufgenommen.

Zudem werden Verhaltensweisen abgefragt, die als Coping-Strategien Einfluss haben können auf die Ergebnisse des SF-36 in seiner Gesamtheit oder auch (nur) auf einzelne Dimensionen dieses Questionnaires. So ist etwa zu vermuten, dass die Tatsache, dass eine Kollegin oder ein Kollege regelmäßig Sport betreibt, Einfluss hat sowohl auf ihre oder seine körperliche Rollenerfüllung als auch auf die Vitalität.

Bei zu gründlicher Befragung stand zu befürchten, dass nur eine geringe Rücklaufquote erzielt werden könnte. Die Antwortbereitschaft wird unter anderem verringert durch Ansprechen von Tabuthemen, oder auch durch viele Seiten umfassende Questionnaires, deren Beantwortung dann ein höheres Maß an Konzentration erfordert und mehr als einige Minuten dauert. Bofinger nennt hier einen empirischen Grenzwert von acht Seiten DIN A 4.²⁸ Positiv beeinflussen hingegen lässt sich die Rücklaufquote durch mehrfache Kontaktaufnahmen in kürzeren Zeitabständen.²⁹ Aus diesem Grunde wurde der ursprünglich sehr umfangreiche Demographieteil gekürzt und das Salant-Dillman-Schema zur Rücklaufmaximierung bei postalischen Befragungen dem Studiendesign der vorliegenden Arbeit angepasst.³⁰

1.2 MOS SF-36 und IQOLA-Gruppe

1.2.1 Historisches

Den zweiten wichtigen Bestandteil des in dieser Arbeit verwandten Instruments stellt der SF-36 dar. Damit wird die gesundheitsbezogene Lebensqualität, ein zentraler Aspekt der vorliegenden Untersuchung, erfasst. In der Medizin stellt die Lebensqualität einen noch recht jungen Parameter dar, der entsprechend noch unter vielfältigen

²⁸ Vgl. Bofinger & Rahlfs 1976, S. 27

²⁹ Vgl. Salant & Dillman 1994, S. 138

³⁰ Siehe Seite, Punkt VII. 5. Salant-Dillman Total Design Methode.

Theorien und Konzepten erfasst wird.³¹ Ich strebe in der vorliegenden Studie nicht eine weitere Definition von Lebensqualität und die entsprechende Entwicklung eines weiteren Instrumentes zu ihrer Erfassung an, sondern benutze dasjenige Konstrukt, das sich in der Vergangenheit vielfältig bewährt hat: Den SF-36 der IQOLA-Gruppe³² (in der Version 1.3, Selbstbeurteilungsbogen, Zeitfenster vier Wochen).

Die IQOLA-Gruppe entstand als Koordinations- und Leitungsorgan im Rahmen der Medical Outcome Study (MOS).³³ Die MOS stellte ein Projekt dar, mit dem seit Beginn der 60er Jahre in den USA die Leistung von Versicherungssystemen geprüft wurde. Aus bereits vorhandenen Instrumenten, die auf den unterschiedlichsten Konzeptionen von Lebensqualität basierten, wurde Mitte der 80er Jahre ein Pool von Fragen gesammelt. Diejenigen Lebensbereiche, zu denen immer wieder Items erschienen, fanden Eingang in den Fragenkatalog des ursprünglichen und umfassenden „Functional Status and Well-Being Questionnaire“. Mit 149 Items für 40 Gesundheitsdimensionen und einer durchschnittlichen Beantwortungsdauer von deutlich mehr als einer Stunde³⁴ war dieses Questionnaire allerdings ausgesprochen unhandlich. Entsprechend gering war die Compliance der Befragten.

Daher wurde 1992 von Aaronson et al. eine Kurzform entwickelt, die Short Form 36 oder SF-36.³⁵ Diese 36 Items beinhaltende Variante des MOS-Fragebogens wurde auf ausgesprochen breiter Basis national wie international untersucht. In über 20.000 Studien mit unzähligen Probanden in Kollektiven unterschiedlichster Zusammensetzung wurde seine testtheoretische Solidität bewiesen, seine „psychometrische Robustheit“³⁶ festgestellt. Der SF-36 kommt zu den gleichen Ergebnissen wie die Vollversion des MOS-Fragebogens.³⁷

³¹ Siehe Seite 1, Punkt VI. 1. Definition der Lebensqualität

³² International Quality of Life Assessment-Gruppe

³³ Vgl. Ware, Gandek, 1998, S. 903-912

³⁴ Vgl. Ware, Snow, Kosinski 1993, S. 2:3-2:5

³⁵ Vgl. Bullinger 1997, S. 76-91 sowie Bullinger & Kirchberger 1998, S. 8

³⁶ Vgl. Bullinger 1996, S. XVII-XXVII; quiz XXVII-XXIX

³⁷ Vgl. Ware & Gandek 1998, S. 903-912

1.2.2 Einsatzmöglichkeiten

Es hat sich gezeigt, dass der ursprünglich als spezifisches Instrument entwickelte SF-36 zuverlässig Verwendung finden kann als generischer Fragebogen. Ursprünglich war der SF-36 entwickelt worden, um an Patienten einer bestimmten Krankheitsgruppe den Erfolg medizinischer Therapien hinsichtlich der gesundheitsbezogenen Lebensqualität zu erfassen. Darüber hinaus wurde in weiteren Untersuchungen nachgewiesen, dass mit dem SF-36 unabhängig von etwaigen Erkrankungen eines jeweils befragten Kollektivs generell die gesundheitsbezogene Lebensqualität sicher erfasst werden kann. Eben wegen seiner hervorragenden und hervorragend abgesicherten psychometrischen Eigenschaften auch als generisches Instrument ist der SF-36 in elf Ländern³⁸ auf Bevölkerungsebene eingesetzt worden zur Erhebung einer Normstichprobe.³⁹

„Die Wahl eines bestimmten Testinstrumentes hängt natürlich ab von der Art der Forschung, dem Forschungsobjekt und den Bedürfnissen des Publikums.“⁴⁰ Da es sich bei den zu befragenden Ärztinnen und Ärzten um potenziell Gesunde handelte, musste ein generischer Fragebogen gewählt werden.

Der Vorteil des SF-36 gegenüber anderen Messinstrumenten liegt meines Erachtens klar auf der Hand: Durch die umfassende Erarbeitung und sorgfältigste Hin- und Rück-Übersetzung ist ein Instrument geschaffen worden, das sowohl testtheoretisch sicher und belastbar ist, als auch eines, das zuverlässig internationale Vergleiche zulässt.⁴¹ Darüber hinaus ist es kurz und präzise.

³⁸ Dänemark, Deutschland, Frankreich, Großbritannien, Italien, Japan, Niederlande, Norwegen, Spanien, Schweden, USA

³⁹ Vgl. Gandek & Ware 1998, S. 953-959

⁴⁰ „The selection of a particular instrument will depend on the nature of the research, the objective of the study and the needs of the audience.“ Berzon, Leplege, Lohr et al. 1997, S. 602

⁴¹ „Der SF-36 ist zweifellos das inzwischen international am häufigsten eingesetzte Instrument zur Messung der gesundheitsbezogenen Lebensqualität geworden. Die deutsche Version wurde (...) erstmalig 1998 im Bundes-Gesundheitssurvey eingesetzt. (...) Infolge seines vielfachen Einsatzes sowie der Übersetzung und Prüfung zahlreicher nationaler Versionen im Rahmen des IQOLA-Projektes <ist der SF-36> das am ehesten international vergleichbare Ergebnisse liefernde generische Instrument geworden.“ Radoschewski, 2000; S. 165-189. Bei den jeweiligen Übersetzungen des SF-36 in die Sprache eines anderen Landes wird auch auf eine kulturelle Äquivalenz der Items geachtet. Vgl. Wagner, Gandek, Aaronson et al. 1998, S. 27. Vgl. auch Hunt 1995, S. 15-24

1.2.3 Aufbau: Dimensionen der LQ

Im System der MOS und damit im SF-36⁴² setzt sich die gesundheitsbezogene Lebensqualität zusammen aus acht Dimensionen oder Skalen, die ihre relevanten Teilaspekte abbilden. Die acht Skalen beschreiben in zwei Teilen die Lebensqualität. Zum einen wird die körperliche Kategorie mit den folgenden Skalen abgebildet:

- Körperliche Funktionsfähigkeit (KÖFU)
- Rollenverhalten wegen körperlicher Funktionsbeeinträchtigung (KÖRO)
- Schmerzen (SCHM)
- Allgemeiner Gesundheitszustand (AGES)

Zum anderen handelt es sich um seelische Kategorien:

- Vitalität / Körperliche Energie (VITA)
- Soziale Funktionsfähigkeit (SOFU)
- Rollenverhalten wegen emotionaler Funktionsbeeinträchtigung (EMRO)
- Psychische Funktionsfähigkeit (PSYC)

KÖFU, die körperliche Funktionsfähigkeit, erfasst dabei körperliche Aktivitäten. KÖRO, die körperliche Rollenfunktion, beschreibt im Unterschied dazu die Beeinflussung täglicher Aktivitäten durch (fehlende) körperliche Gesundheit. Mit SCHM, der Abwesenheit körperlicher Schmerzen, wird der dritte Profilwert bezeichnet; die subjektive Bewertung der eigenen Gesundheit wird mit der Skala AGES (Allgemeine Gesundheitswahrnehmung) erfasst. Die Abwesenheit von Erschöpfung, also die Vitalität, wird mit VITA gemessen; SOFU, die soziale Funktionsfähigkeit, beschreibt den Grad, in dem auch bei körperlicher oder seelischer Beeinträchtigung normale soziale Interaktion, wie Freunde zu besuchen, möglich war. EMRO, die emotionale Rollenerfüllung, spiegelt die Abwesenheit von Beeinträchtigungen durch seelische Probleme bei täglichen Verrichtungen wider und PSYC beschreibt das psychische Wohlbefinden von schwerer Depression bis vollkommenem Glücksgefühl. Die Abbildung 1 auf Seite 14 bereitet das Einfließen der Items in die einzelnen Profilwerte zum besseren Verständnis grafisch auf.

⁴² Die eingesetzte Fassung des SF-36 kann im Anhang auf Seite 132 eingesehen werden.

Abbildung 1: Komposition des SF-36

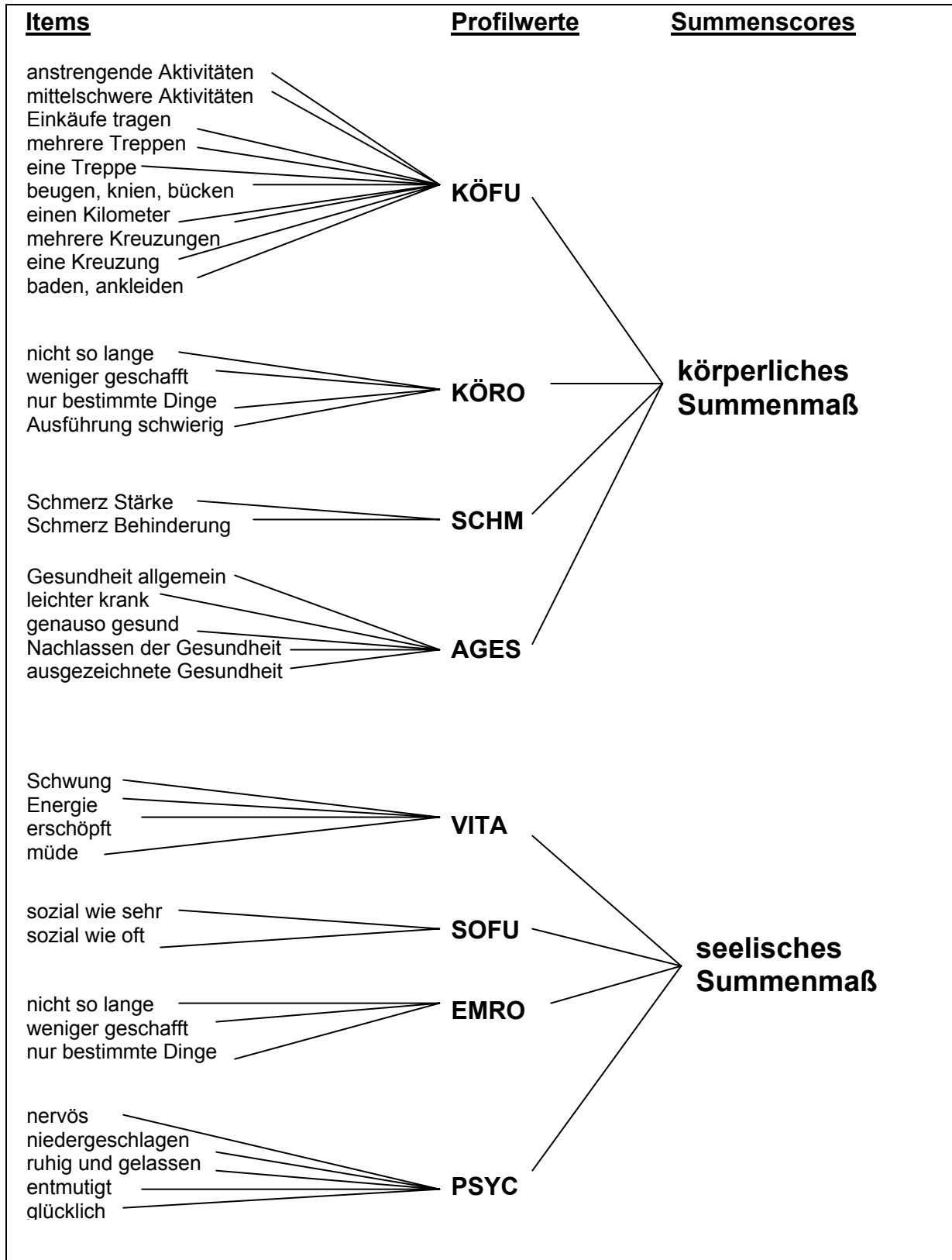


Abbildung nach Ware, Snow, Kosinski et al. 1994

1.2.4 Verarbeitung der Daten

Um aus den Eintragungen auf den jeweiligen Fragebögen diejenigen Werte zu errechnen, die sich im fertigen Diagramm als LQ-Profil ablesen lassen, ist eine handbuchkonforme Aufbereitung der Angaben notwendig. Im Manual finden sich neben den üblichen Anweisungen zum Umgang etwa mit fehlenden Angaben diese Spezifika zur Aufbereitung der Antworten, auf die hier näher eingegangen wird; dazu sind einige Begriffsklärungen notwendig.

Die einzelnen Items des SF-36 geben verschiedene nummerierte Antwortmöglichkeiten vor, die als *Ausgangswerte* oder *precoded values*⁴³ bezeichnet werden.

Beispiel: Ausgangswerte

- Antwort 1: immer
- Antwort 2: meistens
- Antwort 3: manchmal
- Antwort 4: selten
- Antwort 5: nie

Diese Ausgangswerte müssen zu den so genannten *Endwerten* oder *final values* umgerechnet werden. Der dafür stehende Überbegriff der *Rekodierung* umfasst drei unterschiedliche Prozeduren (siehe Abbildung 2), die anhand exemplarisch ausgewählter Items vorgestellt werden. Welche Form der Rekodierung für einen Item durchzuführen ist, schreibt das Handbuch vor.

Abbildung 2: Formen der Rekodierung

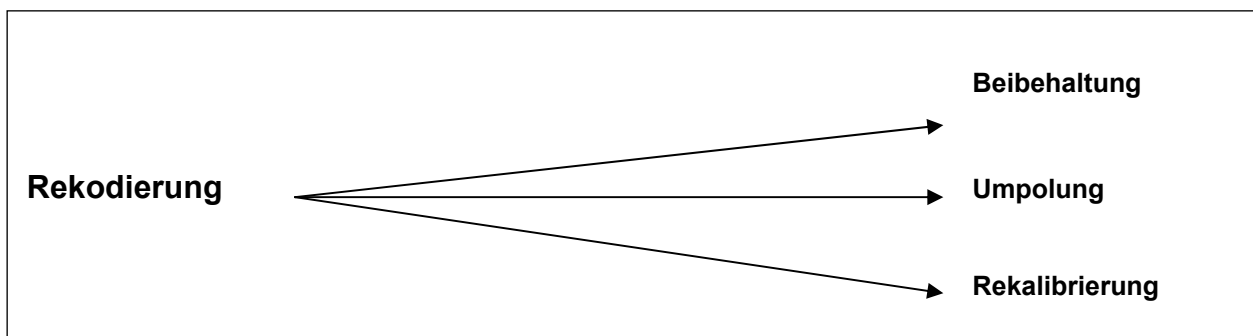


Abbildung nach Ware, Snow, Kosinski et al. 1994

⁴³ Diese Ausgangswerte müssen unterschieden werden von den *Fragennummern* oder *item numbers*, welche in der Auswertung natürlich nicht mathematisch-statistisch bearbeitet werden.

Beispiel einer Beibehaltung:

Frage 3a: „Im Folgenden sind einige Tätigkeiten beschrieben, die Sie vielleicht an einem normalen Tag ausüben. Sind Sie in Ihrem jetzigen Gesundheitszustand bei diesen Tätigkeiten eingeschränkt?“

Tabelle 1: Beibehaltung

Antwortmöglichkeit	Ausgangswert	Endwert nach Rekodierung
Ja, sehr eingeschränkt	1	1
Ja, etwas eingeschränkt	2	2
Nein, überhaupt nicht	3	3

Abbildung nach Ware, Snow, Kosinski et al. 1994

Bei einer Beibehaltung entspricht der Betrag des Endwertes der vorgegebenen Nummer der Antwortmöglichkeit. Die Antwortmöglichkeit mit dem höheren Zahlenwert stellt einen besseren Aspekt der Lebensqualität dar als die Antwortmöglichkeiten mit den niedrigeren Zahlenwerten.

Beispiel einer Umpolung:

Frage 11b: „Inwieweit trifft die folgende Aussage auf Sie zu? Ich bin genauso gesund wie alle anderen.“

Tabelle 2: Umpolung


Antwortmöglichkeit	Ausgangswert	Umpolung	Endwert nach Rekodierung
Trifft ganz zu	1		5
Trifft weitgehend zu	2		4
Weiß nicht	3		3
Trifft weitgehend nicht zu	4		2
Trifft überhaupt nicht zu	5		1

Abbildung nach Ware, Snow, Kosinski et al. 1994

Im Falle einer Umpolung werden also die Zahlenreihen der Ausgangswerte gewissermaßen auf den Kopf gestellt, sodass die Antwort mit der kleinsten präkodierten Nummer nach Umpolung in den Endwert den höchsten Betrag aufweist. Dies ist notwendig, um derjenigen Antwortmöglichkeit einen hohen Zahlenwert zuzuordnen, die zwar als Ausgangswert eine kleine Nummer trägt, aber für die höhere Lebensqualität steht.

Beispiel einer Rekalibrierung:

Frage 7: „Wie stark waren Ihre Schmerzen in den vergangenen vier Wochen?“

Tabelle 3: Rekalibrierung

Antwortmöglichkeit	Ausgangswert	Rekalibrierung	Endwert nach Rekodierung
Keine Schmerzen	1	* <i>u</i>	6,0
Sehr leicht	2	* <i>v</i>	5,4
Leicht	3	* <i>w</i>	4,2
Mäßig	4	* <i>x</i>	3,1
Stark	5	* <i>y</i>	2,2
Sehr stark	6	* <i>z</i>	1,0

Abbildung nach Ware, Snow, Kosinski et al. 1994

Es hat sich gezeigt, dass einige Items rekalibriert werden müssen, um eine lineare Beziehung zwischen dem Endwert und der seiner jeweiligen Profilskala zu erzielen.⁴⁴ Für diese Rekalibrierung gibt es keinen festen Umrechnungsfaktor, sondern es wurde für jeden einzelnen Ausgangswert im Manual ein spezieller, empirisch gefundener Faktor festgelegt.⁴⁵

Der Rekodierung folgt die Berechnung der *Skalenrohwerte*, dafür werden lediglich die Endwerte der zu einer jeweiligen Skala gehörenden Items addiert. Der sich daraus ergebende Wert wird mittels der abgebildeten Formel in einen Wert zwischen 0 und 100 *transformiert*.

Gleichung 1: Transformationsformel

$$\text{Transformierte Skala} = \frac{(\text{tatsächlicher Rohwert} - \text{niedrigst möglicher Rohwert})}{\text{Mögliche Spannweite der Rohwerte}} \times 100$$

Darüber hinaus besteht die auf Abbildung 1 gezeigte Möglichkeit, die acht Gesundheitskonstrukte zu zwei Summenmaßen zusammenzufassen, dem körperlichen und dem seelischen Summenscore.⁴⁶ Durch diese Zusammenfassung werden die Ergeb-

⁴⁴ Bullinger & Kirchberger 1998, S. 16

⁴⁵ Ware, Snow, Kosinski et al. 1994, S. 6:2-6:22

⁴⁶ Vgl. Radoschewski 2000; S. 165-189

nisse vereinfacht dargestellt. Unterschiede in einzelnen Dimensionen werden dadurch allerdings überdeckt. Aus diesem Grunde kamen die Summenscores in der bisherigen Forschung wenig zur Anwendung. Die im Kapitel Ergebnisse vorgestellten Resultate weisen zwar auch in den Summenscores Signifikanzen auf, aber im Rahmen dieser Arbeit wird die präzise Schilderung jeweils im Einzelnen anhand der gesamten acht Dimensionen bevorzugt.

1.3 Visuelle Analogskala (VAS) zur Selbsteinschätzung der Lebensqualität

Wenngleich nach meinem Dafürhalten die durch die IQOLA-Gruppe erstellte Theorie und Erfassung von Lebensqualität das Optimum des zurzeit Möglichen darstellt, so interessierte in der vorliegenden Erhebung doch auch die Frage nach Korrelationen zwischen den Ergebnissen des das interindividuelle Konzept vertretenden SF-36 einerseits und den Ergebnissen der das intraindividuelle Konzept vertretenden VAS.⁴⁷ Die visuelle Analogskala wurde daher in den Fragebogen aufgenommen in Form eines einfachen, 10 cm langen, unskalierten Pfeils, auf dem die befragten Ärztinnen und Ärzte ihre subjektiv empfundene Lebensqualität eintrugen.

2. Auswahl des Befragungskollektivs

Um Stichprobenfehler zu vermeiden und um repräsentatives Datenmaterial zu erhalten, wurde die Untersuchung als Totalerhebung angelegt; es war also beabsichtigt, alle Hausärzte Berlins zu befragen. Als Hausärzte bezeichne ich all diejenigen, die als niedergelassene FachärztInnen für Allgemeinmedizin oder als niedergelassene praktische Ärztinnen oder Ärzte firmierten. Mit einem Verweis auf den Datenschutz lehnte die Berliner Ärztekammer eine Kooperation ab⁴⁸, sodass als Adressquelle auf das Handbuch des Berliner Gesundheitswesens 1999/2000 zurückgegriffen wurde.⁴⁹ Insgesamt wurden auf diese Weise 1709 Personen angeschrieben.

⁴⁷ Siehe Seite 1, Punkt VI. 1. Definition der Lebensqualität

⁴⁸ Bei Stern 1996 jedoch, der ein verwandtes Thema bearbeitete, trug die Ärztekammer Berlin hingegen keine so gearteten Bedenken.

⁴⁹ Stöckel 1999

3. Befragungszeitraum

Durchgeführt wurde die Befragung innerhalb eines Zeitrahmens von zwei Monaten, im September und Oktober 2000; die Erhebung der Daten erfolgte auf dem Postwege.

4. Inhaltliche Überprüfung des Fragebogens

Um die (Unmiss-)Verständlichkeit und grafische Gestaltung des gesamten Fragenkataloges dieser Untersuchung (Demographie, SF-36, VAS) zu überprüfen, wurde eine kleine Pilotstudie an 10 nicht-ärztlichen AkademikerInnen in Berlin durchgeführt. Dabei ergab sich eine durchschnittliche Bearbeitungsdauer der Bögen von ca. 17 Minuten, die Verständlichkeit der Fragen wurde als gut beurteilt, der Inhalt der Fragen als durchaus persönlich, aber nicht indiskret.

5. Salant-Dillman Total Design Methode

5.1 Ziel rücklaufmaximierender Maßnahmen

Die Gültigkeit der im Weiteren getroffenen Aussagen hängt ganz wesentlich ab von der erzielten Response. Üblicherweise wird bei einer standardisierten postalischen Befragung mit einem durchschnittlichen Rücklauf von 20 – 25 % gerechnet. Stern bezeichnet in seiner Studie einen Rücklauf von 35 % als voll befriedigend.⁵⁰ In dieser Größenordnung bewegt sich etwa auch der Rücklauf zweier Studien, die eine ähnliche Thematik wie die vorliegende Arbeit behandeln.⁵¹ Für die vorliegende Untersuchung wurde eine deutlich höhere Response Rate angestrebt.

⁵⁰ Stern 1996, S. 25

⁵¹ Die eine Arbeit untersuchte in einer hessischen Kleinstadt die Lebensqualität von Ärztinnen und Ärzten verschiedener Fachrichtungen (Reimer & Jurkat 1997), in der anderen Erhebung wurde die Arbeitszufriedenheit und Lebenssituation von Kolleginnen und Kollegen verschiedener Fachrichtungen in Dallas County, Texas näher betrachtet (Lewis, Barnhart, Howard et al. 1993)

5.2 Klassische Salant-Dillman Methode

Zur Rücklaufmaximierung bot sich das Verfahren nach Salant/Dillman an, eine optimierte Variante der Dillman Total Design Methode von 1978⁵², die hier zudem den Besonderheiten der vorliegenden Erhebung angepasst wurde. Salant und Dillman haben erkannt, dass eine Antwort der angeschriebenen Person um so wahrscheinlicher ist, je persönlicher das Anschreiben gehalten ist, je häufiger Kontakt mit ihr aufgenommen wird, je größer das Eigeninteresse der Person am Thema ist und je weniger zeitlichen oder finanziellen Aufwand die Antwort verursacht. Um überzeugend zu demonstrieren, wie wichtig den Untersuchern die Antwort der jeweils angeschriebenen, einzelnen Person ist, sollen zwischen den einzelnen Kontaktaufnahmen nur etwa fünf bis sieben Tage liegen.⁵³

5.3 Adaption der Salant-Dillman Methode für die vorliegende Untersuchung

Im vorliegenden Falle wurde insgesamt fünfmal Kontakt zu den Befragten aufgenommen. Der erste Kontakt erfolgte in Form eines kurzen Briefes, in dem zunächst nur ein Fragebogen zum Thema Lebensqualität angekündigt wurde.

Eine Woche später erfolgte der zweite Kontakt: der Fragebogen mit einer ausführlicheren Darstellung des Projekts und einer Erläuterung des Datenschutzverfahrens sowie ein frankierter Rückumschlag. Sobald eine Antwort bei uns eingetroffen war, wurde eine Dankeskarte geschickt; wenn nach weiteren zehn Tagen die Antwort noch ausstand, wurde zum dritten Mal Kontakt aufgenommen, diesmal in Form einer einfachen Postkarte mit der Bitte um Antwort.

Diejenigen, die nach nochmals 10 Tagen nicht geantwortet hatten, erhielten dann als vierten Kontakt ein weiteres Mal den Fragebogen mit Rückumschlag und ein Anschreiben mit der dringenden Bitte um Antwort. Die Tabelle 4 zeigt das Procedere im Überblick.

⁵² Dillman 1978

⁵³ Vgl. Salant & Dillman 1994, S. 138

Tabelle 4: Rücklaufmaximierung nach Salant-Dillman in adaptierter Fassung

#	Zeitpunkt	Zusammensetzung der Postsendung	Inhalt des Anschreibens
1	0	nur Textblatt	„Sie erhalten demnächst einen Fragebogen.“
2	+5d bis +7d	Textblatt, Fragebogen, Rückumschlag, Rückpostkarte	„Hier ist der Fragebogen“ ausführliche Erläuterungen
3	nochmals +5d bis +7d	nur Postkarte	„Bitte antworten Sie uns!“
4	nochmals +5d bis +7d	Textblatt, Fragebogen, Rückumschlag, Rückpostkarte	„Sie haben den Bogen sicher nur verlegt. Hier noch mal.“
5	abschließend	nur Postkarte	„Danke!“

Quelle: eigene Untersuchung

Bei dem Versand der Fragebogen wurde darauf geachtet, an der oberen Grenze der von Salant/Dillman empfohlenen Zeitabstände zu liegen. Das war in diesem speziellen Fall aus mehreren Gründen sinnvoll:

5.3.1 Leistungsfähigkeit der Poststelle

Zum einen musste der Kapazität der Hauspoststelle des Klinikums Benjamin Franklin, an dem die Untersuchung durchgeführt wurde, Rechnung getragen werden. Eine plötzliche Springflut von 2000 Versandstücken liegt jenseits der Mengen, die die Poststelle innerhalb eines Tages bewältigen kann. Es ließ sich daher nicht auf den einzelnen Tag genau festlegen, wann die Sendungen das Klinikum Steglitz verlassen würden.

5.3.2 Uneinheitliche Post-Laufzeiten

Zum zweiten sind die Laufzeiten der Briefe innerhalb der Stadt Berlin uneinheitlich und betragen zwischen einem und drei Tagen. Daher ließ sich nicht mit letzlicher Sicherheit bestimmen, an welchem Tag der jeweilige Empfänger das Anschreiben erhalten würde. Durch die leichte Verlängerung der Intervalle sollte die Wahrscheinlich-

keit erhöht werden, dass der Brief sowohl angekommen als auch gelesen worden war, bevor ein zweites Schreiben zugestellt wurde.

5.3.3 Einschluss eines Wochenendes

Darin liegt auch der dritte Grund für den Versand entlang der Obergrenzen Salant-Dillmans. In ihrem Manual gehen diese beiden Autoren davon aus, dass es sich um eine Umfrage in Privathaushalten handelt. Im Gegensatz zu Privathaushalten erhalten Arztpraxen in der Regel eine große Zahl von Briefen, die gesichtet und bearbeitet werden müssen. Üblicherweise wird ein großer Teil der Post am Wochenende bearbeitet. Daher wurde beim Versand sorgfältig darauf geachtet, dass zwischen den einzelnen Schritten der Kontaktaufnahme jeweils ein Wochenende lag. Schließlich wollte ich zwar die dringende Bitte um Antwort vortragen, nicht aber meine potenziellen Responder belästigen.

6. Datenschutz durch Anonymisierung

Um möglichst breit und detailliert die Lebensumstände der Befragten zu erfassen, stand dem SF-36 ein ausgedehnter Fragenkatalog zu demographischen Variablen, Praxisform und -betrieb voran.⁵⁴ In diesem Teil enthalten waren auch so sensible Fragen wie die nach Alkoholkonsum oder Suizidalität. Es war daher dringend notwendig, die Befragten von der Anonymität ihrer Angaben zu überzeugen.

6.1 Einfluss des Anonymisierungsverfahrens auf die Datenqualität

Üblicherweise werden Fragebögen mit Codenummern versehen, und eine vertrauenswürdige und zur Geheimhaltung verpflichtete Person besitzt den Codeschlüssel, um diejenigen zu benennen, die erneut angeschrieben werden müssen. Dieses Verfahren musste im vorliegenden Fall als unzureichend betrachtet werden, da die Responder keine andere Garantie als mein Wort hatten, dass ihre Daten wirklich geschützt waren. Aus diesem Grunde wurde eine andere Methode angewandt, die

⁵⁴ Siehe Seite 9, Punkt VII. 1.1 Demographische Datenerhebung.

einerseits eben diesen so dringend notwendigen Datenschutz sicherstellte, andererseits aber ermöglichte, ausstehende Antworten gezielt einzufordern.

6.2 Anonymisierung mittels separater Postkarte

Die angeschriebenen Ärztinnen und Ärzte erhielten zusammen mit dem Fragebogen eine an die Allgemeinpraxis im Universitätsklinikum Benjamin Franklin adressierte und frankierte Postkarte, auf der neben dem Text ‚Ich habe geantwortet‘ der Name und die Praxisadresse des Responders vermerkt waren. Diese Postkarte sandten die Befragten dann mit separater Post an das Klinikum zurück. Dadurch kamen die Daten einerseits und die Rückmeldung andererseits getrennt an. Die hohe Zahl täglich eingehender Karten und Bögen stellte sicher, dass etwa aus zeitgleichem Eintreffen der beiden Drucksachen keine Rückschlüsse auf deren Zusammengehörigkeit bzw. einen identischen Absender gezogen werden konnte.⁵⁵

7. Strategien zur Fehlervermeidung

7.1 Adressdaten

Aufgrund fehlenden Zugangs zu den Daten der Berliner Ärztekammer⁵⁶ wurde auf die Anschriften des Handbuchs des Berliner Gesundheitswesens 1999/2000 zurückgegriffen. Es kann billigerd davon ausgegangen werden, dass es sich hierbei um eine aktuelle, hinreichend korrekte Zusammenstellung handelt.

7.2 Computergestützte Datenerfassung

Für die elektronische Erfassung und Bearbeitung der Angaben auf den Fragebögen kam als Statistik- und Analysesoftware die Version 10.0 von SPSS (Statistical Package for Social Sciences) für Windows zum Einsatz.

⁵⁵ Bei akademischer Betrachtung stellen natürlich Pseudoresponder eine potenzielle Fehlerquelle dar, also Menschen, die die Postkarte ‚Ich habe geantwortet‘ zurücksandten, ohne jedoch tatsächlich einen ausgefüllten Fragebogen eingeschickt zu haben. Durch einen schlichten Vergleich der Rücklaufzahlen zeigt sich, dass Pseudoresponse in dieser Arbeit kein Problem darstellt.

⁵⁶ Siehe oben, Seite 18, Punkt VII. 2. Auswahl des Befragungskollektivs.

Nach Eingabe der Daten wurden die gespeicherten Informationen geprüft durch lautes Vorlesen der Angaben auf dem Bogen durch eine Person und die Bestätigung eben dieses Inhalts im elektronischen Speicher durch eine zweite. Insgesamt darf bei dieser strengen Überprüfung davon ausgegangen werden, dass die gespeicherten Daten mit den Angaben auf den Questionnaires weitestgehend übereinstimmen.

7.3 Statistische Auswertung

Um eine statistisch verlässliche Interpretation vornehmen zu können, wurden einige Responder von der Grundauszählung ausgenommen. Dabei handelte es sich zum einen um Fachärzte aus anderen Gebieten der Medizin (n=3), zum anderen um ehemals allgemeinmedizinisch Tätige, die inzwischen nur noch psychotherapeutisch arbeiten (n=4), zum dritten um eine Gruppe, die aus Altersgründen (n=8) sowie aus Krankheitsgründen nicht mehr hausärztlich tätig war (n=4).

Unter den Respondern fand sich eine Gruppe von nur zehn AusländerInnen. Diese Gruppe ist natürlich zu klein, um in irgendeiner Weise zuverlässige Aussagen über sie zu treffen, ich nahm daher auch die AusländerInnen von der Auswertung aus.

Als ebenfalls zu klein stellte sich die Subgruppe der angestellten ÄrztInnen (n=9) und die der rein privatärztlich Tätigen heraus (n=11). Auf eine Auswertung bezüglich angestellter KollegInnen als auch auf den an und für sich hochinteressanten Vergleich der gesundheitsbezogenen Lebensqualität von Kassen- und PrivatärztInnen musste daher aus statistischen Gründen verzichtet werden. Um die gesundheitlich relevante Lebensqualität von im Praxisbereich Angestellten und PrivatärztInnen zu beurteilen, bedarf es eines deutlich größeren Kollektivs.

Nach Bereinigung des Responderkollektivs um die oben genannten Fälle erfolgte die Berechnung der Skalen-Endwerte des SF-36 anhand der im Handbuch vorgeschrie-

benen Syntax.⁵⁷ Für die grafische Präsentation sowohl der SF-36-Resultate als auch der übrigen Daten wurde Word für Windows 2000 verwendet.

Bei normalverteilten quantitativen Daten wurde die Signifikanz der zwischen zwei Gruppen gefundenen Differenzen durch den t-Test für unverbundene Stichproben überprüft. Handelte es sich hingegen um nicht normalverteilte Daten bei unverbundenen Stichproben, so wurde der Mann-Whitney-U-Test verwendet. Bei paarigen Stichproben und fehlender Normalverteilung kam der Wilcoxon-Test zum Einsatz.

Bei differenten Werten von mehr als zwei Gruppen wurde bei Normalverteilung zunächst mit einer Varianzanalyse (=ANOVA) das Vorhandensein von Signifikanzen geprüft. Anschließend klärte der als Post-Hoc-Verfahren eingesetzte Duncan-Test, zwischen welchen der Gruppen genau die Signifikanzen bestanden.

Bei fehlender Normalverteilung wurden die Differenzen zwischen mehr als zwei Gruppen durch den Kruskal-Wallis-Test geprüft. Wenn dabei die Hypothese eines linearen Zusammenhangs aufgestellt werden konnte, kam bei diesen Variablen anschließend der Jonckheere-Terpstra-Test zum Einsatz.

Der Zusammenhang zwischen zwei bivariaten, normalverteilten, quantitativen Variablen wurde anhand des Pearson'schen Korrelationskoeffizienten r ermittelt; erfüllten die zu untersuchenden Variablen diese Kriterien nicht, so wurde stattdessen der Spearman'sche Rang-Korrelationskoeffizient r_s ermittelt. Für die Beurteilung der Zusammenhänge zwischen qualitativen Variablen auf nominalem Skalenniveau wurde der Chi-Quadrat-Test eingesetzt.

Als Signifikanzniveau wurde $p < 0,05$ gewählt. Das heißt, dass unterhalb eines Wertes von $p = 0,05$ nicht von einer Signifikanz gesprochen wird, gekennzeichnet durch ‚nicht signifikant‘ (ns). Ein Wert von $p \leq 0,05$ wird entsprechend signifikant genannt, gekennzeichnet durch (*), ab $p \leq 0,01$ ist von einer sehr signifikanten (**) und ab $p \leq$

⁵⁷ Bullinger & Kirchberger 1998, S. 82-85

0,001 von einer hoch signifikanten Korrelation (***) die Rede. Bei Korrelationen, die die Signifikanz nicht erreichten, wird aufgrund der Größe des befragten Kollektivs bis zu einem Wert von $p=0,1$ von einer nicht signifikanten Tendenz gesprochen (ns).