

## 4. Methode

Die vorliegende Arbeit untersucht Daten der Berliner Altersstudie (BASE; P. Baltes & Mayer, 1999). Die Berliner Altersstudie ist ein interdisziplinäres Projekt mit einem längsschnittlichen Design zur Untersuchung des hohen Alters und des Alterns (70 – 100 Jahre und älter). Im Folgenden werden das allgemeine Erhebungskonzept der Berliner Altersstudie und daran anschließend der BASE Datensatz und die Teilstichprobe von BASE, die dieser Arbeit zugrunde liegt, näher beschrieben. Danach werden die Erhebungsverfahren sowie die mit ihnen erhobenen Konstrukte vorgestellt. Schließlich wird erläutert, wie die erhobenen Daten in dieser Arbeit statistisch ausgewertet worden sind.

### 4.1 Design und Erhebungsvorgehen der Berliner Altersstudie

Die Stichprobe, die dieser Arbeit zugrunde liegt, stammt aus der Berliner Altersstudie (BASE; P. Baltes & Mayer, 1999; Mayer & Baltes, 1996). Die Daten sind den Intensivprotokollen für den ersten (T1), den dritten (T3) und den vierten (T4) Messzeitpunkt der BASE entnommen (für eine Beschreibung dieser Messzeitpunkte siehe auch Smith & Delius, 2003). Die Erhebung der Daten für den ersten Messzeitpunkt fand in der Zeit von Mai 1990 bis Juni 1993, für den dritten Messzeitpunkt von März 1995 bis Mai 1996 und für den vierten Messzeitpunkt von Januar 1997 bis April 1998 statt, so dass für die vorliegende Arbeit auf einen Längsschnittdatensatz von ungefähr sechs Jahren ( $M = 5.51$ ,  $SD = 0.80$ ) zurückgegriffen werden konnte. Der Abstand zwischen dem ersten und dritten Messzeitpunkt beträgt im Mittel 3.75 Jahre ( $Range: 2.37 - 5.18$  Jahre), das Intervall zwischen dem dritten und vierten Messzeitpunkt beträgt im Mittel 1.79 Jahre ( $Range: 0.98 - 2.58$  Jahre). Anhand der Spannweiten ( $Ranges$ ) wird deutlich, dass nicht für alle Personen das gleiche Zeitintervall zwischen den Messzeitpunkten gilt. Differenzen in Zeitintervallen zwischen Messzeitpunkten werden in herkömmlichen Analysen von Messwiederholungen oder Längsschnittdatensätzen häufig nicht berücksichtigt. In der vorliegenden Studie wird das Zeitintervall zwischen den Messungen dagegen für jede Person in die Analyse von Veränderungen miteinbezogen.

Die vorliegende Arbeit konzentriert sich auf ausgewählte Variablen aus dem wesentlich größeren Protokoll der Berliner Altersstudie. Nähere Informationen zum Umfang der Datenerhebung der Berliner Altersstudie sind P. Baltes und Mayer (1999; Mayer & Baltes, 1996) zu entnehmen (zur längsschnittlichen Beschreibung siehe auch Smith & Delius, 2003).

Zu jedem Messzeitpunkt wurde eine Ersterhebung durchgeführt und ein Intensivprotokoll angelegt. Die Ersterhebung ist mit Messverfahren aus den vier Disziplinen, die an der BASE

beteiligt gewesen sind, organisiert worden: Innere Medizin und Geriatrie, Psychiatrie, Psychologie und Soziologie (siehe auch Smith & Helmchen, 1995). Ziel der Ersterhebung war, einen umfassenden Eindruck von den Probanden in ihren psychologischen, soziologischen und medizinischen Charakteristiken zu bekommen. Die Variablen der Ersterhebung decken deshalb ein breites Spektrum ab – von soziodemographischen Daten, subjektivem Wohlbefinden, körperlicher Gesundheit, mentaler Gesundheit bis zu Tests zur kognitiven Funktionsfähigkeit und einem Demenzscreening. Die Konzeption der Ersterhebung war für alle Messzeitpunkte identisch.

Das Intensivprotokoll ermöglicht es, ein detailliertes Profil der Stichprobe zu zeichnen. Es schließt für den ersten Messzeitpunkt 13 Sitzungen ein und zum dritten und vierten Messzeitpunkt je sechs Sitzungen (Smith & Delius, 2003). Die jeweilige Gesamtzahl der Sitzungen ist auf die vier Fachdisziplinen der BASE verteilt worden. In ihnen konnte nach Subdisziplinen unterschieden werden. Die psychologische Einheit umfasst die drei Subdisziplinen Intelligenz und Kognition, soziale Beziehungen, sowie Selbst und Persönlichkeit (siehe auch Smith & Baltes, 1996). Für die vorliegende Arbeit sind Daten aus dem Bereich der psychologischen Subdisziplin „Selbst und Persönlichkeit“ sowie Daten aus der Forschungseinheit der Inneren Medizin herangezogen worden. Zum zweiten Messzeitpunkt (T2) der BASE (Juni 1993 bis Juni 1994) ist im Rahmen einer interdisziplinären Sitzung nur die Ersterhebung durchgeführt worden. Entsprechend liegen für diesen Messzeitpunkt keine Daten zu Persönlichkeitsvariablen vor, die in der vorliegenden Studie im Mittelpunkt des Interesses stehen. Das hat zur Folge, dass in ihr der zweite Messzeitpunkt der BASE nicht berücksichtigt wird.

Die Messzeitpunkte werden in der vorliegenden Studie in Anlehnung an die BASE-Konnotation mit T1 (erster Messzeitpunkt), T3 (dritter Messzeitpunkt) und T4 (vierter Messzeitpunkt) bezeichnet.

Alle Sitzungen wurden in Form von face-to-face Interviews durchgeführt. Sie dauerten jeweils ungefähr eineinhalb Stunden. Soweit möglich, fanden die Sitzungen in der Lebensumwelt der Probanden statt, d. h. bei ihnen zu Hause bzw. in der von ihnen bewohnten Einrichtung. Für die Erhebung wurden Forschungsassistenten und medizinisches Personal, eingeschlossen Internisten und Psychiater, speziell ausgebildet. Jeder Proband wurde über die Zeit der Studie von demselben Forschungsassistenten betreut. Zu weiteren Details des Designs und des Erhebungsvorgehens der Berliner Altersstudie siehe P. Baltes und Mayer (1999).

## 4.2 Stichprobenbeschreibung

Ein Schwerpunkt der BASE liegt auf der Untersuchung interindividueller Unterschiede im hohen Alter und der mit dem Alter verbundenen Abbauprozesse. Im Blick darauf kommt den Variablen Alter und Geschlecht besondere Bedeutung zu. Deshalb wurde die Berliner Altersstudie so angelegt, dass zum ersten Messzeitpunkt pro Altersgruppe (70 - 74, 75 - 79, 80 - 84, 85 - 89, 90 - 94 und 95+ Jahre) und Geschlecht die gleiche Anzahl (jeweils  $n = 43$  in jeder Zelle; 2 Geschlechter x 6 Altersgruppen; zu T1,  $N = 516$ ) von Personen in die Stichprobe einbezogen war. 516 Personen durchliefen das vollständige Untersuchungsprogramm für den ersten Messzeitpunkt. Informationen zur Etablierung der BASE-Kernstichprobe ( $N = 516$ ) sowie zum Stichprobenausfall im hohen Alter sind in den beiden Monographien (P. Baltes & Mayer, 1999; Mayer & Baltes, 1996) enthalten.

Die Beschreibung der Stichproben ergibt: 1.) Es lassen sich Querschnittsstichproben für die genannten drei Messzeitpunkte bilden, d. h. Stichproben mit Personen, die zu T1, T3 und T4 befragt werden konnten (siehe Abbildung 12). 2.) Es ist davon unabhängig möglich, Stichproben zu bilden, in denen Personen nach der Anzahl der Messzeitpunkte, die sie durchlaufen haben, zusammengefasst sind.

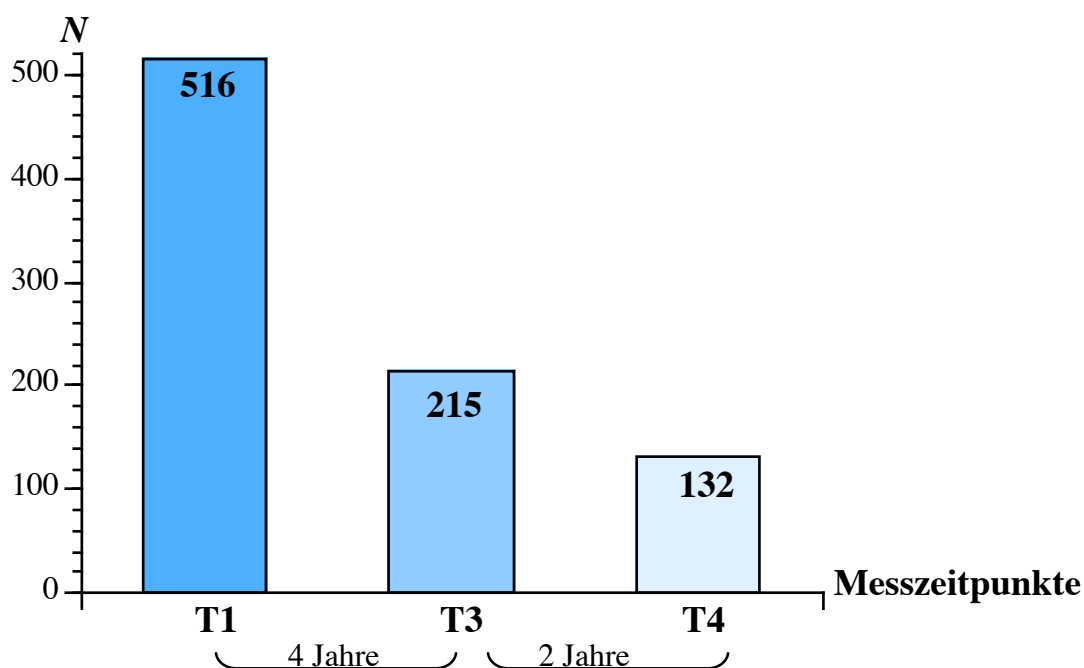


Abbildung 12. Anzahl der Personen zu den einzelnen Messzeitpunkten (T1, T3, T4) des Intensivprotokolls der Berliner Altersstudie

Da in dieser Studie die Methode der Latent Growth Modelle (siehe Abschnitt 4.4.6) Anwendung findet, wurde sich für eine Differenzierung der Stichprobe nach der Anzahl der Messzeitpunkte entschieden. Die Methode der Latent Growth Modelle ermöglicht es, im Unterschied zu herkömmlichen Methoden der Veränderungsmessung zusätzlich zu einem kompletten

Längsschnittdatensatz die Daten von Personen zu analysieren, die zu T1 und T3, nicht aber zu T4 an der Studie beteiligt waren, sowie die Daten von Personen, die nur zum ersten Messzeitpunkt untersucht werden konnten. Entsprechend sind drei Stichproben gebildet worden, die innerhalb der Gesamtheit von 516 Personen, die zum ersten Messzeitpunkt untersucht worden sind, unterschiedliche Gruppen darstellen. Im Folgenden werden sowohl die Längsschnittstichprobe als auch die Stichproben, für die Daten zu einem bzw. zwei Messzeitpunkten vorliegen, näher beschrieben.

#### 4.2.1 Längsschnittstichprobe: Personen, für die Daten zu drei Messzeitpunkten vorliegen

Für die Längsschnittstichprobe liegen Daten zu allen drei Messzeitpunkten (T1, T3 und T4) vor. Die Stichprobe umfasst 132 Personen (72 Frauen und 60 Männer). Das mittlere Alter beträgt in der Stichprobe zum ersten Messzeitpunkt 78 Jahre ( $SD = 5.92$ ), zum vierten Messzeitpunkt 84 Jahre ( $SD = 5.94$ ). Die Probanden lassen sich häufiger den jungen Alten (70- bis 84-Jährige; 64 %) als den alten Alten (85-Jährige und Ältere) zuordnen. Der Anteil der jungen Alten (70 bis 84 Jahre) nimmt vom ersten Messzeitpunkt zum vierten Messzeitpunkt von 88 % auf 64 % ab (siehe Tabelle 10). Da für diese Stichprobe ein kompletter Längsschnitt vorhanden ist, kommt ihr in dieser Studie eine vergleichsweise hervorgehobene Bedeutung zu.

Tabelle 10. Alters- und Geschlechtsverteilung der Längsschnittstichprobe ( $N = 132$ )

		T1	T3	T4
Alter	<i>M</i>	78.3	82.0	83.8
	<i>SD</i>	5.9	6.0	5.9
Frauen	<i>n</i>	72	72	72
	%	54.5	54.5	54.5
Männer	<i>n</i>	60	60	60
	%	45.5	45.5	45.5
junge Alte	<i>n</i>	116	94	84
	%	87.9	71.2	63.6
alte Alte	<i>n</i>	16	38	48
	%	12.1	28.8	36.4

Die 132 Personen der Längsschnittstichprobe stellen 26 % der Intensivstichprobe zu T1 ( $N = 516$ ) dar. In Tabelle 11 sind Ursachen für den Wegfall von Probanden aufgelistet und beschrieben wie sie sich prozentual auf die Ausgangsstichprobe ( $N = 516$ ) verteilen. Die Hälfte der Personen aus der Ausgangsstichprobe von 516 Personen ( $n = 277$ ) ist zwischen dem ersten und vierten Messzeitpunkt verstorben. Zusätzlich konnten 20 % der Ausgangsstichprobe ( $n = 107$ ) aus unterschiedlichen Gründen zu T4 nicht mehr befragt werden. Zu den Gründen zählen Umzug in eine andere Stadt oder Verweigerung der weiteren Teilnahme an der Untersuchung.

Tabelle 11. Gründe für den Stichprobenausfall zwischen dem ersten und vierten Messzeitpunkt

Längsschnittstichprobe	<i>N</i>	%
T4 Intensivprotokoll vorhanden	132	25.6
Gründe für Ausfall:		
a) Verstorben	277	53.7
b) Teilnahme verweigert	63	12.2
c) Verzogen	6	1.2
d) Nur Kurzfragebogen	6	1.2
e) nur T4-Ersterhebung	32	6.2

Anmerkung. Die Prozentangaben beziehen sich auf den Anteil der Gruppen an der Stichprobe zu T1 ( $N=516$ )

#### 4.2.2 Die Stichprobe der Personen, für die Daten zu zwei Messzeitpunkten vorliegen

Die Stichprobe der Personen, für die Daten zu T1 und T3 vorliegen, umfasst 83 Personen. Der Anteil der Männer (54 %) übersteigt geringfügig den der Frauen. Die Hälfte der Mitglieder dieser Stichprobe zählt zu den jungen Alten (70 bis 84 Jahre; 49 %). Das mittlere Alter steigt von 83 Jahren zu T1 auf 87 Jahre zu T3, so dass zum ersten Messzeitpunkt 59 % den jungen Alten zugeordnet werden können, zum dritten Messzeitpunkt hingegen noch die Hälfte (50 %; siehe Tabelle 12).

Tabelle 12. Alters- und Geschlechtsverteilung der Stichprobe mit zwei Messzeitpunkten ( $N = 83$ )

		T1	T3
Alter	<i>M</i>	82.7	86.5
	<i>SD</i>	7.7	7.7
Frauen	<i>n</i>	38	38
	%	45.8	45.8
Männer	<i>n</i>	45	45
	%	54.2	54.2
Junge Alte	<i>n</i>	49	41
	%	59.0	49.4
Alte Alte	<i>n</i>	34	42
	%	41.0	50.6

#### 4.2.3 Die Stichprobe der Personen, für die Daten nur für den ersten Messzeitpunkt vorliegen

Diese Stichprobe umfasst 301 Personen. Auffällig ist, dass in ihr Frauen (49 %) und Männer nahezu gleich vertreten sind, aber wesentlich mehr alte Alte (69 %) als junge Alte. Das mittlere Alter

beträgt 88 Jahre (siehe Tabelle 13). Dies lässt sich mit der Selektivität der Stichproben in Verbindung bringen.

Tabelle 13. *Alters- und Geschlechtsverteilung der Stichprobe mit einem Messzeitpunkt (N = 301)*

		T1
Alter	<i>M</i>	88.4
	<i>SD</i>	8.01
Frauen	<i>n</i>	148
	%	49.2
Männer	<i>n</i>	153
	%	50.8
Junge Alte	<i>n</i>	93
	%	30.9
Alte Alte	<i>n</i>	208
	%	69.1

#### 4.2.4 Zur Selektivität der Stichproben mit einem, zwei oder drei Messzeitpunkten

Für einen Vergleich der drei Stichproben ist festzuhalten, dass Männer und Frauen relativ gleich in den Stichproben vertreten sind. Auffällig ist, dass die Längsschnittstichprobe einen wesentlich größeren Anteil an jungen Alten (70- bis 84-Jährige) umfasst, in der Stichprobe mit nur einem Messzeitpunkt dagegen überwiegen die alten Alten (85-Jährige und Ältere). Viele von den zum ersten Messzeitpunkt besonders alten Alten konnten nur einmal an der Studie teilnehmen. Der Grund dafür liegt zumeist im Ableben der betreffenden Personen zwischen den Messzeitpunkten.

Beim Vergleich des Alters im Mittel für die drei Stichproben zum Zeitpunkt der ersten Erhebung (T1) wird deutlich, dass die Mitglieder der Längsschnittstichprobe mit einem mittleren Alter von 78 Jahren zehn Jahre jünger als diejenigen sind, für die nur zum ersten Messzeitpunkt Daten erfasst werden konnten ( $M = 88$  Jahre,  $t(332.02) = 14.70$ ,  $p < .001$ , zweiseitiger Test). Dies ist als ein Hinweis auf eine positive Selektivität der Längsschnittstichprobe zu werten (siehe auch Abschnitt 4.4.5 zu Selektivitätsberechnung und Abschnitt 5.2 zu den Ergebnissen zur Selektivität).

In Tabelle 14 sind Angaben zum Familienstand, zur Wohnsituation und zum Bildungsniveau der Mitglieder der drei Stichproben zum Zeitpunkt der ersten Untersuchung (T1) abgetragen. Das Bildungsniveau ist über die in Bildungseinrichtungen verbrachte Anzahl von Jahren erfasst worden. In der Längsschnittstichprobe ( $N = 132$ ) sind 33 % verheiratet ( $n = 44$ ), 52 % verwitwet ( $n = 68$ ), 8 % geschieden ( $n = 11$ ) und 7 % ledig ( $n = 9$ ). Von den Mitgliedern dieser Stichprobe leben nur 2 % in einer Einrichtung ( $n = 3$ ). In der Stichprobe der Personen mit zwei Messzeitpunkten ( $N = 83$ ) sind 37 % der Personen ( $n = 31$ ) verheiratet, 45 % ( $n = 37$ ) verwitwet, 10 % ( $n = 8$ ) geschieden und 8 % ( $n = 7$ ) ledig. Die Stichprobe der Personen mit Daten ausschließlich vom ersten

Messzeitpunkt ( $N = 301$ ) umfasst 26 % ( $n = 79$ ) verheiratete Personen, 59 % ( $n = 178$ ) verwitwete, 6 % ( $n = 19$ ) geschiedene und 8 % ( $n = 25$ ) ledige Personen. 21 % ( $n = 63$ ) der Mitglieder dieser Stichprobe leben in einer Institution. Im Vergleich zur Längsschnittstichprobe mit drei Messzeitpunkten zeigt sich, dass ein wesentlich höherer Anteil dieser Stichprobe (mit einem Messzeitpunkt) in Einrichtungen lebt (21 % von  $N = 301$  versus 2 % von  $N = 132$ ). Das kann als ein Hinweis darauf gesehen werden, dass die Längsschnittstichprobe gegenüber der Stichprobe zu T1 ( $N = 516$ ) positiv selektiert ist. Beim Vergleich der drei Stichproben hinsichtlich des für die Bildung aufgewendeten Zeitbudgets zeigt sich, dass die Personen der Stichprobe mit nur einem Messzeitpunkt ( $N = 301$ ) im Mittel ein knappes Jahr weniger in Bildung investiert haben als die der beiden anderen Stichproben ( $F(2) = 8.237, p < .001$ ). In der Längsschnittstichprobe lassen sich auch Veränderungen im Familienstand über die Messzeitpunkte hinweg betrachten. Es zeigt sich, dass vom ersten bis zum vierten Messzeitpunkt die Anzahl der Personen, die verheiratet sind, von 44 Personen auf 36 Personen absinkt und komplementär die Anzahl der Verwitweten ansteigt. Zur Wohnsituation der Mitglieder dieser Stichprobe über die drei Messzeitpunkte hinweg, ergibt sich folgendes Bild: zum ersten Messzeitpunkt leben in Privatwohnungen 91 % der Personen ( $n = 120$ ), zum dritten Messzeitpunkt 89 % (zwei Personen sind in ein Seniorenwohnhaus umgezogen), und zum vierten Messzeitpunkt noch 86 % der Personen (vier Personen sind in ein Seniorenwohnhaus umgezogen).

Tabelle 14. Die drei Stichproben differenziert nach Angaben zu Familienstand, Wohnsituation und Bildung zum ersten Messzeitpunkt

Beschreibung	SP mit drei Messzeitpunkten ( $N = 132$ )		SP mit zwei Messzeitpunkten ( $N = 83$ )		SP mit einem Messzeitpunkt ( $N = 301$ )	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<b>Familienstand</b>						
verheiratet	44	33.3	31	37.3	79	26.2
verwitwet	68	51.5	37	44.6	178	59.1
geschieden	11	8.3	8	9.6	19	6.3
ledig	9	6.8	7	8.4	25	8.3
<b>Wohnsituation</b>						
privat	129	97.7	78	94.0	238	79.1
Institution	3	2.3	5	6.0	63	20.9
<b>Bildung</b>						
in Jahren	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	11.27	2.37	11.19	2.63	10.40	2.18

### 4.3 Erhebungsinstrumente

Die in der Folge vorgestellten Messverfahren sind Teil der Intensivprotokolle der Berliner Altersstudie. Einerseits stammen sie aus der Einheit der Psychologie, und zwar aus der Batterie „Selbst und Persönlichkeit“, andererseits aus der medizinischen Einheit der interdisziplinär konzipierten Studie.

#### 4.3.1 Zur Erhebung von Extraversion und Neurotizismus

Die Dimensionen Extraversion und Neurotizismus wurden über ausgewählte Items des NEO (Costa & McCrae, 1985, siehe auch Borkenau & Ostendorf, 1993) mit jeweils sechs Items erhoben. Für Neurotizismus wurden Items zu den Facetten Ängstlichkeit, Depressivität, Verletzlichkeit und Feindseligkeit dargeboten. Extraversion wurde mit Items zu den Facetten Geselligkeit, positive Emotionalität, Durchsetzungsvermögen und Aktivität beschrieben. In Tabelle 15 sind die ausgewählten Items aufgelistet. In der BASE waren sie mit Items zur Erhebung von „Einsamkeit“ und „Offenheit für Neues“ gemischt, so dass ihre Reihenfolge in Tabelle 15 nicht der Darbietungsreihenfolge bei der Datenerhebung entspricht.

Tabelle 15. *Items zur Erhebung von Extraversion und Neurotizismus in der Berliner Altersstudie*  
(Antwortskala: 1 = trifft sehr gut zu – 5 = trifft überhaupt nicht zu)

Zu bewertende Aussagen
<i>Extraversion</i>
Ich habe gerne viele Leute um mich herum.
Ich unterhalte mich wirklich gerne mit anderen Menschen.
Ich stehe gerne im Mittelpunkt.
Ich bin leicht zum Lachen zu bringen.
Ich bin ein fröhlicher, gut gelaunter Mensch.
Ich bin ein sehr aktiver Mensch.
<i>Neurotizismus</i>
Manchmal fühle ich mich völlig wertlos.
Wenn ich unter starkem Stress stehe, fühle ich mich manchmal als ob ich zusammenbräche.
Ich empfinde oft Furcht oder Angst.
Ich fühle mich oft angespannt und nervös.
Ich ärgere mich oft darüber, wie mich andere Leute behandeln.
Ich fühle mich oft hilflos und wünsche mir jemanden, der meine Probleme löst.

*Anmerkung.* Die Reihenfolge der Items entspricht nicht der Darbietungsform.

Alle Items wurden den Probanden sowohl schriftlich vorgelegt, als auch vorgelesen. Jedes Item konnte über eine fünf-stufige Likert-Skala beantwortet werden (1 = „trifft sehr gut zu“ bis 5 = „trifft überhaupt nicht zu“). Die Probanden bekamen die folgende Instruktion vorgelegt und vorgelesen:



„In der nächsten Aufgabe werde ich Ihnen Sätze vorlesen, mit denen sich anderen Menschen manchmal selbst beschreiben. Ich lese immer einen Satz vor und bitte Sie, mir dann zu sagen, wie gut dieser Satz auf Sie zutrifft.“

Danach erhielt der Proband eine Skala von 1 bis 5 vorgelegt.

„Sie haben dafür eine Skala von 1 bis 5 zur Verfügung, so ähnlich wie sie zur Benotung in der Schule verwendet wird. Mit einer „1“ sagen Sie, dass diese Aussage sehr gut auf Sie zutrifft. Mit einer „5“ sagen Sie, dass diese Aussage auf Sie überhaupt nicht zutrifft.“

Die Probanden konnten ihre Antworten entweder mündlich geben oder sie auf der Skala verorten. Der Untersuchungsleiter hat dann die Antwort in den Fragebogen eingetragen.

Er erhielt, um in schwierigen Situationen regelgerecht handeln zu können, folgende fünf Instruktionen (Arbeitsgruppe "Altern und gesellschaftliche Entwicklung", AGE, 1990):

1. „Bei Entscheidungsschwierigkeiten darauf hinweisen, daß eine allgemeine Einschätzung über verschiedene Situationen hinweg gemeint ist.
2. Bei Verständigungsschwierigkeiten, was mit einem Wort gemeint ist, Item noch einmal vorlesen, auf Betonung achten und darauf hinweisen, dass es darauf ankommt, wie Studienteilnehmer das Wort versteht.
3. Wenn unsicher, ob Notenskala verstanden wurde, den Zahlenwert immer noch einmal verbalisieren.
4. Studienteilnehmer muss selbst die Zahl sagen oder darauf deuten.
5. Falls Studienteilnehmer zuerst nur JA oder NEIN sagt, danach auf der Skala weiter differenzieren.“

Alle Itemantworten zu Extraversion und Neurotizismus sind umgepolt worden, so dass höhere Werte auf einem Item eine höhere Ausprägung auf der Dimension indizieren. Je höher der Punktwert, desto extravertierter oder neurotischer ist die jeweilige Person. Über die Itemantworten einer Dimension hinweg wurde ein Mittelwert gebildet. Die Reliabilitätsberechnung anhand der internen Konsistenz ergibt für die Personen, die zum ersten Messzeitpunkt (T1;  $N = 516$ ) untersucht wurden, ein  $\alpha$  von .64 für Extraversion und ein  $\alpha$  von .75 für Neurotizismus. Für die Stichprobe der Personen mit einem Messzeitpunkt ( $N = 301$ ) liegt die interne Konsistenz für Extraversion bei  $\alpha = .61$ , für Neurotizismus bei  $\alpha = .74$ , in der Stichprobe von Personen mit zwei Messzeitpunkten ( $N = 83$ ) für Extraversion bei  $\alpha = .70$ , für Neurotizismus bei  $\alpha = .75$ ; und für die Längsschnittstichprobe ( $N = 132$ ) für Extraversion bei  $\alpha = .68$ , für Neurotizismus bei  $\alpha = .74$ . Im Vergleich dazu liegt der Median der alphas zur internen Konsistenz bei McCrae und Mitarbeiter (1999) in der deutschen Stichprobe im Alter von 18 bis 83 Jahren ( $N = 3442$ ) bei  $\alpha = .70$ .

### 4.3.2 Verfahren zur Messung der sensorischen Fähigkeiten

#### 4.3.2.1 Verfahren zur Messung der Sehfähigkeit

Das Sehvermögen wurde über die Werte des Nah- und Fernvisus erfasst. Der Nahvisus wurde unter Nutzung von Lesetafeln (siehe Geigy, 1977) bei einer individuell festgelegten Leseentfernung (ungefähr 25 cm) für das linke und rechte Auge getrennt voneinander bestimmt. Bei der Lesetafel handelt es sich um eine Sammlung von Textproben in unterschiedlicher Schriftgröße. Die Untersucher waren trainiert, die Entfernung zur Lesetafel zu standardisieren.

Der Versuchsleiter erfasste den maximal erreichten Tafelwert; dieser entspricht der Textzeile, die der Proband noch richtig lesen konnte. Der maximal erreichte Tafelwert konnte folgende Werte annehmen: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.25 und 1.5 (höhere Werte bedeuten besserer Visus).

Der Fernvisus wurde binokulär bestimmt. Bei der Prüfung sind so genannte „Sehzeichen“ (Optotypen) angeboten worden. Dabei handelte es sich um Optotypentafeln nach Schweigger (siehe Geigy, 1977) mit Buchstaben in unterschiedlicher Größe, die in einem Abstand von mindestens 2.5 m Entfernung aufgestellt waren. Der Untersucher notierte den maximal erreichten Tafelwert - er konnte folgende Werte erreichen: 4, 5, 7.5, 10, 15, 20, 25, 35 und 50 (höhere Werte bedeuten schlechteren Visus) - und die Entfernung der Tafel in Metern. Alle Messungen wurden ohne und falls vorhanden mit eigenen korrigierenden Fern- oder Lesebrillen durchgeführt. Eine Brille trugen über 90 % der BASE-Teilnehmer/innen.

Die Snellen-Dezimale sind als Messeinheit für den Fernvisus so gebildet worden, dass der Wert für die Entfernung zwischen Tafel und Proband durch den maximalen Tafelwert dividiert wurde. Ein Snellen-Dezimal von 1.0 entspricht einer an jungen Erwachsenen normierten unbeeinträchtigten (20/20) Sehschärfe. Der maximal erreichbare Wert des Fernvisus liegt bei 1.6 Snellen; in dieser Stichprobe wurden Werte zwischen .00 und .88 erreicht.

Die Personen dieser Stichprobe variieren hinsichtlich ihres Nahvisus erheblich (*Range*: .00 – 1.00). Der maximal erreichbare Wert beträgt 1.5 Snellen-Dezimale. Er wurde in dieser Stichprobe von keinem Probanden erzielt. Für den Visuswert ergibt sich für die 516 Personen zum ersten Messzeitpunkt eine interne Konsistenz von  $\alpha = .72$ .

#### 4.3.2.2 Verfahren zur Messung der Hörfähigkeit

Um das Hörvermögen zu erfassen, wurde anhand eines Audiometers (Bosch ST-20-1 Reinton) für vier verschiedene Frequenzen (1, 2, 4, 6 kHz in dieser Reihenfolge) die unkorrigierte (ohne

Hörgeräte) Hörschwelle in Dezibel bestimmt. Die korrigierte Hörschwelle konnte nicht festgelegt werden, da die Untersuchung mit Hilfe der Kopfhörer das Tragen von Hörgeräten nicht zuließ. Begonnen wurde die Messung an dem Ohr, das von dem Studienteilnehmer als das „bessere“ bezeichnet worden war oder, wenn der Teilnehmer einen subjektiven Unterschied nicht markieren konnte, am rechten Ohr.

Über das Audiometer wurde ein Ton dargeboten, dessen Schalldruckpegel vom Untersucher zunehmend erhöht wurde. Es wurde für jeden Probanden die Hörschwelle für ausgewählte Frequenzen bestimmt. Als Hörschwelle wird der minimale Schalldruckpegel eines Tones einer bestimmten Frequenz bezeichnet, der vom Probanden wahrgenommen werden kann. Um Fehler beim Ablesen der Markierungen und bei der Schwellen-Zuordnung zu minimieren, wurde eine Zweitkontrolle von unabhängigen Personen durchgeführt. Die Hörschwelle bewegte sich zwischen 0 und 100 dB.

Die Persönlichkeitsdimensionen Extraversion und Neurotizismus sowie Seh- und Hörfähigkeit wurden zu allen drei Messzeitpunkten in der gleichen Reihenfolge erhoben, so dass die Ergebnisse zu allen Bereichen der vorliegenden Studie in einem Längsschnitt vorliegen.

#### **4.4 Datenvorbereitung und statistisches Vorgehen**

##### **4.4.1 Umgang mit Missings**

Die Methode der Latent Growth Modelle ermöglicht es, selbst mit inkompletten Datensätzen (siehe auch Abschnitt 4.4.6) zu arbeiten. Die in der Folge beschriebene Vorgehensweise im Umgang mit fehlenden Werten sollte es ermöglichen, die Anzahl der Personen in den Datensätzen zu den einzelnen Messzeitpunkten konstant zu halten.

Liegen für eine Person Werte zu einem Konstrukt aus Erhebungen zu anderen Messzeitpunkten vor, so werden die für einen Messzeitpunkt fehlenden Werte über eine Schätzung mittels Regressionsgleichung ermittelt. Sie ergeben sich aus der Regression des vorherigen bzw. folgenden Messzeitpunktes mit Alter und Geschlecht als Prädiktor für dieses Konstrukt. D. h. Regressionsgleichung und Regressionskoeffizienten des vorigen bzw. nachfolgenden Messzeitpunktes werden dazu herangezogen. Für Extraversion und Neurotizismus wurde dieses Verfahren bei vier Personen (zu T3), für Hören bei drei Personen (zu T3) und für Sehen bei sieben Personen (zu T1) durchgeführt.

Falls zur Sehfähigkeit Werte fehlten, sind sie, wenn die betreffenden Personen schon zum vorangehenden Messzeitpunkt minimale Werte aufgewiesen hatten, bzw. wenn sie nur zum ersten Messzeitpunkt untersucht werden konnten und dort entsprechende Werte im Nah- bzw. Fernvisus

erreicht hatten, durch eine minimale Ausprägung ersetzt worden. Solche Rechenoperationen sind hinsichtlich des Fernvisus zum ersten Messzeitpunkt für 30, zum dritten Messzeitpunkt für sechs, zum vierten Messzeitpunkt für fünf Personen vorgenommen worden.

Sofern für Personen Werte zu einer Persönlichkeitsdimension bzw. einer sensorischen Fähigkeit allgemein nicht vorlagen, ist für sie der Mittelwert für die Personen derselben Altersgruppe und desselben Geschlechtes eingesetzt worden (zum Vergleich des Vorgehens bei kognitiven Variablen Lindenberger & Baltes, 1997; Lindenberger & Reischies, 1999). Dies ist zu T1 bei einer Person hinsichtlich Extraversion und Neurotizismus erforderlich gewesen, sowie bei je einer für Sehen und Hören.

#### **4.4.2 Konstruktbildung von Extraversion und Neurotizismus**

Die Werte für Extraversion und Neurotizismus wurden über die umgepolten Antworten zu den beschriebenen sechs Items gemittelt. Je höher der Mittelwert war, desto höher fiel die Ausprägung auf der Persönlichkeitsdimension aus. Nach Bildung der Mittelwerte wurden diese zum ersten Messzeitpunkt *t*-standardisiert. Die Mittelwerte für T3 und T4 wurden am Mittelwert und der Standardabweichung von T1 standardisiert, so dass sie mit den entsprechenden Werten für den ersten Messzeitpunkt verglichen werden können. Die Konstrukte von Extraversion und Neurotizismus sind auf der Basis der Mittelwerte aus den Rohdaten gebildet worden und nicht aufgrund von Ergebnissen von Faktorenanalysen. Das ist darin begründet, dass Extraversion und Neurotizismus in dieser Arbeit in Beziehung zu sensorischen Variablen gesetzt werden sollen, die ebenfalls auf der Ebene von Rohdaten beschrieben worden sind. Für die Untersuchung der strukturellen Stabilität wurden faktorenanalytische Modelle anhand von Itemparcels berechnet.

#### **4.4.3 Zur Stabilität der Faktoren (LISREL-Modell, $N = 132$ )**

Ein Verfahren zur Überprüfung von Veränderungen der Struktur von Persönlichkeitsfaktoren über die drei Messzeitpunkte hinweg stellen die Strukturgleichungsmodelle - erstellt mit LISREL (Jöreskog, 1993; Jöreskog & Sörbom, 1989) - dar. Mit Hilfe dieses Verfahrens wurde ein Modell für Extraversion und Neurotizismus erstellt – und zwar mit jeweils drei Messzeitpunkten, so dass sowohl ein Strukturvergleich als auch die Ermittlung der Stabilität von Faktoren möglich sind. Hierzu wurden die Kovarianzmatrizen von jeweils drei Parcels (zur Parcelbildung siehe Anhang C) der Faktorwerte von Extraversion und Neurotizismus zu den drei Messzeitpunkten für die Längsschnittstichprobe von 132 Personen in das LISREL-Programm (Jöreskog, 1993; Jöreskog & Sörbom, 1989) eingelesen.

Zur Untersuchung der Invarianz des Modells über die drei Messzeitpunkte bietet sich eine hierarchisch organisierte Modellspezifizierung an (Schaie, Maitland, Willis & Intrieri, 1998; Meredith, 1993). Im ersten Schritt wird ein konfigurales Invarianzmodell über alle drei Messzeitpunkte berechnet. Das Faktorenmuster ist das gleiche wie in den Modellen, die getrennt für die drei Messzeitpunkte aufgestellt wurden. In einem zweiten Schritt folgt der Test auf metrische Invarianz. Dabei werden die Faktorladungen über die Messzeitpunkte im Modell gleich gesetzt. Der dritte Schritt stellt die Testung der Faktorvarianz-Invarianz dar, für die die Faktorvarianzen ebenfalls über die Messzeitpunkte gleichgesetzt werden. Im vierten Schritt kann die Faktorkovarianz-Invarianz getestet werden, indem zusätzlich die Kovarianzen über die Messzeitpunkte gleich gesetzt werden. Im letzten Schritt ist der Test der Residualvarianz-Invarianz zu überprüfen, wozu zusätzlich die Residualvarianzen der beobachteten Variablen über die Messzeitpunkte gleichgesetzt werden.

Zwischen den einzelnen Modellen ist die Differenz der Fit Indizes zu ermitteln und mit Hilfe eines  $\chi^2$  – Testes auf ihre Signifikanz hin zu testen. Verschlechtert sich der Fit Index nicht signifikant, so kann mit dem Test der nächsten Invarianz (nächster Schritt im hierarchischen Modellvergleich) fortgefahren werden. Unterscheiden sich die Fit Indizes der Modelle signifikant, so ist der Test auf Invarianz zu beenden und es ist möglich, eine Aussage über die vorhandene Invarianz (letztes Modell in der Hierarchie mit signifikanter Verbesserung zu seinem vorangegangenen) zu treffen. Als Signifikanzniveau ist  $p < .01$  analog zur Untersuchung der Invarianz bei Schaie und Kollegen (1998) gewählt worden.

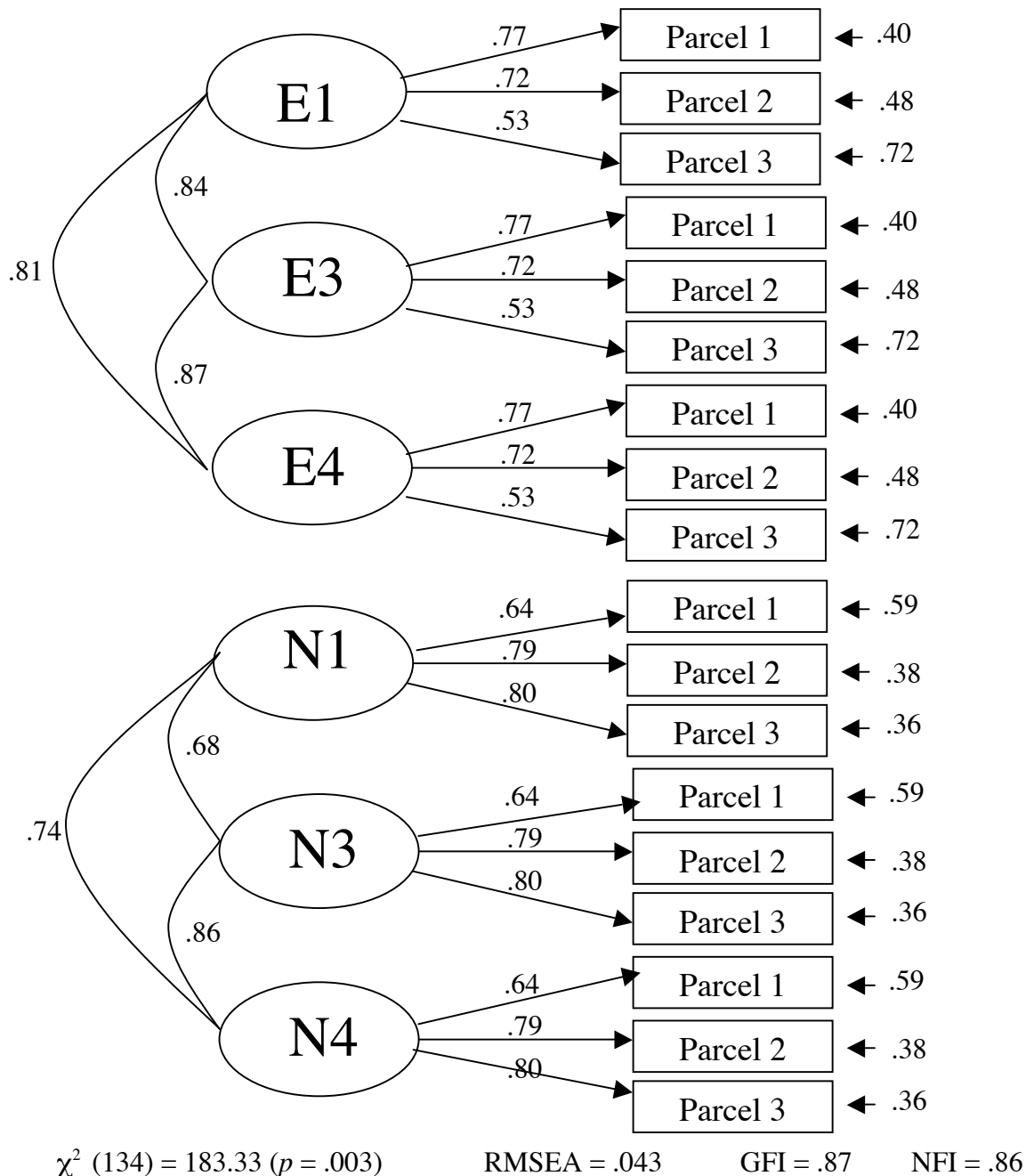
Für die Untersuchung der strukturellen Stabilität von Extraversion und Neurotizismus über die drei Messzeitpunkte hinweg wurde mit Hilfe des LISREL-Programms (Jöreskog & Sörbom, 1989) ein konfirmatorisches faktoranalytisches Modell erstellt. Im hierarchischen Modellvergleich zeigt sich, dass das Modell der Residualvarianzstabilität am besten die Daten widerspiegelt (siehe Tabelle 16).

Tabelle 16. Modelle zur Testung der strukturellen Stabilität über die drei Messzeitpunkte ( $N = 132$ )

Modell	Fit Indizes		Signifikanz der Differenz des $\chi^2$ -Fit Indizes
	RMSEA	$\chi^2$	
Konfigurale Stabilität	0.043	$\chi^2$ (108) = 148.44	
Metrische Stabilität	0.049	$\chi^2$ (116) = 167.45	$\Delta \chi^2$ (8) = 19.06, $p = .015$
Faktorvarianzstabilität	0.047	$\chi^2$ (120) = 168.37	$\Delta \chi^2$ (4) = 0.09, $p = 1.00$
Faktorkovarianzstabilität	0.046	$\chi^2$ (122) = 168.65	$\Delta \chi^2$ (2) = 0.28, $p = .87$
Residualvarianzstabilität	0.043	$\chi^2$ (134) = 183.33	$\Delta \chi^2$ (12) = 14.68, $p = .26$

In Abbildung 13 ist das Modell graphisch dargestellt. Die Kovarianzen sowie die Autokorrelationen zwischen den Residuen sind zur besseren Übersicht in der Graphik weggelassen. In Tabelle 17 sind

zusätzlich die Korrelationen zwischen den latenten Faktoren über die drei Messzeitpunkte zu sehen. Bei einem Vergleich fällt besonders auf, dass Extraversion und Neurotizismus offensichtlich unabhängige Faktoren sind, deren Beziehung zueinander sich nicht über die Zeit hin verändert. Die einzelnen Faktoren erwiesen sich als mit sich selbst über die drei Messzeitpunkte hoch korreliert. Wie Abbildung 13 zu entnehmen ist, zeigen die Fit Indizes einen guten Fit an.



*Anmerkung.* E1, N1 = Extraversion bzw. Neurotizismus zum ersten Messzeitpunkt  
 E3, N3 = Extraversion bzw. Neurotizismus zum dritten Messzeitpunkt  
 E4, N4 = Extraversion bzw. Neurotizismus zum vierten Messzeitpunkt  
 Im Modell sind die Faktorkovarianzen sowie die Autokorrelationen zwischen den Residuen nicht dargestellt.

Abbildung 13. Strukturmodell von Extraversion und Neurotizismus über die drei Messzeitpunkte mit Angaben zum Fit Index des Modells

Tabelle 17. Korrelationen zwischen den latenten Faktoren Extraversion und Neurotizismus über die drei Messzeitpunkte

	N1	N3	N4	E1	E3	E4
N1	1.00					
N3	.68	1.00				
N4	.74	.86	1.00			
E1	.00			1.00		
E3		.00		.84	1.00	
E4			.00	.81	.87	1.00

Festzuhalten ist, dass strukturelle Stabilität über die drei Messzeitpunkte, d. h. über sechs Jahre hinweg, gegeben ist.

#### 4.4.4 Konstruktbildung weiterer Variablen

##### 4.4.4.1 Sehbeeinträchtigung

Bei der Erstellung des Konstruktes Sehbeeinträchtigung ist schrittweise vorgegangen worden. Im Folgenden werden die Schritte bzw. die jeweils daraus resultierenden Variablen nacheinander beschrieben.

Nahvisus. Zur Bestimmung des Nahvisus sind die Tafelwerte – ermittelt in standardisierter Entfernung von 25 cm Leseentfernung – herangezogen worden. Der erreichte Tafelwert entspricht in diesem Fall einem Snellen-Dezimal (*Range*: .00 - 1.00). Für den Nahvisus liegen vier Snellen-Dezimale vor: jeweils zwei für das linke und rechte Auge, einmal mit und einmal ohne Brille. Für die vorliegende Arbeit wurde für den Nahvisus der bessere Wert unabhängig von der Messung mit oder ohne Sehhilfe sowie unabhängig davon, ob das rechte oder linke Auge subjektiv als besser ausgewählt worden war, verwendet. Das bedeutet, dass die Variable „Nahvisus“ über den besten Nahvisuswert bestimmt worden ist.

Fernvisus. Für den Fernvisus liegen zwei Messwerte vor, einer mit und einer ohne Brille. Der Fernvisus wurde über die Entfernung, in der die Tafel dem Probanden dargeboten wurde, geteilt durch den erreichten Tafelwert festgelegt. Dies entspricht dem so genannten Snellen-Dezimal (*Range*: .00 - .88). Auch beim Fernvisus wurde der bessere Wert der Messung mit und ohne Brille ausgewählt, um den Einfluss von schlecht angepassten Sehhilfen auszuschließen.

Visus. Zur Ermittlung des Visus wurde der Mittelwert aus dem besten Nah- und Fernvisus gebildet.

Sehverlust. Diese Variable wurde durch Umpolung der Variable Visus berechnet, weshalb sich ein kontinuierliches Maß an Sehverlust ergibt.

Der Visus sowie der Sehverlust wurden  $t$ -standardisiert. Die Messwerte vom dritten und vierten Messzeitpunkt wurden anhand der Verteilung der Stichprobe zum ersten Messzeitpunkt standardisiert ( $N = 516$ ), so dass eine Abweichung vom Mittelwert ( $M = 50$ ) als Unterschied zwischen dem ersten und dem zu vergleichenden Messzeitpunkt zu betrachten ist und jeweils eine Veränderung indiziert.

Sehbeeinträchtigung. Die Beeinträchtigung des Visus wurde mit Hilfe der Klassifikation der WHO (1980) festgehalten. Danach liegt keine Beeinträchtigung der Sehfähigkeit vor, wenn ein Snellen-Dezimal von 0.8 oder besser gegeben ist. Von einer leichten Beeinträchtigung der Sehkraft wird bei einem Snellen-Dezimal zwischen 0.3 und 0.8 gesprochen. Unter mäßiger Beeinträchtigung der Sehkraft wird ein Snellen-Dezimal zwischen 0.12 und 0.3 verstanden, und unter einer schweren Beeinträchtigung der Sehkraft ein Snellen-Dezimal unter 0.12.

Die Variablen Sehverlust und Sehbeeinträchtigung unterscheiden sich hinsichtlich der Qualität. Der Sehverlust ist ein kontinuierliches Maß an Seheinbußen (in  $t$ -Werten anhand des ersten Messzeitpunktes standardisiert), die Sehbeeinträchtigung ist als eine kategorielle Darstellung der Seheinbußen definiert.

#### 4.4.4.2 Hörbeeinträchtigung

Für alle Parameter des Hörvermögens wurde der Mittelwert der Hörschwellen (in dB) des rechten und des linken Ohres für jede Frequenz (1, 2, 4 und 6 kHz) gebildet, da beim Hören nicht wie beim Sehen von einer Kompensation durch das besser funktionierende Sinnesorgan ausgegangen werden kann. Es kommt vielmehr auf das Vermögen beider Ohren an, damit sich eine Person in der Umwelt, vor allem im Kontext sozialer Interaktionen orientieren kann (Scherer, 2001). Alle weiteren Beschreibungen beziehen sich deshalb auf Mittelwerte für die Hörschwelle (beider Ohren bei allen Frequenzen).

Hören. Das Konstrukt Hören umfasst den Mittelwert der Hörschwellen für die Frequenzen 1, 2, 4 und 6 kHz. Die Variable Hören wurde nach einer Umpolung errechnet, so dass hohe Werte einer besseren Hörfähigkeit entsprechen. Diese wurden wie die Visusvariablen  $t$ -standardisiert. Werte des dritten und vierten Messzeitpunktes wurden in Bezug auf ihre Verteilung des ersten Messzeitpunktes standardisiert, so dass die zu den drei Messzeitpunkten erhobenen Werte im Blick auf Veränderungen miteinander verglichen werden können.

Hörverlust. Die Variable Hörverlust beinhaltet den Mittelwert der Hörschwellen der Frequenzen 1, 2, 4 und 6 kHz und wurde im Gegensatz zur Variable Hören nicht umgepolt, so dass höhere Werte einem höheren Verlust entsprechen. Sie wurden ebenfalls unter Bezug auf den ersten Messzeitpunkt  $t$ -standardisiert.



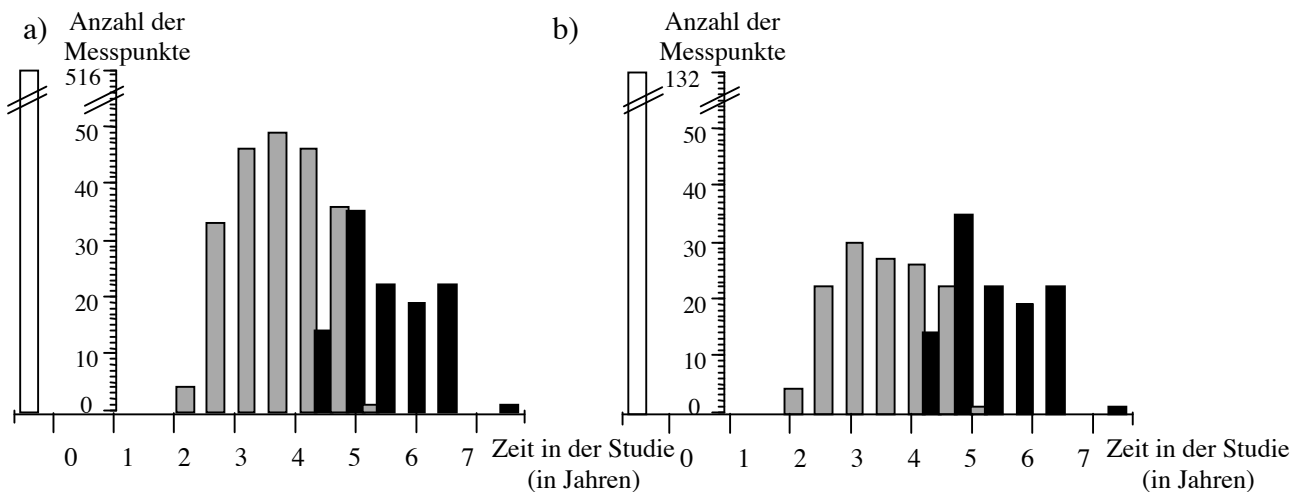
**Hörbeeinträchtigung.** Eine Beeinträchtigung des Hörvermögens wurde unter Bezug auf die Klassifikation der WHO (1980) festgelegt. Danach liegt keine Beeinträchtigung des Hörvermögens vor, wenn die Hörschwelle zwischen 0 und 25 dB liegt. Von einer leichten Beeinträchtigung des Hörvermögens wird bei einer Hörschwelle zwischen 26 und 40 dB gesprochen. Unter mäßiger Beeinträchtigung des Hörvermögens wird eine Hörschwelle zwischen 41 und 55 dB verstanden. Eine schwere Beeinträchtigung des Hörvermögens liegt vor, wenn die Hörschwelle über 56 dB liegt.

Die Variablen Hörverlust und Hörbeeinträchtigung unterscheiden sich in ihrer Darstellung für Höreinbußen. Hörverlust gibt eine Höreinbuße kontinuierlich als Hörschwelle an, Hörbeeinträchtigung wird dagegen nach vier Graden unterschieden. Beide Variablen stellen die gleichen Informationen auf unterschiedliche Art und Weise dar.

#### 4.4.4.3 Zeitintervalle zwischen Messzeitpunkten

Da die Untersuchungstermine exakt protokolliert worden sind, ist es möglich, den Abstand zwischen den einzelnen Messzeitpunkten für jede Person zu bestimmen. Als Variable beinhaltet „Zeit“ die Information, wie lange der Proband an der Berliner Altersstudie teilgenommen hat. Dazu erhalten alle Personen, egal wann für sie der erste Untersuchungstermin gegeben war, zu T1 einen Wert von 0 hinsichtlich der Zeit in der Studie. Damit ist der individuelle Beginn der Studie auf Null gesetzt. Individuell unterschiedlich, abhängig vom tatsächlichen Abstand zwischen dem ersten und dritten Messzeitpunkt, fallen die Werte für das Intervall zum dritten Messzeitpunkt aus (*Range*: 2.37 – 5.18 Jahre). Das gleiche gilt für das Intervall zwischen dem ersten und vierten Messzeitpunkt (*Range*: 4.2 – 7.02 Jahre). Die Zeitvariable wird für die Analyse mit den Latent Growth Modellen benötigt. Ihr kommt in dieser Studie im Blick auf Veränderungen über die Zeit besondere Bedeutung zu.

Der vorliegende Längsschnitt umfasst den Zeitraum von sechs Jahren. In Abbildung 3 ist zu sehen, wie sich die Messpunkte (siehe Abschnitt 4.4.6 zur Erklärung der Messpunkte) über die Zeit in der Studie verteilen. Für die Darstellung wurden Halbjahres-Abschnitte gewählt, so dass ein Jahr mit zwei gleichfarbigen Balken gekennzeichnet ist. Die hellen Balken stellen die Abstände der Messungen von T1 zu T3, die schwarzen den Zeitraum zwischen T1 und T4 dar. In Abbildung 14a sind die jeweiligen Zeitbudgets aller an der Studie beteiligter Personen abzulesen, in Abbildung 14b für diejenigen, die Mitglieder der Längsschnittsstichprobe gewesen sind ( $N = 132$ ). Eine nähere Beschreibung dieser Gruppen findet sich in Abschnitt 5.1. Keine der Personen hat länger als acht Jahre an der Studie teilgenommen.



*Anmerkung.* Der erste Messzeitpunkt wurde auf null gesetzt.  
Die grauen Balken stellen die Messpunkte des dritten, die schwarzen die Messpunkte des vierten Messzeitpunktes dar.

Abbildung 14. Verteilung der Messpunkte über die Zeit in der Studie (in Halbjahresabschnitten) für a) die Stichprobe des inkompletten Datensatzes (863 Messpunkte) und b) die Längsschnittstichprobe (kompletter Datensatz: 396 Messpunkte)

#### 4.4.4.4 Demenz als Kontrollvariable

Personen mit einer Demenzdiagnose zeigen häufig Persönlichkeitsveränderungen (siehe auch Hennerici, 1997). Aus diesem Grund erscheint es sinnvoll, die Stichprobe bezüglich des Vorliegens einer Demenz zu kontrollieren. Das geschieht, indem neben allen Analysen, die anhand der Gesamtstichprobe berechnet wurden, solche für die Stichprobe abzüglich der Personen mit einer Demenzdiagnose durchgeführt und die Ergebnisse beider Rechnungen miteinander verglichen werden. Die Längsschnittstichprobe ( $N = 132$ ) umfasst drei Personen (2.3 %, eine alte und zwei sehr alte Personen), die die Kriterien einer Demenzdiagnose erfüllen. In der Stichprobe mit zwei Messzeitpunkten ( $N = 83$ ) befinden sich 11 Personen mit einer Demenzdiagnose (13.2 %, davon sechs junge und fünf alte Alte) und in der Stichprobe, für die nur Daten zum ersten Messzeitpunkt erhoben werden konnten, sind 95 Personen an einer Demenz erkrankt (31.5 %, davon 15 junge und 80 alte Alte). Bei Ausschluss der 109 Personen mit einer Demenzdiagnose reduziert sich die Anzahl der Messpunkte von 863 auf 737 (siehe Abschnitt 4.4.6 zur Erklärung der Messpunkte). Die Demenzdiagnosen wurden anhand der Diagnosekriterien nach der DSM-III-R (American Psychiatric Association, 1987; Wittchen, Saß, Zaudig & Koehler, 1989) erstellt. Zur Diagnosefindung diente ein halbstrukturiertes Interview mit Hilfe der Geriatric Mental State, Version A/History and Aetiology Schedule (GMS-A/HAS; Copeland, Dewey & Griffiths-Jones, 1986; McWilliam, Copeland, Dewey & Wood, 1988), die 1. Symptome, Verhalten und 2.

Vorgeschichte und Verlauf, Lebensereignisse, Beeinträchtigung im Alltag erfasst. Personen mit leichter, mittlerer und schwerer Demenz wurden aus dem Datensatz herausgenommen.

#### 4.4.4.5 Variablen der Folgeanalysen

Die *subjektive Wahrnehmung einer Veränderung der Sehfähigkeit* wurde als Variable zum vierten Messzeitpunkt gesondert erhoben, indem die Probanden gefragt wurden: „Hat sich die Sehkraft von Ihnen in den letzten 12 Monaten verändert? Sie können mir wieder sagen, welche Antwort auf Sie zutrifft.“ (Arbeitsgruppe "Altern und gesellschaftliche Entwicklung", AGE, 1990). Folgende Liste wurde vorgelegt bzw. vorgelesen: -ist viel besser geworden (1), -ist besser geworden (2), - ist gleich geblieben (3), -ist schlechter geworden (4), - ist viel schlechter geworden (5). Der Untersuchungsleiter erhielt dazu folgende Instruktion: „Bitte nach Veränderung der Sehkraft ohne Korrektur/Brille/neue Gläser fragen! Wenn die Frage nicht verstanden wird, zunächst vereinfacht abfragen, dann durch Nachfragen differenziertere Zuordnung versuchen. Falls anders nicht möglich, vereinfachte Abfrage verwenden und kodieren“. Die Antworten auf der 5-stufigen Likert-Skala wurden zu einer kategoriellen Variable umgeformt: Vorliegen negativer Veränderungen der Sehkraft (ja: Antwortkategorien vier und fünf; bzw. nein: Antwortkategorien eins bis drei).

Die *subjektive Wahrnehmung einer Veränderung der Hörfähigkeit* wurde auf dieselbe Weise erhoben und ist als Variable analog konstruiert. Die spezifische Frage zur subjektiven Veränderung der Hörfähigkeit lautete: „Hat sich das Hörvermögen von Ihnen in den letzten 12 Monaten verändert? Sie können mir wieder sagen, welche Antwort auf Sie zutrifft.“ Folgende Liste wurde vorgelegt bzw. vorgelesen: -ist viel besser geworden (1), -ist besser geworden (2), - ist gleich geblieben (3), -ist schlechter geworden (4), - ist viel schlechter geworden (5). Der Untersuchungsleiter bekam folgende Instruktion: „Bitte nach Veränderung des Hörvermögens ohne Korrektur und ohne Hörhilfe fragen! Wenn die Frage nicht verstanden wird, zunächst vereinfacht abfragen, dann durch Nachfragen differenziertere Zuordnung versuchen. Falls anders nicht möglich, vereinfachte Abfrage verwenden und kodieren“. Die Antworten auf dieser 5-stufigen Likert-Skala wurden zu einer kategoriellen Variable umgeformt: Vorliegen negativer Veränderungen des Hörvermögens (ja: Antwortkategorien vier und fünf; bzw. nein: Antwortkategorien eins bis drei).

Als weitere Variablen wurden in der soziologischen Einheit von BASE *Freizeitaktivitäten* angesprochen: „Denken Sie bitte einmal daran, wie Sie Ihre (freie) Zeit in den letzten 12 Monaten verbracht haben“. Den Probanden wurden Karten mit Angaben zu den folgenden Bereichen vorgelegt: 1.) Sport treiben, z. B. Gymnastik machen, schwimmen gehen, wandern oder etwas anderes; 2.) Cafés/Restaurant besuchen; 3.) Tanzen gehen; 4.) Tagesausflüge machen (bis drei Tage), wandern oder etwas anderes; 5.) Ein Konzert, ein Theater, ein Museum, Ausstellungen,

Messen oder Filmvorstellungen besuchen; 6.) Ein Hobby ausüben, z. B. Briefmarken sammeln, Handarbeiten; 7.) Ehrenamtlich tätig sein, z. B. ehrenamtlich als Kassenswart tätig sein oder etwas anderes; 8.) Reisen, Verwandte oder Bekannte an einem anderen Wohnort besuchen (mind. vier Tage); 9.) Eine künstlerische Tätigkeit ausüben, z. B. malen, singen, töpfern, ein Musikinstrument spielen; 10.) Spielen, z. B. Gesellschaftsspiele, Skat, Kegeln; 11.) Etwas lernen, sich weiterbilden, z. B. zu Hause oder durch den Besuch von Vorträgen und Kursen; 12.) An politischen Veranstaltungen teilnehmen, z. B. einer Partei, Bürgerinitiative, Gewerkschaft.

Die Probanden konnten Angaben machen, ob die genannte Aktivität für sie zutrifft. Über die Bereiche wurde aufsummiert und ein Index für Freizeitaktivität gebildet. Dieser wurde zu allen drei Messzeitpunkten erhoben. Die Indizes sind am ersten Messzeitpunkt  $t$ -standardisiert worden.

Die Freizeitaktivitäten sind in der Auswertung zu Gruppen zusammengefasst worden: solche, die eher in Gesellschaft ausgeübt (Sport, Tanzen, Ausflüge bis drei Tage, Konzert, Ausflüge ab vier Tage und Spielen) und als *gesellige Freizeitaktivitäten* bezeichnet werden, sowie solche, die eher alleine durchgeführt werden (Hobby, Ehrenamt, Kunst, Lernen, Café, Politik), und die als *weniger gesellige Freizeitaktivitäten* bezeichnet werden. Für beide Gruppen wurden Summen über die Aktivitäten gebildet. Die Indizes für gesellige und weniger gesellige Freizeitaktivitäten wurden für jeden der drei Messzeitpunkte gebildet und am ersten Messzeitpunkt  $t$ -standardisiert.

#### 4.4.5 Selektivitätsanalysen

Im Rahmen der deskriptiven Analyse wird des weiteren die Stichprobenselektivität bestimmt, d. h. das Ausmaß, in dem Merkmale, die den Stichprobenausfall seit Beginn der Berliner Altersstudie vorhersagen, mit untersuchungsrelevanten Merkmalen zusammenhängen (Nesselroade, 1988). Die Selektivität der Stichprobe kann unterteilt werden in die *mortalitätsassoziierte* und die *experimentelle* Selektivität (Lindenberger, Singer und Baltes, 2002).

Unter mortalitätsassoziierte Selektivität wird die Selektivität verstanden, die durch den Tod von Personen begründet ist. Experimentelle Selektivität bezeichnet den Ausfall von Personen aufgrund von Desinteresse, zeitlichen Schwierigkeiten oder geistiger und körperlicher Gebrechlichkeit.

Die Gesamtselektivität wie auch die mortalitätsassoziierte und die experimentelle Selektivität lassen sich unterscheiden und getrennt voneinander berechnen (siehe Anhang D). Bei der Gesamtselektivität wird für Personen, für die Daten zu allen Messzeitpunkten vorliegen, der Mittelwert des interessierenden Merkmals zum ersten Messzeitpunkt mit dem Mittelwert aller Personen für das gleiche Merkmal zum ersten Messzeitpunkt verglichen. Dabei interessiert, inwiefern sich die Personen, für die Daten zu allen Messzeitpunkten vorliegen, sich bereits zum

ersten Messzeitpunkt von Personen der Ausgangsstichprobe ( $N = 516$ ) unterscheiden. Die Gesamtselektivität wird mit folgender Formel berechnet (Lindenberger et al., 2002):

$$\text{Gesamtselektivität: } \frac{M_{\text{Teilnehmer}} - M_{\text{Ausgangsstichprobe}}}{SD_{\text{Ausgangsstichprobe}}}$$

Selektivitätsanalysen haben Grenzen. Einerseits können nur Aussagen über die beobachtete Selektivität, nicht aber zur prinzipiell denkbaren Selektivität gemacht werden. Andererseits wird im Folgenden die Längsschnittstichprobe ( $N = 132$ ) mit der Ausgangsstichprobe von T1 ( $N = 516$ ) der Berliner Altersstudie verglichen. Diese ist zwar, was die Stichprobenziehung betrifft, repräsentativ - sie erfolgte über eine Zufallsziehung von Adressen aus dem Berliner Einwohnermeldeamt - doch mit experimenteller Selektivität behaftet. Das hat zur Folge, dass sie als Population nicht repräsentativ sein kann für die 70- bis über 100-Jährigen allgemein (P. Baltes, Mayer, Helmchen & Steinhagen-Thiessen, 1996). Weiterhin zeigen die Ergebnisse der entsprechenden Untersuchung, dass der Umfang der beobachteten Selektivitätseffekte bei keinem Merkmal eine halbe Standardabweichung überschritten hat und es mit Ausnahme der Demenzprävalenz, keine Hinweise für eine ausgeprägte altersdifferentielle Selektivität der Ausgangsstichprobe gegeben hat (vgl. Lindenberger, Gilbert, Pötter, Little & Baltes, 1996).

#### 4.4.6 Methode der Latent Growth Modelle (LGM)

Um bei der Analyse von Veränderungen nicht nur auf die - wie in Abschnitt 4.2.4 angedeutet - für die Variable Alter hoch selektierte Längsschnittstichprobe angewiesen zu sein, wurden in der vorliegenden Studie Latent Growth Modelle gerechnet. Sie werden in diesem Abschnitt näher erläutert.

In den letzten Jahren wurden Methoden entwickelt, die eine Simulation von Entwicklungen über die Zeit, vor allem intraindividuelle Veränderung, ermöglichen (Bryk & Raudenbusch, 1992; McArdle & Anderson, 1990; Nesselroade & Boker, 1994; Rogosa et al., 1982; Rogosa & Willett, 1985). Für die Untersuchung intraindividuelle Veränderungen der Persönlichkeitsdimensionen Extraversion und Neurotizismus sowie der Seh- und Hörfähigkeit über die Zeit wurde sich in der vorliegenden Arbeit für die Technik der Latent Growth Modelle (LGM; latente Entwicklungsmodelle; Jones & Meredith, 1996, McArdle et al., 1998; Ware, 1985; Willett et al., 1998; Singer, 1998; Willett, 1997) entschieden. Sie wird für die Untersuchung intraindividuelle Veränderungen von Persönlichkeit von Pedersen und Reynolds (1998) empfohlen. Ihr Titel besagt, dass sie auf latenter (nicht erfasster bzw. gemessener) Ebene Entwicklung über die Zeit darstellen

kann. Dabei werden latente Variablen als solche verstanden, die frei von random- oder systematischen Messfehlern sind, mit welchen „observed variables“ (beobachtete Variablen) behaftet sind (Bollen, 1989).

Aus dem Repertoire der LGMs werden für die Untersuchung mittlerer intraindividuelle Veränderungen auf der Ebene der Persönlichkeitsdimensionen Extraversion und Neurotizismus in der vorliegenden Arbeit die „gemischten Modelle“ benutzt. Sie werden anhand der SAS-Verfahren PROC MIXED konstruiert und ermöglichen es, jede Veränderung einer Persönlichkeitsdimension auf eine Funktion der Zeit zurückzuführen. Dabei werden sowohl Parameter bestimmt, die den Verlauf der gesamten Stichprobe über die Zeit definieren (fixed effects), als auch solche, die individuelle Unterschiede in den persönlichen Verläufen (random effects) abbilden. Die Modelle werden sowohl für Alters- und Geschlechtsgruppen als auch für Seh- und Hörbeeinträchtigung als Prädiktoren interindividueller Unterschiede intraindividuelle Veränderungen berechnet.

Stark verbreitet ist bei Versuchen, Veränderungen zu untersuchen, das Verfahren, Differenzwerte zu bilden und diese durch den Wert für den Abstand zwischen Test und Retest zu dividieren. So gebildete Veränderungsrate werden in vielen Datenanalysen als outcome bezeichnet und als abhängige Variable für weitergehende Datenanalysen genutzt (zum Beispiel Giambra, Arenberg, Zonderman, Kawas & Costa, 1995; McCrae, Arenberg & Costa, 1987; Sullivan, Rosenbloom, Lim & Pfeffman, 2000). Der einfache und praktische Gebrauch dieser Veränderungsrateberechnung geht einher mit einem „Deal“ theoretischer Komplexität und Konfusion. Methodologische Studien konnten zeigen, dass, wenn die so berechneten Raten als abhängige Variable in multiplen Regressionsanalysen genutzt werden, der resultierende Koeffizient durch unterschiedliche Faktoren beeinflusst wird, und zwar durch Residuenfehler, Messfehler, Regression (und Egression) zum Mittelwert (z. B. Allison, 1990; Nesselroade & Bartsch, 1977, Nesselroade & Cable, 1974; Nesselroade, Stigler & Baltes, 1980; Raykov, 1999, Willett, 1990; Williams & Zimmerman, 1996). Diesen Problemen kann mittels linearer Regressionsmodelle mit zeitabhängigen Variablen begegnet werden (z. B. Boker & McArdle, 1995; Hamagami & McArdle, 2001; McArdle, Hamagami, Elias & Robbins, 1991). Methodologisch begründete Lösungen sind im Konzept der Latent Growth Modelle zu finden (siehe McArdle & Bell, 2000; McArdle, Ferrer-Caja, Hamagami & Woodcock, 2002; McArdle & Hamagami, 2001; Meredith & Tisak, 1990). Bei ihnen handelt es sich mathematisch gesehen um eine Erweiterung der ANOVA und der Profilanalysen. Latent Growth Modelle sind im Vergleich zur ANOVA mit Messwiederholung flexibler, da mit ihrer Hilfe die Analyse *inkompletter* Datensätze möglich ist (McArdle, 1989). „Inkomplett“ hat in diesem Zusammenhang mehrere Bedeutungen. Nicht alle Probanden sind zu den drei Messzeitpunkten der Studie untersucht worden: Wenn die Analysen auf die Längsschnittstichprobe reduziert werden müssten, so hätten sie sich auf eine kleinere Stichprobe zu beziehen (in diesem

Fall:  $N = 132$ ). Angaben von Personen, die nur zu zwei Messzeitpunkten ( $N = 83$ ) – und solchen, die nur einmal ( $N = 301$ ) – untersucht wurden, blieben bei der ANOVA unberücksichtigt. Bei den Latent Growth Modellen zur Berechnung des Niveaus und der Veränderung von Extraversion und Neurotizismus können sie dagegen einbezogen werden.

Die Berechnung der Parameter erfolgt beim LGM-Verfahren jedoch nicht über Daten zu Personen, sondern über Messpunkte (Observations), d. h. Messinformationen. Der Datensatz der vorliegenden Studie weist 863 Messpunkte auf. Von den 863 Messpunkten stammen 301 aus der Stichprobe, für die Daten nur zum ersten Messzeitpunkt vorliegen, 166 Messpunkte von den 83 Befragten, für die Daten zum ersten und dritten Messzeitpunkt erhoben werden konnten ( $2 \times 83$ ), und 396 Messpunkte steuert die Längsschnittstichprobe mit 132 Personen bei, für die Daten zu drei Messzeitpunkten vorliegen ( $3 \times 132$ ). Die Stichprobe mit 863 Messpunkten wird im weiteren Verlauf der Arbeit als Gesamtstichprobe (inkompletter längsschnittlicher Datensatz) bezeichnet – im Vergleich zur Längsschnittstichprobe (kompletter längsschnittlicher Datensatz) von 396 Messpunkten. Davon abzugrenzen ist die Stichprobe von 516 Personen des ersten Messzeitpunktes, die im Weiteren als Ausgangsstichprobe bezeichnet wird. Unter Kontrolle von Demenz umfasst die Stichprobe noch 737 Messpunkte (zur Begründung des Wegfalls von 126 Messpunkten: siehe Abschnitt 4.4.4.4).

Mit der Wahl dieser Analysemethode wird zusätzlich der Tatsache Rechnung getragen, dass in den inkompletten und kompletten Längsschnittdatensätzen einzelne Personen in unterschiedlich großen Zeitintervallen untersucht worden sind. Der vorliegende Längsschnitt erstreckt sich über sechs Jahre. In Abbildung 14 war zu sehen, wie die Messpunkte sich über die Zeitspanne der Studie verteilen. In der herkömmlichen ANOVA mit Messwiederholung wird davon ausgegangen, dass Zeitintervalle zwischen Erhebungen nicht variieren und nur die Personen in die Stichprobe aufgenommen werden können, für die Daten zu drei Messzeitpunkten vorliegen.

Die Grundparameter der Latent Growth Modelle beschreiben systematische Muster individueller Unterschiede in der Veränderung über die Zeit (McArdle & Anderson, 1990). Sie ermöglichen es, eine Entwicklungsfunktion über die Zeit hinweg zu errechnen.

Folgende mathematische Gleichung liegt ihnen zugrunde (McArdle & Anderson, 1990):

$$Y(t,n) = L(n) + B(t) S(n) + E(t,n).$$

Danach ist Variable  $Y$  an  $n$  Personen über  $t$  Zeit in der Studie erhoben worden. In der Gleichung geht der gemessene Wert  $Y(t,n)$  auf drei latente (nicht gemessene) Faktoren individueller Unterschiede zurück: a)  $L(n)$  = Niveau (level, intercept), der so genannte  $Y$ -Achsenabschnitt bzw. die Ausprägung, b)  $S(n)$  = Steigung (slope), des weiteren auch als Veränderung bzw. je nach

Richtung als Abnahme, Stabilität oder Zunahme bezeichnet, gewichtet durch Koeffizienten  $B(t)$  und c)  $E(t,n)$  = Fehler (error).

In den Latent Growth Modellen (siehe auch Stoolmiller, 1995) wird angenommen, dass die latenten Faktoren L und S anhand des Mittelwerts und der Varianz sowie deren Korrelation beschrieben werden können. In der Folge ergeben sich die Parameter  $M_L$  (Mittelwert des Niveaus),  $M_S$  (Mittelwert der Steigung),  $V_L$  (Varianz des Niveaus),  $V_S$  (Varianz der Steigung) &  $r_{L,S}$  (Korrelation zwischen dem Niveau und der Steigung). Der Fehler E ist für jeden Messzeitpunkt spezifisch und über die Messzeitpunkte unkorreliert. Er hat einen Mittelwert von Null, so dass die Abweichung  $V_e$  für individuelle Unterschiede stehen kann, die über die Zeit hinweg unabhängig voneinander bleiben. Diese Fehlervarianz ist in jedem Modell automatisch enthalten und wird im Weiteren als  $V_U$  oder als U bezeichnet (McArdle, 1989; McArdle & Anderson, 1990).

Die Parameter dieses Modells beschreiben sowohl Gruppen- als auch individuelle Veränderungen über die Zeit. Die Modelle kennen drei Indikatoren für Gruppenveränderungen:  $M_L$ ,  $M_S$  &  $B(t)$ , sowie drei für individuelle Unterschiede im Falle der Veränderungsmuster der Gruppen:  $V_L$ ,  $V_S$  &  $r_{L,S}$ . Die Indikatoren der Gruppenveränderung ( $M_L$ ,  $M_S$ ) werden auch als „fixed effect“ bezeichnet und die Parameter der individuellen Unterschiede bei der Veränderung ( $V_L$ ,  $V_S$ ) als „random effect“ (McArdle & Anderson, 1990).

#### 4.4.6.1 Spezifische Analysestrategie und Analyseschritte

Bei der Spezifikation der Latent Growth Modelle wird hierarchisch vorgegangen. Damit ist nicht wie bei der hierarchischen Regression gemeint, dass hinzugenommenen Faktoren zusätzliche Aufklärungsanteile zugeschrieben werden können. Vielmehr wird der Modellvergleich in dem Sinn hierarchisch vorgenommen, als mit einfachen Modellen begonnen wird und diese mit immer komplexeren verglichen werden. In jedem Schritt wird möglichst nur ein Parameter hinzugenommen, so dass sich die zu vergleichenden Modelle nur in ihm unterscheiden. Sein Einfluss wird über die Testung der Signifikanz der Verbesserung des Fit Index mit Hilfe eines  $\chi^2$  - Testes überprüft.

##### 4.4.6.1.1 Bestimmung des Grundmodells

In einem ersten Schritt werden die Niveau-Modelle bestimmt. Dabei dient das Fixed-Niveau Modell als Ausgangsmodell (siehe auch Ghisletta & McArdle, 2001): Das Modell (außer der Varianz des Fehlers wird der Mittelwert des Niveaus bestimmt,  $M_L$ ) beinhaltet, dass alle Messpunkte das gleiche Niveau über die Zeit hinweg zeigen und keine Veränderung über die Zeit bestimmt wird. Im



nächsten Schritt wird die Varianz des Niveaus hinzugenommen (Random-Niveau Modell:  $M_L, V_L$ ). Mittels dieses Modells wird ein Mittelwert für das Niveau, aber auch die Varianz um diesen Wert berechnet, was bedeutet, dass davon ausgegangen wird, dass es Personen gibt, die sich in ihren Ausprägungen von diesem Mittelwert des Niveaus unterscheiden. Bei den folgenden linearen Modellen wird das Fixed-Niveau Fixed-Steigungs Modell, bei dem  $M_L, M_S$  berechnet werden, vom Random-Niveau Fixed-Steigungs Modell mit der Bestimmung von  $M_L, M_S, V_L$  unterschieden sowie vom Random-Niveau Random-Steigungs Modell, in das  $M_L, M_S, V_L, V_S, r_{LS}$  eingehen. Das Fixed-Niveau Fixed-Steigungs Modell impliziert, dass alle Messpunkte das gleiche Niveau und die gleiche Steigung aufweisen. Das Random-Niveau Fixed-Steigungs Modell hingegen geht davon aus, dass sich die Messpunkte zwar hinsichtlich des Niveaus unterscheiden, aber alle die gleiche Steigung aufweisen. Beim letzten Vergleich wird davon ausgegangen, dass sich die Messpunkte hinsichtlich des Niveaus und der Steigung unterscheiden.

Von diesen Modellen können solche unterschieden werden, die dazu dienen, zusätzlich noch den Einfluss einer Kovariaten auf das Niveau und die Steigung zu untersuchen (Eine detaillierte Beschreibung findet sich in Abschnitt 4.4.6.1.2).

Das Modell, das für die Berechnung des Verlaufs von Extraversion bzw. Neurotizismus angewendet wird, kann über ein Pfaddiagramm (siehe Abbildung 15) dargestellt werden. Dort ist das am weitesten detaillierte Grundmodell mit einer maximalen Anzahl von Parametern zu sehen. Die vorliegende Arbeit orientiert sich deshalb an diesem Modell, weil es unter anderem ein Ziel der Untersuchung ist, das Modell zu bestimmen, mit Hilfe dessen in der Modellhierarchie die vorliegenden Daten in Blick auf die Forschungsfragen am besten beschrieben werden können.

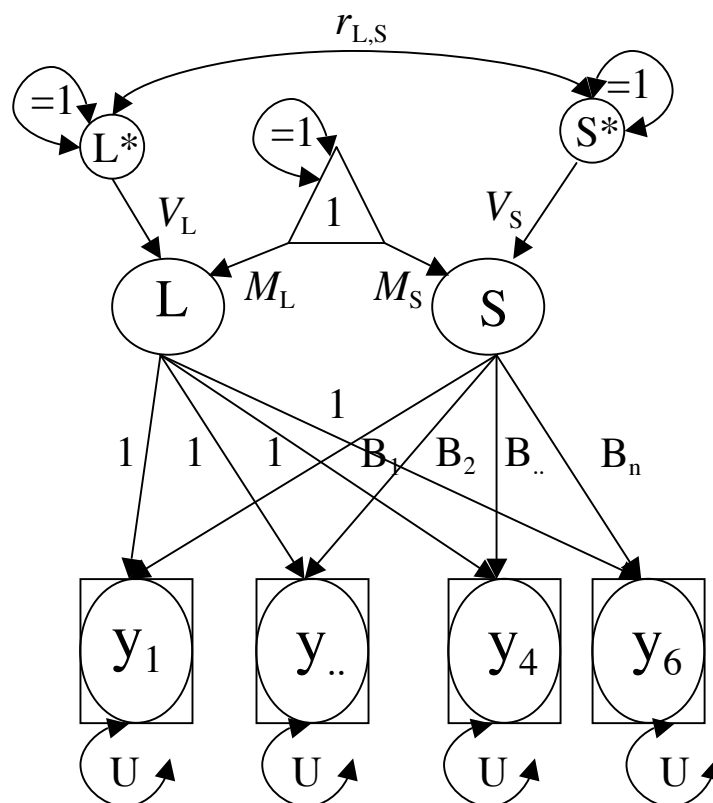
Für die in der Graphik mit Rechtecken eingerahmten Variablen konnten in der Untersuchung Werte ermittelt werden, während mit Kreisen gekennzeichneten latente Variablen darstellen. Sei sind in der Untersuchung nicht erfasst worden. Bei den Variablen, die mit einem Kreis und einem Rechteck eingerahmt sind, handelt es sich um solche, die nicht bei allen Teilnehmern der Studie erfasst werden konnten.

Die latente Variable „L“ steht für das Niveau der Ausprägung („Level“). Alle möglichen Varianzen der Zeit in Studie „y“ sind auf „1“ fixiert. Das beinhaltet, dass dieser Parameter keine Veränderungsinformation einschließt. Der Faktor „L\*“ ist in dieses Modell aufgenommen worden, um die Varianz ( $V_L$ ) des Niveaufaktors „L“ zu bestimmen. Ebenfalls gehört zu dem Modell der Mittelwert ( $M_L$ ) des Niveaufaktors „L“. Das Dreieck mit der „1“ symbolisiert eine Konstante mit dem Wert 1. Analog zu den Faktoren „L“, „L\*“ umfasst das Modell die Faktoren „S“ und „S\*“, d. h. „Steigung“ („slope“). Über sie wird der Mittelwert und die Varianz der Veränderung ( $M_S, V_S$ ) erfasst. Weitere Parameter, die das Modell einschließt, sind die Fehlervarianz, durch „ $V_U$ “ bzw. „U“ symbolisiert, „B“ als Gewichtungsfaktor und die Korrelation zwischen dem Niveau und der

Veränderung:  $r_{L,S}$  (siehe auch Stoolmiller, 1995). In der vorliegenden Arbeit wird bei der Ergebnisdarstellung nicht auf diese Korrelation eingegangen werden, da ihre Aussagekraft im wissenschaftlichen Diskurs umstritten ist (Kreft, Leeuw & Aiken, 1995; Mehta & West, 2000; Wainer, 2000).

Schließlich ist zu unterstreichen, dass „y“ die Zeitvariable darstellt, die als „Zeit in der Studie“ definiert ist. Da die Zeit in der Studie je nach Proband variiert, ist im Modell einerseits „y“ von einem Kreis und einem Rechteck eingerahmt und andererseits sind Werte von „y<sub>1</sub>“ bis „y<sub>n</sub>“ angenommen (kontinuierliche Variable der Zeit).

Der in dieser Studie zu untersuchende lineare Zusammenhang von Veränderung von Extraversion und Neurotizismus und der Zeit ist über „B<sub>1</sub>“ bis „B<sub>n</sub>“ zu spezifizieren.



Anmerkung. L = Level, Niveau; S = Slope, Steigung; y = Zeit in der Studie; U = Fehlervarianz; B = Gewichtungsfaktor; M = Mittelwert; V = Varianz; 1 = Konstante, Wert auf 1 gesetzt.

Abbildung 15. Pfaddiagramm des Latent Growth Modells zum ersten Hypothesenblock für den Verlauf von Extraversion und Neurotizismus (nach McArdle & Anderson, 1990, S. 25)

Zur Berechnung der LGMs wurde in der vorliegenden Arbeit auf das Programm SAS PROC MIXED zurückgegriffen. Mit dem SAS PROC MIXED werden zu jedem Modell Informationen zum Fit Index ermittelt. Danach gibt es mehrere Fit Indizes, die für jedes Modell zu bestimmen sind (siehe Verbeke & Molenberghs, 1997, 2000). Der „Akaike's Information Criterion“ (AIC; Akaike,

1974) wird folgendermaßen berechnet:  $AIC = p(\theta) - q$ , wobei  $p(\theta)$  entweder für den REML (=restricted maximum likelihood) oder den ML (=Maximum likelihood) log likelihood steht und  $q$  für die Anzahl der Kovarianzparameter. REML und ML stehen für die Methode der „Estimation“ der Kovarianzparameter. Der AIC kann dazu genutzt werden, Modelle mit den gleichen „fixed effects“, aber unterschiedlichen Kovarianzstrukturen miteinander zu vergleichen. Je höher der AIC, desto besser bildet das Modell die Daten ab.

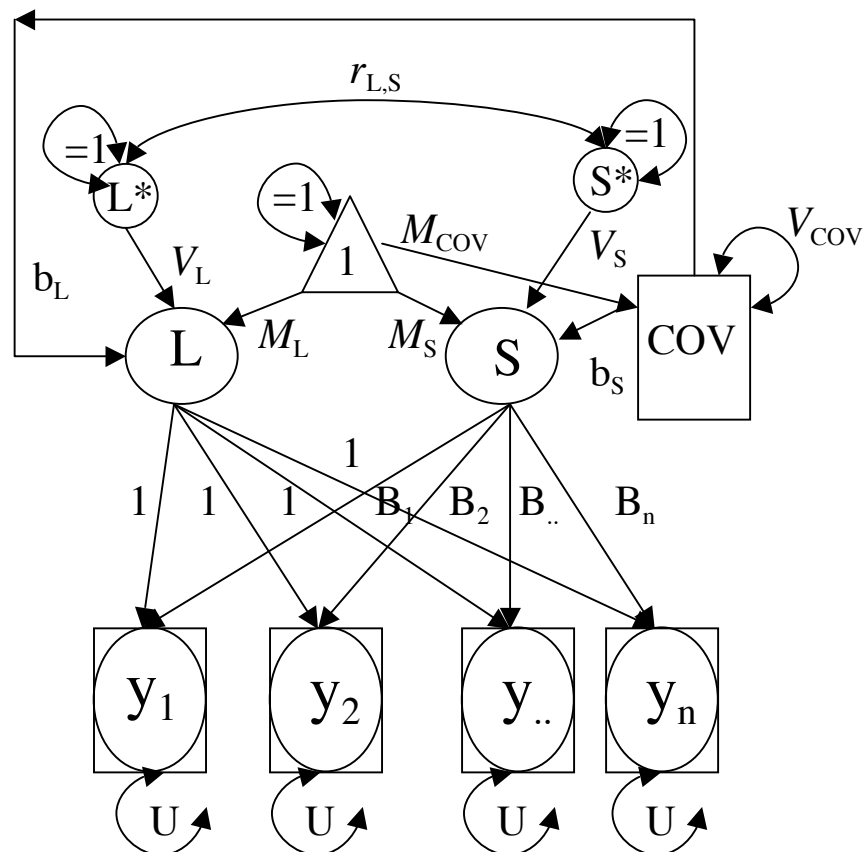
Der „*Schwartz's Bayesian Criterion*“ (SBC; Schwartz, 1978) wird folgendermaßen berechnet:  $SBC = p(\theta) - 1/2 q \log N^*$ , wobei  $N^*$  gleich  $N$  für ML und  $N - p$  für REML. Modelle mit höherem SBC werden bevorzugt.

Der Vergleich der unterschiedlichen linearen Modelle wird in der vorliegenden Arbeit über den Fit Index des „*-2 natural log likelihood*“ (-2LL) vorgenommen. Je kleiner er ist, desto besser bildet das Modell die Rohdaten ab. „-2LL“ ist der natürliche Logarithmus der „maximum likelihood“-Funktion und dient als Fit Index. Dieser ist abhängig von der Stichprobengröße bzw. der Anzahl der Messpunkte. Aus diesem Grund beinhaltet der Fit Index an sich keine Informationen von großem Gehalt für das Modell. Er kann aber zum Vergleich von Modellen herangezogen werden. Ein solcher Vergleich ist nur unter der Voraussetzung möglich, dass in den beiden Modellen die gleiche Stichprobe bzw. die gleichen Messpunkte untersucht wurden und diese „nested“ sind. Unter „nested“ wird in diesem Kontext verstanden, dass zwei Modelle sich – wie oben gezeigt - nur durch Hinzunahme von Parametern unterscheiden. Das heißt, dass das Modell mit der geringeren Anzahl von Parametern im Modell mit der größeren Anzahl von Parametern bereits enthalten ist. Zum Vergleich zweier Modelle wird die Differenz zwischen den Fit Indizes -2LL und die Differenz der Freiheitsgrade ermittelt. Mit diesen Informationen wird ein  $\chi^2$ -Vergleich durchgeführt. Über ihn wird getestet, ob die Verbesserung des Fit Index unter Berücksichtigung der Erhöhung der Freiheitsgrade signifikant ist (siehe Ghisletta & McArdle, 2001).

Der „Null Model LRT Chi-Square“-Wert berechnet sich aus der Differenz zwischen dem „-2LL“ des berechneten Modells und dem Nullmodell, welches ausschließlich „Fixed effects“ zulässt. Diese Statistik hat eine asymptotische  $\chi^2$ -Verteilung mit  $q - 1$  Freiheitsgraden. Der  $p$ -Wert teilt das obere Areal dieser Verteilung ab. Der  $p$ -Wert kann dann zur Bestimmung der Signifikanz des Fit Index des Modells genutzt werden. Dieses Verfahren wird in der Literatur als zu liberal eingestuft (zur Kritik siehe Hu & Bentler, 1999). Aus diesem Grund wird in der vorliegenden Arbeit zum Vergleich von Modellen auf den „-2LL“ zurückgegriffen.

#### 4.4.6.1.2 Untersuchung des Einflusses von Kovariaten

Nachdem das Grundmodell bestimmt ist, mit dem die vorhandenen Daten am besten beschrieben werden können, können nun mögliche Einflussfaktoren durch die Hinzunahme von Kovariaten ermittelt werden (siehe Singer, 1998). In dieser Arbeit werden das folgende Variablen sein: Altersgruppe, Geschlecht, Sehbeeinträchtigung und Hörbeeinträchtigung. In Abbildung 15 ist das Pfadmodell um eine Kovariante ergänzt dargestellt. Die Modelle von Abbildung 15 und Abbildung 16 sind Beispiele für „nested models“. Das Modell aus Abbildung 15 ist in dem von Abbildung 16 enthalten. Der einzige Unterschied zwischen den beiden besteht darin, dass in Abbildung 16 eine Kovariante ins Modell integriert worden ist. Damit soll ermittelt werden, ob ihrem Einfluss auf das Niveau und die Veränderung Bedeutung zukommt, d. h. ob eine signifikante Verbesserung durch die Hinzunahme der Kovariaten erreicht wird. In der Graphik ist das am stärksten detaillierte Grundmodell inklusive einer Kovariaten mit maximaler Anzahl von Parametern dargestellt.



Anmerkung. L = Level, Niveau; S = Slope, Steigung; y = Zeit in Studie; U = Fehlervarianz; COV = Kovariante; M = Mittelwert; V = Varianz; 1 = Konstante, Wert auf 1 gesetzt; B = Gewichtungsfaktor; b = Kovariationseffekt.

Abbildung 16. Pfaddiagramm des Latent Growth Modells zum zweiten Hypothesenblock für den Einfluss einer Kovariaten auf die Ausprägung und deren Verlauf von Extraversion und Neurotizismus (nach McArdle & Anderson, 1990, S. 25)

Zur Berechnung des Einflusses von Kovariaten auf das Niveau und die Veränderung bzw. auf die Stabilität der abhängigen Variablen sind kategorielle Kovariaten kontinuierlichen vorzuziehen, da dabei bestimmt wird, in welchem Maß sich zum Beispiel das Niveau der abhängigen Variable in unterschiedlichen Kategorien einer Kovariaten unterscheidet (McArdle & Aber, 1990). Dies ist insofern analog zu einer Varianzanalyse (ANOVA) zu sehen, bei der ebenfalls Gruppenunterschiede im Hinblick auf eine abhängige Variable untersucht werden. Aus diesem Grund werden in der vorliegenden Studie Alter, Seh- und Hörfähigkeit als Kovariaten in kategorieller Form in die Grundmodelle eingeführt.

Nach Bestimmung des Grundmodells und der Kovariaten mit Einfluss auf das Niveau oder die Veränderung lässt sich die Varianzaufklärung des Niveaus oder der Veränderung durch die Kovariate ermitteln. Die Varianzaufklärung der Kovariaten (Beispiel an der Aufklärung der Varianz vom Niveau durch eine Kovariate; Singer, 1998) wird folgendermaßen berechnet:

$$\text{aufgeklärte Varianz des Niveaus in \%} = \frac{(\text{Niveauvarianz ohne Kovariate} - \text{Niveauvarianz mit Kovariate})}{\text{Niveauvarianz ohne Kovariate}} \times 100$$

Zur Berechnung der Varianzanteile, die sich zwei Kovariaten teilen (shared effect), und der Anteile jeder einzelnen Kovariate (unique effect) wird von dem Anteil der Varianzaufklärung des Niveaus, der von allen Kovariaten inklusive der interessierenden Variable aufgeklärt wird, der Anteil der Varianzaufklärung des Niveaus subtrahiert, der von allen Kovariaten ohne die interessierende Kovariate aufgeklärt wird (Hatcher & Stepanski, 1994).

Zum Beispiel ist von Interesse, inwiefern Altersgruppenzugehörigkeit und Sehbeeinträchtigung gemeinsam Niveauvarianz aufklären. Um den Effekt der Altersgruppenzugehörigkeit, den sie nicht mit Sehbeeinträchtigung teilt (unique effect), auszurechnen, wird die Varianzaufklärung von Sehbeeinträchtigung von der, die Altersgruppe und Sehbeeinträchtigung zusammen hinsichtlich des Niveaus aufklären, subtrahiert. Analog wird beim unique effect von Sehbeeinträchtigung verfahren. In einer Formel ausgedrückt sieht das folgendermaßen aus:

$$\text{Full equation} - \text{reduced equation} = \text{unique effect der interessierenden Variable.}$$

Die „full equation“ ist die aufgeklärte Varianz aller Kovariaten. Die „reduced equation“ steht für die aufgeklärte Varianz aller Kovariaten, ohne die besonders interessierende (Hatcher & Stepanski, 1994).

Wie bereits erwähnt, gibt es zwei Gründe, die Analyse mit Hilfe der Latent Growth Modelle durchzuführen: 1.) Sie ermöglichen die Analyse inkompletter Datensätze. Das heißt, dass im

Vergleich zu einer ANOVA die Information der Veränderung zwischen ersten und dritten Messzeitpunkt und die Information des Niveaus aller Personen des ersten Messzeitpunktes in die Analysen eingehen. 2.) Die LGMs machen es möglich, unterschiedlich lange Intervalle zwischen den Messzeitpunkten in der Analyse zu berücksichtigen.

Die Analyse eines kompletten Datensatzes, dessen Personen alle zu den gleichen Messzeitpunkten erhoben wurden, müsste anhand der LGM und einer ANOVA mit Messwiederholung identische Ergebnisse hervorbringen. Das ist der Fall, wenn das LGM so spezifiziert wird, dass keine random effects berechnet werden. Dies wird durch das Einfügen des REPEAT-Befehls bewirkt (siehe Verbeke & Molenberghs, 1997; Verbeke & Molenberghs, 2000).

Boker und Maxwell (2001) zeigen eindrücklich, dass unter Verwendung eines Datensatzes, der komplett und zu gleichen Messzeitpunkten erhoben wurde, die LGMs einer Regression und einer ANOVA gleichen. Boker und Maxwell (2001) interessieren sich dafür, ob und wie weit ein IQ Veränderungen als Funktion der Erhebungswelle indiziert. Daher lautet ihre Frage, ob eine Erhebungswelle geeignet ist, IQ-Werte vorhersagen zu können. Für die Analyse, die diese Frage beantworten soll, entschieden sie sich für ein Regressionsmodell. Als Alternative untersuchten sie ihren Datensatz anhand einer ANOVA. Bei der Untersuchung des LGMs ohne random effects ergibt sich eine Bestätigung der Ergebnisse des Regressionsmodells, was bestätigt, dass ein LGM ohne random effects einem normalen Regressionsmodell entspricht. Die traditionelle ANOVA ist äquivalent zu einem random intercept Modell (eines kompletten Datensatzes). Entsprechungen dieser Art sind nur unter der Bedingung gegeben, dass es sich um einen kompletten Längsschnittdatensatz handelt und die Zeitvariable in Form von Messzeitpunkten und nicht als kontinuierliche Variable angenommen wird. Für Details eines Vergleichs von LGM und MANOVA siehe Roger und Kenward (1993) und Roger (1993).

Zusammenfassend lässt sich festhalten, dass die Veränderungsanalyse anhand des Längsschnittdatensatzes von 132 Personen mit Hilfe einer Varianzanalyse mit Messwiederholung zu gleichen Ergebnissen wie eine Analyse mit Hilfe von LGMs kommen müsste. Für diese Studie wurde sich deshalb für die Analyse der Daten mit Hilfe von LGMs entschieden, weil

1. die LGMs es erlauben, Unterschieden in den Zeitintervallen zwischen den Messzeitpunkten für die einzelnen Versuchsteilnehmer gerecht zu werden, und
2. sie es ermöglichen, zusätzlich zu den 132 Personen der Längsschnittstichprobe, Personen in die Analyse einzubeziehen, die nur zu zwei ( $N = 83$ ) bzw. nur zu einem ( $N = 301$ ) Messzeitpunkt befragt werden konnten.

Dies ist für diese Studie umso bedeutender, als es zur Konsequenz hat, dass damit das Selektivitätsproblem, das bei einer Untersuchung allein der Längsschnittstichprobe von 132 Personen gegeben ist, reduziert werden kann. Deshalb wird bei den LGMs nicht mehr von

Stichproben mit einer bestimmten Anzahl von Personen gesprochen, sondern von Stichproben, die eine bestimmte Anzahl von Messpunkten einbeziehen. Der vorliegenden Studie kann folglich ein inkompletter Datensatz zugrunde gelegt werden, in dem alle Messinformationen, die zu den ausgewählten Variablen in der BASE erhoben worden sind, berücksichtigt werden. Er umfasst 863 Messpunkte (396 von der Längsschnittstichprobe, 166 von der Stichprobe mit zwei Messzeitpunkten, und 301 von den Personen, die nur zu einem Messzeitpunkt erhoben wurden). Für die Längsschnittstichprobe werden Modelle über 396 Messpunkte berechnet und - nach Kontrolle der Demenz - solche über 737 Messpunkte. Schließlich ist hervorzuheben, dass die Modelle latente Parameter berechnen. Bollen (1989) führt dazu aus, dass latente Variablen – im Unterschied zu „observed variables“ (beobachtete Variablen) – frei sind von random- oder systematischen Messfehlern. Je geringer die Erhebung der „observed variables“ mit Messfehlern behaftet ist, desto weniger sollten sich Unterschiede zwischen den Ergebnissen zeigen, die bei der Untersuchung latenter Variablen und „observed variables“ anfallen. Vor diesem Hintergrund haben die oben ausgeführten Vorzüge der Latent Growth Modelle bei der Wahl der Methode für die Untersuchung der Stichproben der vorliegenden Studie den Ausschlag gegeben. Im Weiteren wird bei Mittelwert und Varianz in den LGM von latenten Parametern gesprochen.