III METHOD

The Method chapter is structured around three major topics. First, detailed information about the sample and the data collection procedure is provided. This is followed by a description of the measures to be used. Finally, to provide the background for further steps planned in the present dissertation, a set of preliminary analyses examined issues of differential sample attrition in BASE over a period of six years.

3. 1 Sample and Data Protocol

To examine the research questions formulated in the preceding section, data are used from a longitudinal subsample of the Berlin Aging Study (N = 132), which was measured three times over a period of six years (i.e., between 1990/93 and 1997/98; for more extensive information, see P. B. Baltes & Mayer, 1999; Smith & Delius, 2003). Table 4 contains detailed information about baseline and longitudinal study design and assessment procedure. It can be seen, for example, that the Berlin Aging Study involved an in-depth and intensive assessment procedure with 14 sessions of measurement at baseline and 6 sessions at 6-year follow-up (plus 6 sessions at 4-year follow-up, which results in 26 sessions over the years). At each wave, data collection started with a multi-disciplinary Intake Assessment (MIA), which corresponded in time and effort to a typical multidisciplinary survey study of older adults. This was followed by several single sessions (Intensive Protocol) of discipline-specific assessment. Each of the research units involved in BASE were primarily responsible for these sessions. Sessions required an average of 90 minutes and, when necessary, where split up to shorter units of assessment. Testing was carried out by trained research assistants and medical personnel at the participant's place of residence (i.e., private household or institution) with the exception of those sessions that involved geriatrics and dentistry, in which participants were taken to various places at the Free University Berlin.

Table 4 also shows the sample sizes at each occasion as well as information regarding study drop-out (e.g., mortality, refusal). For example, 46% of the total cross-sectional sample were still alive in 1997 and 55% of those took part in the third wave of data collection.

Table 4
Longitudinal Follow-Up of the Total Cross-Sectional BASE Sample: Overview of Assessment Procedure and Samples

	Measurement Occasions	
	Total Sample (1990–93)	3rd Wave (1997–98)
Number of Sessions	14	6
Content of Session	MIA[a] (1) Psychiatry (3) Sociology (3) Geriatrics (3) Dentistry (1) Psychology (3)	MIA (1) Psychiatry (1) Sociology (1) Geriatrics (1) Psychology (1) Everyday Competence (1)
Survivor of Total Sample (%)	—	*N* = 239 (46.3%)
Refused	—	*N* = 63
Moved / not Reached	—	*N* = 6
Incomplete MIA	—	*N* = 6
Complete MIA	— [a]	*N* = 32
Participants with IP (% of Survivors)	*N* = 516	*N* = 132 (55%)

Note. MIA = Multidisciplinary Intake Assessment. IP = Multidisciplinary Intensive Protocol. [a] At baseline assessment, MIA was carried out with additional 412 participants. Due to research pragmatics and financial reasons, these participants were not contacted again after baseline. Adapted from Smith & Delius, 2003.

Figure 3 shows the sample distribution of age groups/cohorts over time. At baseline assessment, the study design in BASE involved an age-stratified sample of men and women aged 70–100+. That is, there was an equal number of male and female participants (*n* = 43) in each of the six age-cohort groups (*n* = 86 in total in each age group: 70–74, 75–79, 80–84, 85–89, 90–95, and 95+ years) resulting in an average age of 85 years. Due to sample attrition (and particularly the death of the oldest; Lindenberger et al., 2002), the 6-year longitudinal sample primarily involved participants in their late seventies and early eighties and clearly lacked individuals older than age 90. In fact, the BASE sample became younger, on average, over time. For example, the average age of participants from the 6-year longitudinal sample was 84 years in 1997. In other words, age stratification only holds for the total cross-sectional BASE sample, whereas the 6-year longitudinal sample had a skewed age distribution and participants from this sample were considerably younger (78 years at baseline, on average).

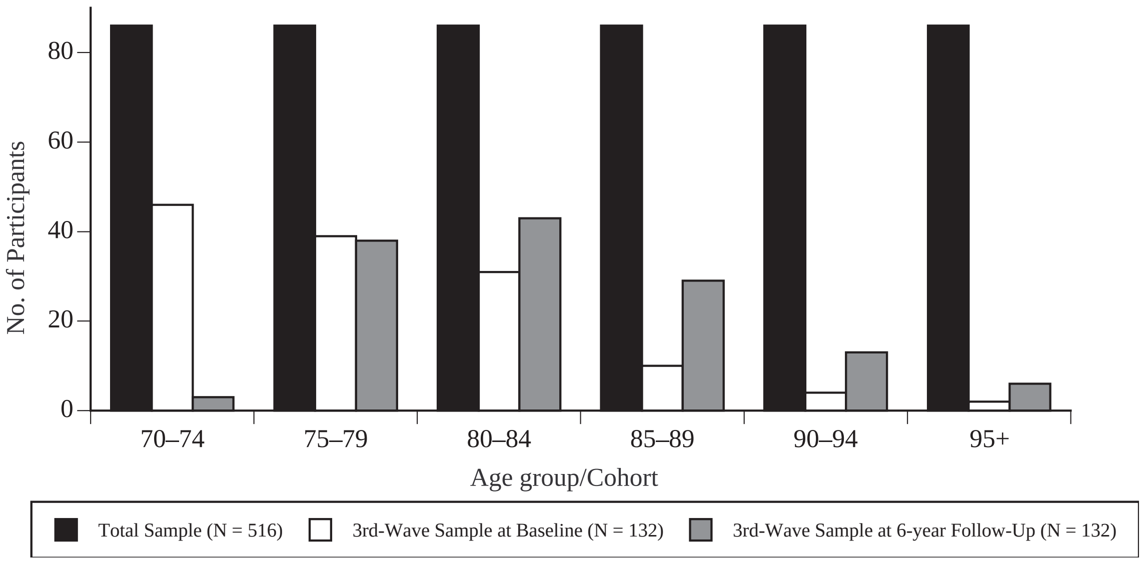

Figure 3 Evolution of the BASE sample over 6 years: Changes in shape and size by age-cohort group.

Note. The total cross-sectional BASE sample involved an equal number of 86 participants in each of the six age-cohort groups. In contrast, the 6-year longitudinal sample primarily included participants who were in their 70s and early 80s at baseline assessment, but there were few participants older than age 90. As a consequence, the majority of the sample was six years later in their late 70s and mid 80s. Adapted from Smith and Delius (2003).

Table 5 gives demographic characteristics of the total cross-sectional and both longitudinal follow-up samples. In contrast to the skewed age distribution in both longitudinal samples, there was still an almost equal distribution of men (*n* = 60; 45%) and women (*n* = 72; 55%) in the 6-year sample. This Table also indicates that there were no major differences in several demographic characteristics between the samples (i.e., education, occupational training, income, marital status, and type of residence).

Table 5
Demographic Characteristics of the Total Cross-Sectional BASE Sample and the 3-Wave Longitudinal Sample at Baseline Assessment and at 6-Year Follow-Up

	Total Sample (1990–93)			3-Wave Sample at Baseline (1990–93)			3-Wave Sample at 6-year Follow-Up (1997–98)		
Demographic characteristics	M *N* = 258	W *N* = 258	Total *N* = 516	M *N* = 60	W *N* = 72	Total *N* = 132	M *N* = 60	W *N* = 72	Total *N* = 132
Age (in years)									
M	84.7	85.1	84.9	78.1	78.4	78.3	83.5	84.0	83.8
SD	8.4	8.9	8.7	5.8	6.1	5.9	5.8	6.1	5.9
Education in years									
up to 8 years	16.0	37.0	26.2	12.1	22.2	17.7			
8 to 10 years	33.6	20.9	27.4	25.9	26.4	26.2	—	—	—
10 to 13 years	38.8	39.6	39.2	44.8	45.8	45.4			
more than 13 years	11.6	2.6	7.2	17.2	5.6	10.8			
Occupational training (in %)	74.8	42.0	58.4	73.3	54.2	62.9	—	—	—
Income (in %)									
< 1.000 DM	4.7	5.0	4.8	1.7	1.4	1.5			
1.000 – 1.399 DM	18.2	19.8	19.0	11.7	13.9	12.9			
1.400 – 1.799 DM	18.2	20.5	19.4	11.7	8.3	9.8			
1.800 – 2.199 DM	15.1	26.4	20.7	15.0	25.0	20.5	—	—	—
> 2.200 DM	43.8	28.3	36.0	60.0	51.4	55.3			
M (in DM)	2208	1877	2042	2707	2328	2500			
SD (in DM)	1249	664	980	1659	943	1325			
Marital status (in %)									
Married	52.3	7.4	29.8	63.3	8.3	33.3	56.7	5.6	28.8
Widowed	39.9	69.8	54.8	30.0	69.4	51.5	38.3	73.6	57.6
Divorced	4.3	10.5	7.4	—	15.3	8.4	—	13.9	7.6
Single	3.5	12.4	7.9	6.7	7.0	6.8	5.0	6.9	6.1
Type of residence (%)									
Living alone	36.8	65.1	50.9	35.0	83.3	61.4	33.3	79.2	58.3
Living with others	53.5	17.1	35.3	63.3	1.4	29.5	60.0	—	27.7
institutionalized	9.7	17.8	13.8	1.7	15.3	9.1	6.7	20.8	14.4

Note. M = men, W = women. Education, occupation, and income were omitted for the 3-wave sample at follow-up because these measures were stable over time. Adapted from Smith and Delius (2003).

3. 2 Measures

3. 2. 1 Psychological Measures

To be in congruence with the previous cluster analysis carried out in the context of BASE (Smith & Baltes, 1997), the present study opted for a broadly defined assessment of psychological functioning in old age. To do so, 11 constructs including cognition, personality and self-related functioning, and social integration were selected from the BASE data protocol. These constructs were available for all three measurement occasions, baseline assessment: 1990–1993, second wave: 1995–1996, and third wave: 1997–1998. Data for one clustering variable (perceived receipt of instrumental and emotional support) used in the Smith and Baltes study was neither available at the second wave of data collection nor at the third wave. All other constructs were assessed at the three measurement occasions in BASE in exactly the same way.

Cognitive functioning was assessed using a touch-screen computerized battery of intelligence tests. Three factor (ability) scores of psychometric intelligence were computed to represent the two-component model of intelligence (P. B. Baltes, 1987; Cattell, 1971; Horn, 1982): Performance on perceptual speed and memory characterized fluid abilities of intelligence, and performance on knowledge characterized crystallized abilities of intelligence (for details, see Lindenberger & Baltes, 1997; Lindenberger, Mayr, & Kliegl, 1993; T. Singer et al., 2003). Each ability was represented by a unit-weighted composite of two tests. Perceptual Speed was measured by the Digit-Letter Test and Identical Pictures. Memory was measured by Memory for Text and a Paired-Associates Task. Knowledge was measured by Spot-a-Word and a vocabulary test. It has to be noted that, in the total cross-sectional analysis at baseline assessment as reported by Lindenberger & Baltes (1997), each ability was measured by three tests. Due to time restrictions at follow-up, only two tests per ability were administered both at the second wave and at the third wave. Results reported in the present study are based on unit-weighted composite measures for each ability that are based on two tests.

To assess *personality* dispositions, items were selected from the NEO (Costa & McCrae, 1985; for details, see Smith & Baltes, 1999). Neuroticism was derived from responses to six items assessing the facets of anxiety, depressivity, vulnerability, and hostility. Extraversion was derived from responses to six items assessing the facets of gregariousness, positive emotionality, assertiveness, and activity.

Self-related functioning was measured using two factor scores representing general control beliefs (internal and external control) and one factor score of goal investment (for details, see Kunzmann et al., 2000; Smith & Baltes, 1999; Staudinger, Freund et al., 1999). Internal control refers to the extent to which individuals believed that the good things in life are due to their own actions. External control represents beliefs that the actions of other people determine what happens to oneself. Based on seminal work from Levenson (1981), the BASE project (Smith, Marsiske, & Maier, 1996) developed a scale measuring internal control with three items and external control with four items. Goal investment indexed personal engagement (investment) in projects and goals. Participants were asked to rate the time and effort they currently invested into 10 areas of life including health, well-being of close relatives, mental performance, relationships with friends and acquaintances, thinking about life, hobbies and interests, independence, death and dying, occupational or comparable activities, and sexuality.

Social integration was assessed using two loneliness measures and the reported number of close confidants (for details, see Lang, 2000; Smith & Baltes, 1999; Wagner, Schütze, Lang, 1999). The literature suggests that feelings of social isolation and emotional distance from other persons constitute two distinct dimensions of loneliness (Smith & Baltes, 1999; Weiss, 1982). Accordingly, eight items from the UCLA Loneliness scale (Russell et al., 1984) were selected to assess social loneliness (perceptions of belonging to a social group and the general availability of trusted others) and emotional loneliness (feelings of isolation, being alone, and being secluded from contact with others). The Circle Task of Kahn and Antonucci (1986) was used to measure the reported number of close confidants. Participants were shown a diagram of concentric circles and they were told that they stood in the center. They were then asked to name the people they considered to be extremely close and important in their lives (first inner circle). The outer two circles (second circle: Somewhat less close people; and third circle: Still important, but more distant people) were not considered in this analysis.

Measures of personality and self-related functioning, and social integration were obtained in individual tape-recorded interviews with a verbal response format. For the latter measures, participants were asked to indicate how well items described them using a five-point Likert-scale with 1 labeled as 'does not apply to me at all' and 5 labeled as 'applies very well to me.' Each item was read aloud by trained research assistants.

3. 2. 1. 1 Data Preparation and Reliability

The first step in data preparation involved examining and dealing with missing data in the longitudinal sample. The highest percentage of missing data was found for the cognitive tests. Overall, 367 out of a total of 2340 attainable data points (i.e., 130 participants x 6 tests x 3 occasions), or 16%, were missing from the cognitive data set. More specifically, between 1.5% and 12% were missing for the Digit Letter over time, 6% – 8% for Identical Pictures, 0% – 8% for Memory for Text, 0% – 7% for Paired Associates, 3% – 5% for Spot-A-Word, and 0% – 10% for Vocabulary. For all other measures, 'missingness' was below 3%. For the cognitive measures, there were two main reasons for missing data. First, most research participants who produced missing data on cognitive measures had very poor vision so that computerized testing was totally or partly impossible (see also Lindenberger & Baltes, 1997). Second, some of the participants were intellectually unable to understand the instructions for some of the tests.

Following the procedures that were applied in the research group of BASE, missing data were estimated by linear regression analyses (Rovine & Delaney, 1990). Specifically, missing values were regressed by age, gender, and, if available, data for other items measuring the same underlying construct. For example, the estimated value for the perceptual speed indicator of Digit Letter was not only based on the age and gender regression weights, but also on the regression weight of the other speed indicator, Identical Pictures (for details, see also Kunzmann, 1997; Lindenberger & Baltes, 1997). Using this strategy is less likely to result in distortions of the data structure than listwise deletion (Beale & Little, 1975). All of the data reported replaced missing data through estimates based on linear regression. A routinely check revealed virtually the same results when using listwise deletion rather than replacement through regression estimates.

In a second step of data preparation, both psychological and cross-disciplinary measures were scaled by a linear transformation such that scores at the first measurement occasion for the total sample conform to a *T* metric with a mean of 50 and a standard deviation of 10. On the four dimensions of neuroticism, external control beliefs, social loneliness, and emotional loneliness scores were reverse-coded so that higher scores consistently refer to higher functionality (desirability) for adjustment in old age. Details about the tests and the wording of items can be obtained from Appendix B (see Section B.1).

In a third step, issues of reliability and test-retest stability were addressed. The reliability of the cognitive tests has been demonstrated to be high. Lindenberger and colleagues (1993) report the following Cronbach's alphas: Digit-Letter Test: $\alpha = .93$; Identical Pictures: $\alpha = .91$; Memory for Text: $\alpha = .59$; Paired-Associates: $\alpha = .87$; Spot-a-Word: $\alpha = .91$; Vocabulary: $\alpha = .80$. With the exception of goal investment and number of close others, the constructs of personality and self-related functioning, and social integration were based on item parcel scores rather than items. There is an extensive literature to demonstrate that such procedure is advantageous in several respects: (a) Parcel scores can be expected to have greater reliability and generality, (b) response biases and other characteristics that are specific to single items are likely to be of less importance, and (c) the distributions of variables are less likely to cause problems (e.g., Kishton & Widaman, 1994; Marsh, Antill, & Cunningham, 1989; cf. Kunzmann, 1997). There are a number of different procedures that can be used to collapse items into parcels. In the context of BASE, it was decided to represent each construct by three indicators (e.g., because this is an optimal situation in confirmatory factor analyses) and it was opted for an assignment of items to parcels that combined those items with the most extreme reliabilities. For example, perceived others control was measured by four items, and the item with the highest reliability was combined with the item with the lowest reliability to form a common parcel, while both other items defined a parcel on their own.[1]

Test-retest-stability coefficients for the 11 profile-defining psychological measures over time can be obtained from Table 6. These correlations range from r = .38 for number of close others over the 4-year period between baseline assessment and the second wave to r = .75 for perceptual speed over the 2-year period between the last two occasions as well as for knowledge between baseline and the second wave. For the domain-specific composite measures of psychological functioning, test-retest correlations were particularly substantive (i.e., range from r = .64 for social integration between baseline and the third wave to r = .84 for cognition between the second wave and the third wave). With the exception for number of close others, test-retest stability coefficients can be regarded satisfactory. Low stability for number of close others may be interpreted with regard to both measurement properties and conceptual arguments. Methodologically, low reliability may be due to the special format of a single item asked rather than using several items that are more reliable (Spearman-Brown formula; Brown, 1910; Spearman, 1910). Conceptually, low stability may simply reflect the fact that old age has been shown to be associated with social losses; dealing with and compensating for such losses can be considered one of the major regulative achievements in this phase of life (Carstensen, 1991; Lang, 2001). In the context of the longitudinal analyses in BASE, an intensive amount of work has been devoted to demonstrate the invariance of measures across time (e.g., factorial invariance of control beliefs: Kunzmann, 1997; factorial invariance of personality: Lißmann, 2002).

[1] Cronbach's alphas for each scale when using items rather than parcels can be obtained from Appendix B (see Section B.3).

Table 6
Test-Retest Correlations Over Time for the 11 Profile-Defining Psychological Constructs Entered Into the Cluster Analysis, and for Composite Measures of Desirability: 6-Year Longitudinal BASE Sample

	Test-Retest Correlation		
Constructs	Baseline – 2nd wave	2nd wave – 3rd wave	Baseline – 3rd wave
Profile-Defining			
1. Speed	.68	.75	.68
2. Memory	.66	.73	.73
3. Knowledge	.75	.73	.74
4. Neuroticism	.58	.74	.62
5. Extraversion	.69	.73	.68
6. Internal Control	.51	.48	.34
7. External Control	.60	.64	.60
8. Goal Investment	.59	.59	.50
9. Social Loneliness	.62	.66	.53
10. Emotional Loneliness	.68	.68	.60
11. Close Others	.38	.52	.42
Composite			
Cognition	.77	.84	.82
Self & Personality	.72	.75	.67
Social Integration	.71	.74	.64
Overall Desirability	.79	.83	.78

Note. $N = 130$. All correlations were significant at $p < .01$ or below.

3. 2. 2 Cross-Disciplinary Measures

Details about the assessment procedure for the *cross-disciplinary* constructs can also be obtained from P. B. Baltes and Mayer (1999; see also Appendix B for Section B.2). Life-history and socio-cultural status was measured using a unit-weighted composite of three measures: Equivalent income, occupational prestige, and number of years of education (for details, see Mayer, Maas et al., 1999). Equivalent income represents the net household income weighted by the number of people sharing the household. Occupational prestige was based on a standard rating scale in Germany (Wegener, 1985), which indicates the participant's last occupation before retirement. If individuals were not part of the labor force, the prestige of the last occupation of the spouse (former spouse if widowed) was used.

Two measures were used to index sensory functioning (for details, see Marsiske et al., 1999; Lindenberger & Baltes, 1997). Auditory acuity was measured separately for each ear by using a Bosch ST-20-1 pure-tone audiometer with headphones. To estimate hearing ability, an inverted average score of thresholds in dB across both ears and four frequencies (1.00, 2.00, 4.00, and 6.00 kHz) was computed. Visual acuity represents a composite based on the unit-weighted mean of close vision acuity and distance visual acuity. Close visual acuity was measured separately for both eyes with a reading table presented at reading distance (Geigy Pharmaceuticals, 1977); distance visual acuity was assessed binocularly with a reading table presented at a standard distance of 2.5 m to the participant. Visual acuity was measured both with and without the best optical correction provided by the participant (i.e., corrective glasses); the best value of the two was used in the present analyses.

BASE assessment also included comprehensive medical examinations and laboratory tests carried out by physicians (for details, see Steinhagen-Thiessen & Borchelt, 1999). The index of multimorbidity represents the number of medically diagnosed moderate to severe chronic illnesses. Clinical diagnosis of dementia was determined according to *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R;3rd ed., rev; American Psychiatric Association [APA], 1987) criteria using standard clinical interview and assessment procedure. An experienced clinician who was unaware of the results of the cognitive and neuropsychological assessments made the diagnosis (for details, see Helmchen et al., 1999). Measures of Activities of Daily Living/Instrumental Activities of Daily Living (ADL/IADL) were self-reports on a standardized instrument (Katz, Downs, Cash, & Grotz, 1970; Lawton & Brody, 1969).

Overall subjective well-being was measured using a German translation of the Philadelphia Geriatric Morale Scale (PGCMS), which was specifically designed for use with older adults (Lawton, 1975; for details, see Smith, Fleeson, Geiselmann, Settersten, & Kunzmann, 1999). The scale comprised items for three dimension: Non-agitation (6 items), aging satisfaction (5 items), and life satisfaction (4 items). Cronbach's alphas were high for the scale at each measurement occasion (baseline: $\alpha = .80$; second wave: $\alpha = .84$; third wave: $\alpha = .86$) and the test-retest-stability-coefficients were satisfactory (baseline – second wave: $r = .69$; second wave – third wave: $r = .72$; baseline – third wave: $r = .60$). In a number of control analyses, age was used as a categorical variable contrasting participants in their seventies at baseline assessment ($n = 84$) versus those older than age 80 ($n = 46$).

Information about mortality status and date of death for deceased participants were obtained from the City Registry (update from July, 11th , 2002). Mortality status was missing for

6 participants because they had moved out of the Berlin area. These individuals were not considered in the mortality analyses. Overall, 52 (42%) participants were deceased by July 2002 and 72 were still alive.

3. 3 Differential Sample Attrition, Measurement Issues and Their Implications for the Present Study

Previous work in BASE has demonstrated that there was substantive sample attrition over the successive measurement occasions (e.g., T. Singer et al., 2003). For example, the 6-year longitudinal BASE sample represented only 25% of the heterogeneous and locally representative total sample of BASE and 55% of the survivors of this sample (*n* = 239). Such attrition can have major implications for the extraction of new clusters as well as for the type and amount of change that can be observed. For that reason, attrition analyses were carried out preliminary to the main analyses reported in the dissertation. Results from the analyses serve as a basis for adequately interpreting subgroup differences in psychological profiles and change over time.

First, sample differences at the subgroup level are reported. The analyses extend the previous literature on differential sample attrition by a systemic-wholistic approach. They determine to what extent the subgroups identified in the total cross-sectional BASE sample by Smith and Baltes (1997) were differentially vulnerable to drop out of the study. Effects of sample attrition over six years were expected to be strongest among the less desirable profile subgroups from the total cross-sectional BASE. This is followed by examining sample differences at the level of the subgroup-defining measures. Effects of sample attrition over six years were expected for cognitive functioning as well as for measures of personality and self-related functioning, and social integration. Attrition analyses were carried out to determine differences both at the mean level and in the covariance structure. Examining these questions extends previous work from BASE by including three waves of data collection (rather than two waves) and by using psychological measures across different domains including personality and self-related functioning, and social integration (rather than cognitive measures only).

3. 3. 1 Attrition Effects at the Subgroup Level

In the total cross-sectional BASE sample, Smith and Baltes (1997) empirically identified nine subgroups, which differed from one another substantively in their psychological profiles (see Section 2.1.1 for Table 2). In an attempt to evaluate the findings, Smith and Baltes (1997) heuristically established a criteria that resulted in an almost equal distribution of desirable and less desirable profile subgroups: Based on the overall mean across all subgroup-defining measures (12 constructs in total), the subgroups were rank-ordered. Four of the nine subgroups were above the sample mean on the so-defined overall measure of functioning/desirability and were thus referred to as representing desirable psychological profiles. The other five subgroups were below the sample mean and were thus considered as representing less desirable psychological profiles. Panel A of Figure 4 shows these groups pooled and indicates that the criteria split up the total cross-sectional BASE sample into two almost equal halves (*n* = 240 vs. *n* = 270).[2]

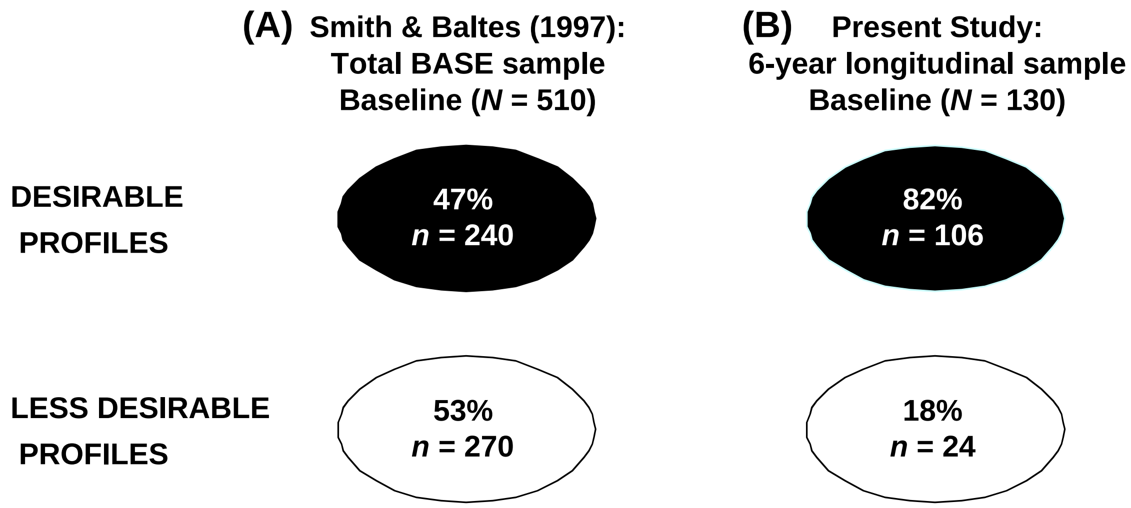

Figure 4 Positive selection of the 6-year longitudinal BASE sample: Subgroup level.
Note. Effects of differential sample attrition: For the total BASE sample, Smith and Baltes (1997) established criteria that resulted in an almost equal distribution of desirable and less desirable profile subgroups (47% vs. 53%). In the present study, using the same criteria for the 6-year longitudinal sample revealed that 82% of these participants were in desirable profile subgroups, whereas only 18% were in the less desirable profile subgroups. Note that Smith and Baltes (1997) identified four subgroups with desirable profiles and five subgroups with less desirable profiles. For clarity, these subgroups were pooled.

To investigate sample attrition at the subgroup level, the question is to what extent these subgroups were differentially likely to be part of the 6-year longitudinal sample. Accordingly, I examined how many participants from each subgroup were left in the longitudinal sample. As can be seen from Panel B of Figure 4, the 6-year longitudinal sample showed a considerable degree of positive sample attrition at the subgroup-level. Consistent with the prediction, the effects of sample attrition were strongest among the less desirable profile subgroups. Put differently, BASE participants who were part of the 6-year longitudinal sample were about four

[2] In the Smith and Baltes study (1997), six participants were excluded from cluster analysis because they were identified as multivariate outliers. For the same reason, those two cases who remained in the longitudinal sample were also excluded.

times more likely to come from the desirable profile subgroups in the total sample (82%) than to come from less desirable profile subgroups (18%), χ^2 (1, *N* = 510) = 83.3, *p* < .001. In both subgroups, mortality-associated attrition was the major factor underlying attrition. Among participants who were *not* available for repeated assessment, 64% in the desirable subgroups (*n* = 86) and 77% in the less desirable subgroups (*n* = 189) had died over the 6-year period.

From these differences, it follows that (1) longitudinal data in BASE are, in essence, restricted to different variants of successful aging, (2) the effects of sample attrition cannot only be found at the variable level, but also at the subgroup level, and (3) it is highly unlikely that the nine subgroups previously identified in the total cross-sectional BASE sample (Smith & Baltes, 1997), particularly those at the lower end of the desirability spectrum, can be observed in the present longitudinal sample. If one were to generalize these findings about differential sample attrition, one could conclude that longitudinal samples likely miss participants with profiles of lower psychological functioning/desirability across different domains. This finding is particularly important as it extends previous research (e.g., Cooney, Schaie, & Willis, 1988): Attrition effects are not only associated with single variables of cognitive functioning and socio-economic status, but also generalizes to multiple domains and profiles of psychological functioning.

3. 3. 2 Attrition Effects at the Variable Level

In a second step of preliminary analyses, I examined to what extent the measurement space and data structure at baseline assessment differed between the total cross-sectional BASE sample (*N* = 510) and the 6-year longitudinal sample (*n* = 130). To determine sample attrition effects at mean levels of performance, the amount of total selectivity (attrition) was examined and decomposed into mortality-associated and experimental attrition based on a method described by Lindenberger et al. (2002). To examine the statistical significance of differences in the covariance structure, analyses compared participants who took part three times in BASE with those who did not (i.e., 6-year drop-outs; *n* = 380). Sample attrition effects on mean levels at baseline assessment are reported first, followed by presenting differences in the covariance structure.

3. 3. 2. 1 Sample Differences at the Mean Level

In the context of the design of BASE, Lindenberger et al. (1999, 2002; see also T. Singer et al., 2003) distinguished mortality-associated and experimental selectivity. *Mortality-*

associated selectivity refers to differences between individuals who survived and those who died. *Experimental* selectivity, on the contrary, relates to differences among individuals who survived: Persons who participated in a study and those who declined participation due to unwillingness or incapability. Mortality-associated selectivity reflects a (natural) population process that does not compromise the validity of observations.[3] Experimental attrition, in contrast, leads to a nonrandom subsample of the surviving population and thus limits generalization of research findings (P. B. Baltes et al., 1977). To compute these components, data from baseline assessment for the 6-year longitudinal BASE sample (n = 130) was compared to those of individuals still alive at the third occasion (n = 239), and by comparing these survivors with data from the total cross-sectional sample (N = 510). Differences between the three groups were standardized on the standard deviation units of the total cross-sectional sample; following Lindenberger et al. (2002):

total selectivity = $(M_{select} - M_{parent}) / SD_{total\ sample}$;

experimental selectivity = $(M_{select} - M_{survivor}) / SD_{total\ sample}$;

mortality-associated selectivity = $(M_{survivor} - M_{parent}) / SD_{total\ sample}$.

It has to be noted that the effect size is a descriptive measure that is derived directly from the group level, so that there is no variance associated with it. For that reason, it is not possible to apply significance tests.

Figure 5 displays the observed selectivity effects. In analogy to expectation, positive selection effects were found across a number of different measures of psychological functioning and cross-disciplinary correlates. Among the measures defining the psychological profile (see Figure 5, Panels A through C), the magnitude of total selectivity was 0.77 *SD* units for perceptual speed, 0.57 *SD* units for memory, 0.50 *SD* units for knowledge, 0.14 *SD* units for neuroticism, 0.27 *SD* units for extraversion, – 0.09 *SD* units for internal control beliefs, 0.38 *SD* units for external control beliefs, 0.22 *SD* units for goal investment, 0.20 *SD* units for social loneliness, 0.34 *SD* units for emotional loneliness, and 0.19 *SD* units for close others. According to statistical convention (e.g., Cohen, 1977), observed selectivity corresponds to medium effects for intellectual functioning and small effects for measures of personality and self-related functioning, and social integration. Of the total amount of observed selectivity in most psy-

[3] Some demographers might argue that this ‘validity’ depends on cause of death. For example, prior to age 85 cause of death primarily represent extrinsic factors including diseases, whereas after age 85 causes of death primarily represent natural factors and can thus be considered intrinsic to aging per se (Manton, 1990; Manton & Stellard, 1990).

chological measures, the majority was due to mortality. For perceptual speed, for example, the relative amount of mortality-associated selectivity was 64%.[4]

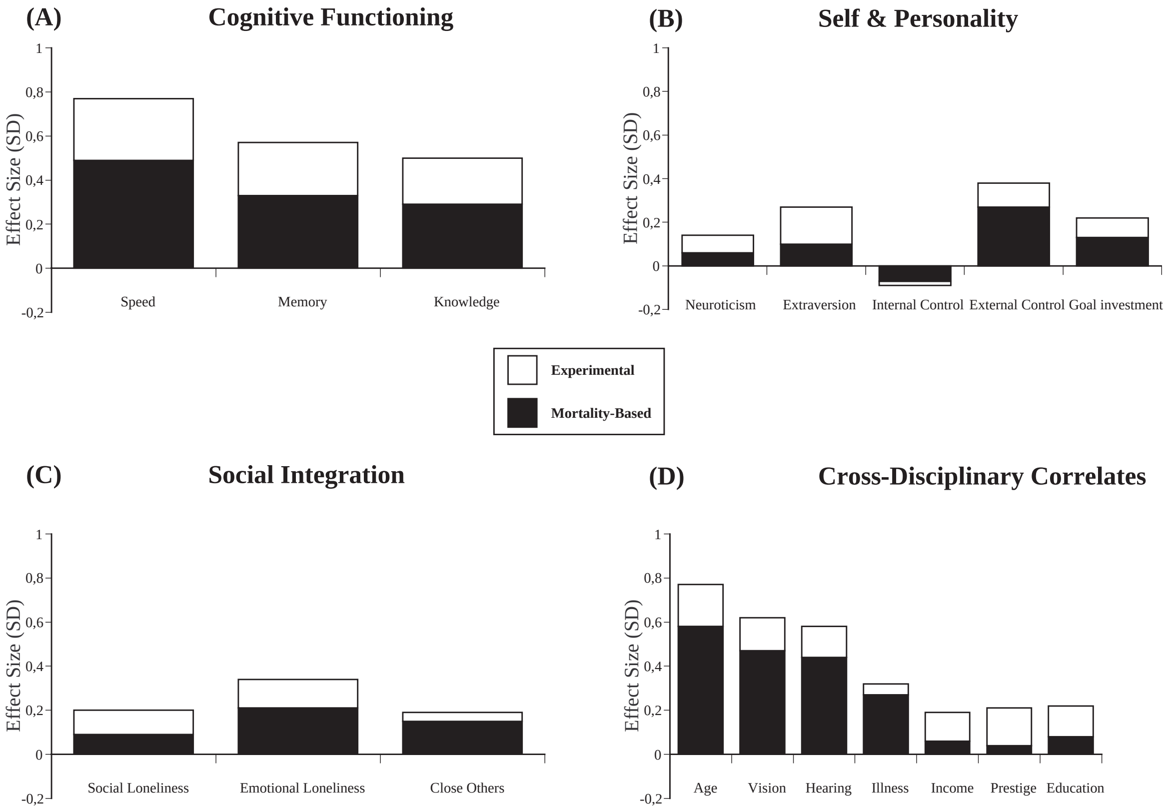

Figure 5 Positive selection of the 6-year longitudinal BASE sample: Variable level - Mortality-associated and experimental selectivity effects for measures of cognitive functioning (A), personality and self-related functioning (B), social integration (C), and cross-disciplinary correlates (D).

Note. Effects of sample attrition on mean level differences at baseline assessment were not only found for cross-disciplinary correlates and cognitive functioning, but also generalize to other measures of psychological functioning such as personality and self-related functioning, and social integration. For the majority of constructs, mortality-associated selectivity effects were stronger than those of experimental selectivity.

Effects of sample attrition were also found for the cross-disciplinary constructs (see Figure 5, Panel D). The magnitude of total selectivity was 0.77 *SD* units for chronological age, 0.62 *SD* units for visual abilities, 0.58 *SD* units for hearing abilities, 0.32 *SD* units for multimorbidity, 0.19 for *SD* units for income, 0.21 for *SD* units for occupational prestige, and 0.22 *SD* units for education. According to convention (e.g., Cohen, 1977), the observed effect sizes were medium for age and sensory functioning, and correspond to small effects for multimorbidity and the indices of income, prestige, and education. With the exception of the three

[4] Follow-up analyses indicated that selection effects were particularly strong among the oldest-old participants suggesting an increased age effect on selection (see also Lindenberger et al., 2002; T. Singer et al., 2003).

life-history factors, most selectivity was due to mortality (e.g., multimorbidity: 84%). For the life-history factors, only between 19% and 36% were mortality-associated selectivity. In sum, these subsample differences indicate that the 6-year longitudinal sample had a much higher average level of functioning at baseline assessment as compared to the rest of the total cross-sectional sample, particularly those who died in the mean time. The effects of sample attrition over time are on top of the fact that the total cross-sectional BASE sample already reflected a slight positive selection at baseline assessment for cognition, personality and self-related functioning, and social integration as well as health and mortality (Lindenberger et al., 1999).[5]

3. 3. 2. 2 Sample Differences in the Covariance Structure

In a final set of preliminary analyses, subsample differences in the covariance structure were determined. The objective was to examine whether the measurement space differed between the total cross-sectional BASE sample and the 6-year longitudinal sample. To do so, three subsamples of BASE were compared to one another: The total cross-sectional BASE sample (N = 510), the 6-year longitudinal sample (n = 130), and the 6-year drop-outs (n = 380). The latter group was necessary to test for statistical significance of differences in the covariance structure.

Table 7 shows the zero-order (rows A) and age-adjusted (rows B) intercorrelations of the 11 measures used to define the psychological profile across the three subsamples. From rows A of Table 7 can be seen that the data set of the 6-year longitudinal BASE sample is characterized by (a) lower age correlations for the cognitive variables, (b) lower intercorrelations among the cognitive variables, and (c) lower correlations between most cognitive and non-cognitive variables. For example, (a) the age correlation for knowledge drops from r = –.34 in the total cross-sectional sample to a non-significant value of r = –.03 in the 6-year longitudinal sample; (b) intercorrelations among the cognitive variables ranged between r = .55 and r = .64 in the total cross-sectional sample and dropped down to r = .39 and r =.43 in the 6-year longitudinal sample; and (c) the correlation between perceptual speed and external control was r = .36 in the total cross-sectional sample, but was only r = .16 in the 6-year longitudinal sample.

[5] Lindenberger et al. (1999) reported that selection effects were present at baseline assessment and amounted to half a standard deviation. This was within the acceptable range to conclude that the total cross-sectional BASE sample was locally representative.

Table 7
Intercorrelations of the 11 Profile-Defining Psychological Constructs Entered Into the Cluster Analysis: Comparing Different BASE Samples at Baseline Assessment – Total Cross-Sectional Sample (N = 510), 6-Year Longitudinal Sample (n = 130), and 6-Year Drop-Outs (n = 380)

Construct		Age			1			2			3		
		T	L	D	T	L	D	T	L	D	T	L	D
1. Speed	A	**-.58**	**-.30**$_a$	**-.51**$_b$	–								
	B	–			–								
2. Memory	A	**-.43**	**-.22**$_a$	**-.36**$_a$	**.62**	**.43**$_a$	**.61**$_b$	–					
	B	–			**.51**	**.39**$_a$	**.52**$_a$	–					
3. Knowledge	A	**-.34**	-.03$_a$	**-.28**$_b$	**.64**	**.39**$_a$	**.63**$_b$	**.55**	**.39**$_a$	**.54**$_b$	–		
	B	–			**.57**	**.40**$_a$	**.60**$_b$	**.48**	**.39**$_a$	**.49**$_a$	–		
4. Neuroticism[a]	A	-.08	-.12$_a$	-.04$_a$	**.15**	.13$_a$	**.14**$_a$	.09	.03$_a$	.07$_a$	**.21**	**.18**$_a$	**.21**$_a$
	B	–			**.13**	.09$_a$	**.13**$_a$	.06	.01$_a$	.06$_a$	**.20**	**.18**$_a$	**.20**$_a$
5. Extraversion	A	**-.19**	-.10$_a$	**-.15**$_a$	**.23**	**.29**$_a$	**.15**$_a$	**.13**	**.20**$_a$	.04$_a$	.03	-.02$_a$	-.01$_a$
	B	–			**.15**	**.27**$_a$	.09$_b$	.05	**.18**$_a$	-.01$_b$	-.04	-.02$_a$	-.06$_a$
6. Internal Control	A	.02	-.14$_a$	.02$_a$	-.06	-.07$_a$	-.04$_a$	**-.09**	-.06$_a$	-.08$_a$	**-.20**	**-.24**$_a$	**-.18**$_a$
	B	–			-.06	-.11$_a$	-.03$_a$	**-.09**	-.09$_a$	-.08$_a$	**-.20**	**-.25**$_a$	**-.18**$_a$
7. External Control[a]	A	**-.32**	-.16$_a$	**-.27**$_a$	**.36**	.16$_a$	**.32**$_b$	**.26**	**.22**$_a$	**.20**$_a$	**.25**	.09$_a$	**.22**$_a$
	B	–			**.22**	.12$_a$	**.22**$_a$	**.14**	**.20**$_a$	**.11**$_a$	**.16**	.08$_a$	**.15**$_a$
8. Goal Investment	A	**-.17**	**-.17**$_a$	**-.13**$_a$	**.24**	**.24**$_a$	**.20**$_a$	**.17**	.06$_a$	**.16**$_a$	**.17**	.16$_a$	**.14**$_a$
	B	–			**.18**	**.21**$_a$	**.16**$_a$	**.11**	.02$_a$	**.12**$_a$	**.12**	.16$_a$	**.11**$_a$
9. Social Loneliness[a]	A	**-.13**	**-.17**$_a$	-.08$_a$	**.20**	.06$_a$	**.20**$_a$	**.21**	.11$_a$	**.21**$_a$	**.18**	.13$_a$	**.16**$_a$
	B	–			**.16**	.01$_a$	**.19**$_b$	**.17**	.08$_a$	**.20**$_a$	**.14**	.13$_a$	**.14**$_a$
10. Emotional Loneliness[a]	A	**-.29**	**-.28**$_a$	**-.21**$_a$	**.29**	**.17**$_a$	**.24**$_a$	**.22**	**.18**$_a$	**.16**$_a$	**.25**	.15$_a$	**.22**$_a$
	B	–			**.15**	.09$_a$	**.16**$_a$	**.11**	.12$_a$	.09$_a$	**.17**	.15$_a$	**.17**$_a$
11. Close Others	A	**-.14**	-.15$_a$	-.09$_a$	**.17**	.04$_a$	**.17**$_a$	**.15**	-.03$_a$	**.18**$_b$	**.12**	-.04$_a$	**.13**$_b$
	B	–			**.11**	-.01$_a$	**.14**$_b$	**.10**	-.07$_a$	**.16**$_b$	.08	-.04$_a$	**.11**$_a$

Note. T = Total cross-sectional sample (*N* = 510), L = 6-year longitudinal sample (*n* = 130), D = 6-year drop-outs (*n* = 380). A = zero-order correlation; B = age-partialed correlation. Six BASE participants were excluded from cluster analyses and were thus not considered. [a] Scores on these dimensions were reverse-coded to calculate the desirability (functional status) score.
Correlations different from zero at $p < .05$ or below in bold. Correlations with different subscripts differ at $p < .05$ or below, one-tailed.

The overall pattern of findings in the covariance structure was in line with the expectation that positive selection effects can be seen across a number of different measures of psychological functioning: The covariance structure among the group-defining measures at baseline assessment was of lesser magnitude in the longitudinal BASE sample as compared wit both other samples. Differences were strongest on cognitive measures, but were also found on measures of personality and self-related functioning, and social integration.

Without speculating too much about potential mechanisms underlying these findings, one might argue that structural differences in the subgroup-defining measures are a result of age differences between the samples. Table 8 shows that the mean age of the 6-year longitudinal BASE sample at baseline was 78 years as compared to 85 years for the total cross-sectional sample and 87 years for those who dropped out over time. In addition, the longitudinal sample had less age variation (*SD* = 6.0) than both the total sample (*SD* = 8.6) and the drop-out sample (*SD* = 8.3). It has long been established that variance restrictions (e.g., in chronological age) reduce the magnitude of the covariation (e.g., Nesselroade & Thompson, 1995; see also Lindenberger et al., 1999). Another factor reducing variance and so contributing to reduced magnitude of the covariance structure is the loss of subjects due to experimental and biological mortality. Proposals of dedifferentiation in old age (e.g., P. B. Baltes et al., 1980) suggest that participants in advanced old age would have contributed to a more homogenous psychological structure. However, attrition effects were more pronounced among the oldest-old as compared to the young-old (T. Singer et al., 2003) so that it were particularly participants from the older age-cohort who dropped out over time and were not available for repeated testing (Smith & Delius, 2003; see Section 3.1 for Figure 3).

To examine such post-hoc interpretation, covariance structures of the samples were examined that were age-free, that is age-partialed. Results are shown in Rows B of Table 7. Findings indicate that not all correlation differences between the samples vanished after age was partialed out (particularly among the cognitive variables), which suggests that subsample differences were not exclusively due to differences in chronological age.[6] For example, perceptual speed and knowledge showed much lower intercorrelations among participants from the 6-year longitudinal sample as compared to those who dropped out of BASE, even after differences in chronological age were partialed out. This pattern suggests that old age might be just one of several driving forces underlying the altered correlation pattern. As can be seen in Figure 5 and in Table 8, the largest selectivity effects at the mean level were observed for the cognitive variables, external control beliefs, and emotional loneliness. This suggests that baseline participants who were cognitively impaired, perceived themselves as externally controlled and emotionally lonely had a higher risk of sample attrition over time. To the extent to which

[6] One could reasonable argue that differences may also be a function of dementia (see Sliwinski, Hofer, Hall, Buschke, & Lipton, 2003). After covarying out the effects of dementia diagnosis, a similar pattern was found: Most, but not all differences in the correlation pattern vanished, which suggests that neither age nor dementia alone accounted for these differences.

this was the case, aging processes, cognitive impairment, and other less desirable features of old age such as loneliness did not play a major role at baseline assessment for the highly select group of 6-year longitudinal BASE participants.

Table 8
Descriptive Statistics for the 11 Profile-Defining Psychological Constructs Entered Into the Cluster Analysis, for Composite Measures of Desirability, and for Age: Comparing Different BASE Samples at Baseline Assessment – Total Cross-Sectional Sample (N = 510), 6-Year Longitudinal Sample (n = 130), and 6-Year Drop-Outs (n = 380)

	M			*SD*		
	Samples in BASE			Samples in BASE		
Measure	*N* = 510	*n* = 130	*n* = 380	*N* = 510	*n* = 130	*n* = 380
Profile-Defining						
1. Speed	50.0	57.6_a	47.3_b	*10.0*	7.4_a	9.4_b
2. Memory	50.0	55.6_a	48.0_b	*10.0*	9.7_a	9.3_a
3. Knowledge	50.0	54.9_a	48.3_b	*10.0*	8.1_a	10.0_b
4. Neuroticism *	50.0	51.3_a	49.5_a	*10.0*	9.0_a	10.3_b
5. Extraversion	50.0	52.6_a	49.1_b	*10.0*	8.9_a	10.2_a
6. Internal Control	50.0	49.1_a	50.3_a	*10.0*	9.0_a	10.3_a
7. External Control *	50.0	53.8_a	48.7_b	*10.0*	8.3_a	10.2_b
8. Goal Investment	50.0	52.1_a	49.2_b	*10.0*	8.4_a	10.4_a
9. Social Loneliness *	50.0	51.9_a	49.3_b	*10.0*	9.3_a	10.2_a
10. Emotional Loneliness *	50.0	53.3_a	48.8_b	*10.0*	9.5_a	9.9_a
11. Close Others	50.0	51.2_a	49.3_a	*10.0*	9.2_a	9.7_a
Composite						
Overall Desirability	50.0	53.0_a	48.8_b	*4.9*	4.2_a	4.7_a
Cognition	50.0	56.1_a	47.9_b	*8.5*	6.6_a	8.1_b
Self & Personality	50.0	51.8_a	49.3_b	*5.2*	4.7_a	5.2_a
Social Integration	50.0	52.1_a	49.0_b	*6.9*	6.9_a	6.8_a
Age	84.9	78.3_a	87.2_b	*8.6*	6.0_a	8.3_b

Note. * Scores on these dimensions were reverse-coded to calculate the desirability (functional status) score. Six BASE participants were excluded from cluster analysis and were thus not considered in this Table. Apart from age, all variables were standardized to the *T* metric.
Means (*M*) and standard deviations (*SD*), respectively, in the same row that do not share subscripts differ at the $p < .05$ level or below. Mean level differences remained significant after age had been partialed out for each cognitive variable, external control, and each composite measure of psychological functioning.

Coming back to the effects of variance restrictions on the magnitude of correlations, it has to be acknowledged that the 6-year longitudinal BASE sample had not only less variance in age, but also in perceptual speed, knowledge, neuroticism, and external control (see Table 8).

From this follows that it is an open question whether or not the relative variance restrictions found in the longitudinal BASE sample have primarily caused the reductions in intercorrelations. However, it would also be possible to argue against such concern. Comparing Table 7 and Table 8 with one another reveals that findings of lower age correlations and intercorrelations for the profile-defining measures were not restricted to these more homogenous measures, but also have been found on measures that do not show subsample differences in variability (e.g., memory, close others). To come to a close, it seems reasonable to infer that the patterning of differences between the total cross-sectional BASE sample and the 6-year longitudinal sample reflected a complex function of sample attrition processes and changes in the age distribution rather than being exclusively due to variance restrictions. Sample attrition seems to be a domain-generalized phenomenon and its interplay with functional dedifferentiation of psychological profiles is worth further consideration.

3. 3. 3 Implications for the Present Study

What implications arise from the positive sample selection of the longitudinal sample and its restricted age range for the steps planned in the current study? First, longitudinal data in BASE is limited to profiles of psychological functioning that represent different forms of aging successfully. Second, it is not possible to examine stability and change over time for the subgroups identified previously in the total cross-sectional BASE sample (Smith & Baltes, 1997), particularly when operating in a *complete* data analysis framework. Instead, it is necessary to extract new subgroups by using baseline assessment data of the 6-year longitudinal BASE sample. Here, the objective is to link the newly extracted subgroups to those identified in the much larger and more heterogeneous total BASE sample. Because of these restrictions and differences in measurement space (i.e., means and covariance structure), the overlap between cluster groups from the total cross-sectional sample and the one generated from the 6-year longitudinal sample can only be partial. Instead, it is expected that subgroups identified in the 6-year longitudinal BASE sample primarily represent different profiles of desirable psychological functioning and that far fewer subgroups at the lower end of the functioning/desirability spectrum will be found (Q_{1a}). Differential sample attrition also has implications for the type and amount of change to be observed. For example, it may not be possible to fully test proposals about qualitative transitions in advanced old age (Q_{2b}) because the majority of the longitudinal sample is functioning too well to have reached some kind of lower functional limits.

Despite the limitations outlined, the key argument for employing a systemic-wholistic approach to investigate functional change over time in the context of BASE is as follows. Heterogeneity can be expected to be still preserved when one considers the positively select 6-year longitudinal BASE sample only. The four more desirable subgroups in the Smith and Baltes study (1997) can be interpreted to reflect the outcomes of conjoint processes of different biogenetics of aging as well as lifestyles and pathways into old age that all have resulted in highly functional profiles in old age. All four subgroups represent different profiles of successful aging. The underlying idea here is that given particular contexts, very different profiles may be more or less functionally effective though they are not the best possible. So, these subgroups have aged differently up to the (first) occasion of measurement and one could expect that they also age differently in the future (i.e., after recruitment in the longitudinal study). Based on this line of argument, it seems reasonable to expect in the present study that the subgroups extracted from the longitudinal sample will differ substantively from one another in psychological and cross-disciplinary profile characteristics (Q_1) and that questions about differential development in old age over the subsequent six years and its underlying mechanisms (Q_2) and consequences can be examined (Q_3).