Sample Selectivity and Generalizability of the Results of the Berlin Aging Study

Ulman Lindenberger, Reiner Gilberg, Todd D. Little, Reinhard Nuthmann, Ulrich Pötter, and Paul B. Baltes

In epidemiological investigations, one common but rarely analyzed threat to generalizability is sample selectivity or nonrandom sample attrition. In this chapter, we describe our approach to the study of selectivity and provide in-depth analyses of the magnitude of sample selectivity in the Berlin Aging Study. Of all individuals eligible for participation (the verified parent sample, $N = 1,908$), 27% reached the highest level of participation (the Intensive Protocol, $N = 516$). With respect to levels of performance, projection of selectivity observed on lower levels of participation onto Intensive Protocol constructs indicates that the Intensive Protocol sample was, indeed, positively selected on medical, social, and psychological dimensions. However, the magnitude of observed selectivity effects did not exceed 0.5 standard deviations for any construct. In addition, variances and covariance relations observed in the Intensive Protocol sample were not markedly different from those found at lower levels of participation. We conclude that the degree of selectivity in BASE fell within the usual range and did not result in a decrease of sample heterogeneity. Given the magnitude of sample attrition and the high mean age of the sample, this is a satisfactory result.

1   Introduction

A major goal in science is to ensure that the validity of empirical patterns does not remain restricted to the observed events, but can be generalized to a larger space of potential measurements. In this sense, *measurement representativeness* characterizes the degree to which observations can stand for other nonmeasured events (cf. McArdle, 1994).

In this chapter, our main intention is to document and analyze one important factor endangering representativeness and generalizability: the sample selectivity that can occur with sample loss or sample attrition. *Sample attrition* describes the fact that not all persons asked to take part in the Berlin Aging Study (BASE) passed through the entire assessment protocol. This sample loss can lead to *selectivity* (or bias) of the sample if participants differ from dropouts in characteristics relevant to the study (Kessler, Little, & Groves, 1995; Little, 1995; Little & Rubin, 1987). This selectivity associated with sample attrition should not be confused with selectivity related to the *sampling* procedure itself ("selective sampling" vs. "selective dropout"; cf. P. B. Baltes, Reese, & Nesselroade, 1988). The random sampling procedure used in BASE, by which every element has an equal chance of being selected, is generally accepted as being the best way of minimizing systematic sampling-related biases (Kruskal & Mosteller, 1979a, 1979b, 1979c).[1]

Independent of the degree of generalizability associated with *sampling* – which is large in BASE – is the risk of *sample attrition*, which makes it possible that the statements made on the basis of reduced samples can no longer be applied to the parent sample. In that case, sample attrition would lead to sample selectivity, and the generalizability of BASE results would need to be qualified. The main task of the analyses reported in this chapter is to determine the type and extent of sample selectivity and, in turn, contribute to the adequate interpretation of statements based on the BASE data set.

In the following, we first define our concept of selectivity. We then present an overview of the participation levels in BASE and introduce the statistical methods and variables used in the analyses. After reporting the results of the selectivity analyses, we finally discuss how the findings from the BASE Intensive Protocol data are influenced by sample selectivity.

2 Focuses of Interest

2.1 On the Examination of Sample Selectivity

In a study such as BASE, values on variables are assigned to persons and groups of persons. Depending on the topic of examination and the type of variable, these assignments can usually be summarized by statistical reference values such as means, frequency distributions (prevalence rates), variances, and correlations. The following examples illustrate this: (a) How large are the social networks of men and women? (mean); (b) What proportion of individuals aged 95 and above have dementia? (prevalence); (c) How large are individual differences in intellectual abilities among 70- to 80-year-olds? (variance); (d) How closely are sensory and intellectual functioning linked? (correlation).

The validity of these statements is initially restricted to those participants who were actually measured on the relevant variables. However, not all of the people selected for BASE went through the whole assessment sequence; therefore, the question arises as to whether the examination of the entire parent sample would have yielded different results. This would be the case if variables predicting dropout versus further participation in the study were systematically related to the variables under scrutiny (Little & Rubin, 1987). For example, persons with dementia could be less likely to reach the part of the assessment during which they would receive a clinical diagnosis of dementia (i.e., the Intensive Protocol). In this case, statements on dementia rates based on the Intensive

---

[1] Specifically, the verified sample of BASE is based on random sampling of addresses from the obligatory Berlin city register according to certain criteria. This form of random sampling is advantageous because systematic biases connected to the investigators' research interests are much less likely to occur than with nonrandom samples. However, even if 100% of the parent sample took part in the study (i.e., if there were no sample losses), there would be no guarantee that the persons selected by chance would represent a perfect "miniature" of the Berlin population aged 70 and above (cf. Kruskal & Mosteller, 1979a, 1979b, 1979c; Rendtel & Pötter, 1992; Rudinger & Wood, 1990). Therefore, questions about generalizability of the statements based on analyses of BASE data cannot be answered with a sweeping reply. Instead, generalizability will vary by domain, the type of statement made, and our current state of knowledge, and will sometimes refer to rather small (e.g., the elderly population of West Berlin) or larger (e.g., old people in Western industrial societies) groups (cf. P. B. Baltes et al., Chapter 1 in this volume).

Protocol would underestimate the prevalence rate in the parent sample (cf. Helmchen et al., Chapter 6 in this volume).

Thus, the analysis of sample selectivity represents a methodological precaution to reduce the likelihood of false conclusions and misleading generalizations. Do we overestimate educational levels because fewer people with a low education level take part in the study than those with a higher level? Do we underestimate the variability of intelligence because both good performers and persons with dementia participate less often than individuals with average abilities?

In trying to answer these questions, the *analysis* of sample selectivity is confronted with a fundamental paradox: In order to quantify optimally the degree and nature of selectivity, we would need precisely the information that is missing. We would have to know the characteristics of the nonparticipants – but if they provided this information, they would no longer be nonparticipants. In the methodological literature, this paradoxical situation has led to the demand that at least some basic pieces of information should be gathered on *all* persons, including potential nonparticipants (Dalenius, 1988; Esser, Grohman, Müller, & Schäffer, 1989; von Eye, 1989; Herzog & Rodgers, 1988; Oh & Scheuren, 1983; Panel on Incomplete Data, 1983; Tennstedt, Dettling, & McKinlay, 1992; Weaver, Holmes, & Glenn, 1975). An empirical examination of selectivity – without reference to external sources such as census information – is only possible when this condition is fulfilled.

For a study stretching over a longer period of time such as BASE, it makes sense to replace the dichotomy of participation versus nonparticipation by the graded concept of "participation levels" or *participation depth*. From the outset, the design of BASE was structured to allow the distinction of an ordered sequence of participation levels with increasing amounts of information (cf. P. B. Baltes et al., Chapter 1). For selectivity analyses, this has the important advantage that, at every step from one participation level to the next, the persons continuing participation can be compared with the dropouts on previously measured variables. In this way, we can identify variables on which the two groups differ. This eventually allows the calculation of estimates for constructs assessed later in the protocol which take account of observed sample selectivity at all previous levels.

2.2   Definition of Participation Levels

The parent sample of BASE is based on addresses drawn randomly from the West Berlin city registry. (In Germany, every citizen must register with the police.) The Intensive Protocol sample constitutes the highest participation level and offers the data for most analyses reported in this volume ($N = 516$). It is stratified by age and gender, meaning that the manifestations of these variables are distributed equally. Therefore, on the level of the Intensive Protocol, there are 43 men and 43 women in each of six age groups (70–74, 75–79, 80–84, 85–89, 90–94, and 95+ years).

As opposed to a random distribution, sample stratification has the advantage that age differences can be registered across the entire age range and in both sexes with equal reliability. A comparison with the expected numbers in a nonstratified random sample may illustrate this: If 516 persons aged 70 and above had been drawn in a nonstratified manner from the West Berlin population, there would have been approximately 94 70- to 74-

Figure 2.1. Sample attrition by age and gender. The columns in the 12 cells represent the five participation levels: (1) city registry ($N = 1{,}908$); (2) Short Initial Assessment ($n = 1{,}264$); (3) complete Intake Assessment ($n = 928$); (4) consent to take part in the Intensive Protocol ($n = 638$); (5) complete Intensive Protocol ($N = 516$). There are exactly 43 people in each of the 12 cells on the level of the Intensive Protocol.
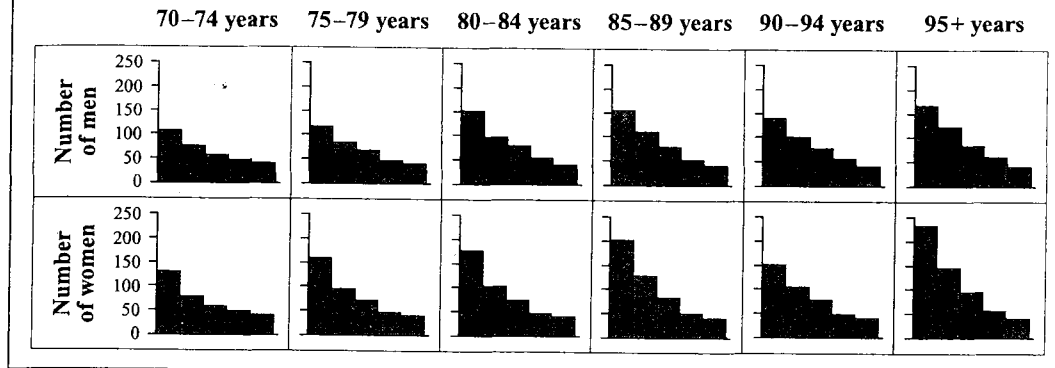
year-old women and 42 70- to 74-year-old men, but only 5 women and 1 man aged 95 and above in the sample.

The equal distribution by age and gender on the level of the Intensive Protocol was achieved by drawing a different number of addresses for each of the 12 design cells, thereby compensating for differences in sample attrition between the cells (see Figure 2.1). As outlined above, we distinguish five participation levels (see also Fig. 1.2 in P. B. Baltes et al., Chapter 1). From May 1990 until May 1993, a total of 2,297 persons received letters asking them to participate in BASE. *Participation level 1* comprises the subset of 1,908 individuals who could be contacted personally or indirectly (via information from relatives or third parties such as friends, neighbors, or nursing personnel). The fact that 389 persons could not be reached indicates that even an obligatory register does not necessarily constitute a perfect record of the population. Specifically, 50% of those 389 had already died. Others had moved to an unknown address (28%) or away from Berlin (4%). The remaining 18% could not be contacted at the given address despite many attempts. Therefore, we feel justified in designating those 1,908 individuals who were still alive and, in principle, recruitable, as constituents of the *verified parent sample*. By definition, participation at this level was 100%.

Participation level 2 ($n = 1{,}264$, or 66% of the verified sample of $N = 1{,}908$) contained all persons who completed all or most of the Short Initial Assessment. This consists of the first 16 questions of the multidisciplinary Intake Assessment, which were either posed by the interviewers during an initial contact or sent to participants as a questionnaire. In addition, interviewers carried out a standardized concurrent observation of the participants and their living conditions. For the selectivity analyses, we considered the 1,219 participants who answered all of the 16 questions as well as an additional group consisting of 45 individuals for whom information on most of those questions was available (sometimes from interviews with relatives, but always including

direct observational information from the interviewers). The criterion for reaching *participation level 3* (*n* = 928, or 49% of *N* = 1,908) was the completion of the Intake Assessment (an instrument developed in BASE assessing core constructs of all involved disciplines in 100 questions; cf. P. B. Baltes et al., Chapter 1).

The fourth and fifth participation levels are distinguished by the amount of information available from the Intensive Protocol. All participants who agreed to take part in the Intensive Protocol reached *participation level 4* (*n* = 638, or 33% of *N* = 1,908). Only those who actually *completed the entire protocol* consisting of the Intake Assessment and an additional 13 sessions were assigned to *participation level 5* (*N* = 516, 27% of *N* = 1,908). The multidisciplinary data set available for these persons is extensive (see P. B. Baltes et al., Chapter 1, Tables 1.4–7, for an overview of the measured constructs). *Participation level 5 constitutes the core of the Berlin Aging Study.* The extent of sample attrition at this level underscores the necessity of conducting selectivity analyses in order to safeguard generalizability.

The main aim of the BASE field coordination was to ensure stratification by age and gender at level 5 (cf. Nuthmann & Wahl, 1996, 1997). As can be seen in Figure 2.1, this was achieved, with 43 individuals in each of the 12 cells of the study design's highest participation level. Variations from equal distributions at lower levels show that different numbers of addresses had to be entered in each cell to reach stratification at the highest level.[2] For example, 235 women aged 95 and above had to be selected, whereas only 109 70- to 74-year-old men had to be selected on participation level 1 to achieve equal distribution on the level of completed Intensive Protocols.

The variation of sample attrition by age and gender raises the question whether the degree of sample *selectivity* is also associated with age and gender. For example, sample selectivity could be larger for the very old (aged 85 and above) than for the old (aged 70 to 84), or larger for women than men. This makes it methodologically necessary to consider the design variables age and gender systematically in the selectivity analyses.

2.3   Summary

The validity of statements made on the basis of the Intensive Protocol is initially restricted to those individuals who actually went through the entire measurement process. Therefore, the question arises as to whether the observations made about these persons are also true for the parent sample. Validity would be limited according to the extent that observed (but also nonobserved) characteristics that predict study participation or dropout are correlated with variables of interest, such as constructs from the Intensive Protocol. In itself, the fact that a complete data set from the Intensive Protocol is not available for 73% of the verified parent sample does not prove the existence of *selec-*

---

[2] From a theoretical perspective, redrawing addresses at a later time according to participation rates is problematic because observations can no longer be seen as being stochastically independent of each other. In the context of selectivity analyses, a dependency of observations would have to be examined more closely if persons recruited later in time differed in measured characteristics from those selected earlier. To test this, the variables presented in Table 2.1 were correlated with the date of the letter asking people to participate. No strong associations between that date and the observed characteristics were found within or across the 12 design cells. This indicates that the redrawing of addresses according to participation rates did not have any significant influence on the composition of the sample.

tivity, it only documents sample attrition. One can only speak of selectivity if sample loss was not random. The statistical methods presented below concentrate on this question.

3 Methods

First, we present the variables used in the selectivity analyses (Section 3.1). Second, we summarize expectations regarding the direction and pattern of selectivity effects on the basis of previous gerontological research (Section 3.2). Finally, we give an overview of statistical methods applied for the analysis of selectivity (Section 3.3).

3.1 Choice of Variables

The variables used in the following analyses are listed in Table 2.1. Our choice of variables was aimed at capturing selectivity as early (i.e., on low participation levels) and as comprehensively as possible.

All information available on level 1 was included. As can be seen in Table 2.1, the data from the city register were supplemented by an assessment of the living area based on an index developed for West Berlin by Meinlschmidt, Imme, and Kramer (1990), as well as an interviewers' description of the living quarters, which allowed the distinction between community dwellers (private households or sheltered housing for old people) and the institutionalized (senior citizens' homes, nursing homes, or hospitals for the chronically ill).

The one-year mortality variable is particularly important. It conveys information on mortality given to us by the city registry later on, and shows whether a person was still alive one year after he or she was sent the initial letter asking him or her to participate in BASE.

At the higher levels of participation, the selection of additional variables was guided by the consideration that information pertinent to important constructs be available as early as possible (e.g., health, functional capacity, intelligence, everyday competence, social networks, education, and well-being).

3.2 Predictions

The analyses reported in this chapter are mainly exploratory and descriptive. Nonetheless, some predictions on the direction of selectivity effects can be made on the basis of selectivity analyses carried out in previous gerontological research. In these (mainly longitudinal) studies, it has been shown that participation likelihood and duration are usually positively correlated with the many dimensions that can be subsumed under the terms of "fitness" or "competence." Thus, on average, persons taking part in studies for a longer time are *younger* (DeMaio, 1980; Hawkins, 1975; Lowe & McCormick, 1955; Mercer & Butler, 1967/68; Weaver et al., 1975) and *healthier* (Goudy, 1976; Hertzog, Schaie, & Gribbin, 1978; McArdle, Hamagami, Elias, & Robbins, 1991; Norris, 1985; Powers & Bultena, 1972; Schaie, Labouvie, & Barrett, 1973; Siegler & Botwinick, 1979) and come from a *higher social class* (Goudy, 1976; Powers & Bultena, 1972; Streib, 1966) than persons who do not agree to participate or who drop out during the course of the protocol. Moreover, participants often have *higher intelligence* (P. B. Baltes, Schaie, & Nardi, 1971; Cooney, Schaie, & Willis, 1988; Goudy, 1976; Norris,

Table 2.1. *List of variables used in selectivity analyses*

	Participation level								
	1 N = 1,908		2 n = 1,264		3 n = 928		4 n = 638		5 N = 516	
City registry and additional information									
Age at time of contact (in years)	86.1	(8.6)	86.3	(8.7)	86.0	(8.6)	85.5	(8.8)	84.9	(8.7)
Gender (proportion of women in %)	55.6		52.6		50.9		48.4		50.0	
One-year mortality (in %)	13.5		12.0		10.7		9.9		5.6	
Married (in %)	28.6		28.6		28.0		29.5		30.4	
Widowed (in %)	55.9		56.8		55.3		54.4		53.9	
Single (in %)	8.2		7.8		9.1		7.8		7.9	
Divorced (in %)	7.2		6.8		7.7		8.3		7.7	
Institutionalized (in %)	14.9		15.1		15.4		16.3		14.5	
Social Index of Living Area	24.2	(117.8)	24.7	(114.7)	31.4	(110.1)	36.1	(109.6)	40.7	(107.2)
Short Initial Assessment									
Observer-rated ADL[a]	—		3.9	(1.3)	4.0	(1.3)	4.1	(1.2)	4.2	(1.2)
Observer-rated signs of dementia	—		4.3	(5.8)	3.7	(5.6)	3.7	(5.5)	3.1	(4.9)
Observer-rated sensory impairment	—		3.2	(0.8)	3.2	(0.8)	3.2	(0.8)	3.3	(0.8)
Rough index of education	—		1.9	(0.7)	1.9	(0.7)	1.9	(0.7)	1.9	(0.7)
Life satisfaction (item)	—		3.5	(1.1)	3.6	(1.0)	3.6	(1.0)	3.7	(1.0)
Subjective physical health (item)	—		2.7	(1.1)	2.8	(1.1)	2.9	(1.1)	2.9	(1.1)
Subjective mental health (item)	—		3.3	(1.1)	3.5	(1.0)	3.5	(1.0)	3.5	(1.0)

Intake Assessment

ADL score	—	—	4.4	(1.3)	4.4	(1.3)	4.5	(1.2)
Subjective walking distance (in km)	—	—	4.2	(1.4)	4.3	(1.4)	4.4	(1.3)
Body Mass Index	—	—	1.2	(0.2)	1.2	(0.2)	1.2	(0.2)
SMMS[b]	—	—	0.45	(1.3)	0.51	(1.3)	0.59	(1.2)
Digit Letter test	—	—	68.1	(28.3)	70.3	(29.1)	73.3	(28.2)
Equivalent income (in DM)	—	—	1,979	(981)	2,020	(1,078)	2,042	(1,037)
Number of close companions	—	—	2.1	(3.0)	2.2	(3.3)	2.2	(3.3)
Depressivity (short scale)	—	—	10.4	(5.8)	10.1	(5.6)	10.0	(5.5)
Nonagitation	—	—	4.0	(0.9)	4.0	(0.9)	4.1	(0.8)
Satisfaction with aging	—	—	4.0	(0.8)	4.1	(0.8)	4.1	(0.8)
Life satisfaction (scale)	—	—	4.0	(0.8)	4.1	(0.8)	4.1	(0.8)

Intensive Protocol

Number of moderate to severe illnesses	—	—	—	—	8.1	(4.0)
ADL/IADL[a] (T-score)	—	—	—	—	50.0	(10.0)
Visual acuity (in Snellen decimals)	—	—	—	—	0.32	(0.18)
Hearing (threshold in decibels)	—	—	—	—	54.4	(16.0)
Dementia (clinical diagnosis; proportion in %)	—	—	—	—	21.1	
Intellectual functioning (T-score)	—	—	—	—	50.0	(10.0)
Education (in years)	—	—	—	—	10.8	(2.4)
Activities (number in previous year)	—	—	—	—	2.9	(2.4)
Social network size (number of persons)	—	—	—	—	10.9	(7.2)
Neuroticism (T-score)	—	—	—	—	50.0	(10.0)
Openness (T-score)	—	—	—	—	50.0	(10.0)
Depressivity (HAMD[c])	—	—	—	—	5.7	(6.1)
Depression (clinical diagnosis; proportion in %)	—	—	—	—	25.6	

Note. Data are given in means, with standard deviations in parentheses. All variables are introduced in more detail in the discipline-specific chapters of this book (cf. Steinhagen-Thiessen & Borchelt, Chapter 5; Helmchen et al., Chapter 6; Smith & Baltes, Chapter 7; Mayer et al., Chapter 8).

[a]ADL/IADL: Activities of Daily Living/Instrumental Activities of Daily Living.

[b]SMMS: Short Mini Mental State Examination; dementia screening.

[c]HAMD: Hamilton Depression Scale.

1985; Powers & Bultena, 1972; Schaie et al., 1973; Siegler & Botwinick, 1979), and a *lower mortality risk* (Cooney et al., 1988; Manton & Woodbury, 1983; Powell et al., 1990; Siegler & Botwinick, 1979). Finally, there are indications that the duration of participation is positively correlated with desirable personality characteristics, such as flexibility (Cooney et al., 1988).[3]

Selectivity in BASE most likely resembles this pattern. However, BASE has two special features which make it likely that the *observed* selectivity effects should be *larger* than in most other studies: (a) the stratification by age – leading to an unusually high proportion of very old people in the sample – and (b) the initial random sampling of addresses by the city registry. First, in the examined age range, age is negatively correlated with health and intelligence, and positively associated with the likelihood of death (Siegler & Botwinick, 1979). Therefore, health- and mortality-related selectivity should be stronger in an elderly sample stratified by age than in a random sample of older adults. Second, it can be argued that the random sampling by the city registry has probably resulted in a parent sample of little bias. There is no reason to suppose that the persons selected randomly were particularly intelligent, healthy, or educated. This may stand in contrast to studies whose parent samples were selected using other criteria, such as membership in a private health insurance (Schaie, 1983) or profession (Shock et al., 1984). Thus, random sampling of the parent sample reduces the likelihood of selectivity having taken place before it could even be observed. Compared with other studies, then, both the age stratification and the random drawing of addresses should increase, rather than decrease, the extent of *observable* selectivity effects in BASE.

3.3   Methodological Considerations

3.3.1   Overview of Statistical Methods Used

Sample selectivity is examined using three interrelated methods:[4] (a) logistic regressions to determine differences in means and frequency distributions (Aldrich & Nelson, 1984; Kühnel, Jagodzinski, & Terwey, 1989); (b) the comparison of variance-covariance matrices to measure differences in variability and correlations (Bentler, 1989); (c) Pearson-Lawley formulae to estimate statistical reference values for the parent sample on Intensive Protocol variables (Lawley, 1943; Neale, 1991). Looking at results obtained with all three methods should comprehensively document selectivity effects in BASE.

To clarify the relations between the three methods, it is helpful to formalize the data structure. With x_i, we refer to vectors of variables related to the four different participation levels (e.g., x_1 = city registry variables, x_2 = variables from the Short Initial Assessment, x_3 = variables from the Intake Assessment, x_5 = variables from the Intensive Protocol; x_4 is excluded because the willingness to participate in the Intensive Protocol is not associated with a unique vector of variables). For example, for persons at the highest level of participation ($N = 516$ of $N = 1,908$), we have observations (i.e., value assign-

---

[3] So far, not many empirical findings covering selectivity effects on correlation patterns have been reported in the gerontological literature (but see McArdle et al., 1991). Furthermore, there have been hardly any attempts to quantify selectivity effects on variables measured later using estimation techniques (but see McArdle & Hamagami, 1991; McArdle et al., 1991).

[4] The methods are recapitulated in Section 4. Readers not interested in procedural details may skip this section.

ments on variables) for vectors x_1, x_2, x_3, and x_5. For persons who did not even reach level 2, only observations on the vector x_1 are available.

Furthermore, let y_i be an indicator of observation loss (i.e., dropout) immediately after the ith participation level. Thus, $y_i = 1$ if a person reaches level i, but not level $i + 1$. If a person reaches level $i + 1$, $y_i = 0$. If a person is without observation on level i, y_i is not defined.

The three statistical methods are based on the assumption that x_i and y_i can be described as samples from a "superpopulation" characterized by a probability distribution (Cassel, Särndal, & Wretman, 1977). Assuming a superpopulation makes it possible to express aspects of the population as parameters and to use classical statistical procedures for their estimation. In particular, after making further assumptions, three different questions can be posed to clarify sample selectivity:

(1) *Sample attrition depending on means and frequency distributions of previously observed variables.* Can the likelihood of observation loss be described as a function of values on previously observed variables?

$$E\,(y_i \mid x_i) = Pr\,(y_i = 1 \mid x_i) \tag{1}$$

(2) *Differences in variances and covariances between samples of different participation levels.* Do continuing participants differ from dropouts in terms of variances and covariances on the previously observed variables?

$$cov\,(x_i \mid y_i = 0) \neq cov\,(x_i \mid y_i = 1) \tag{2}$$

(3) *Cumulative effects of sample attrition on Intensive Protocol constructs.* Are there differences in the expected values E (i.e., in means and frequency distributions) on Intensive Protocol variables between persons who reached this level and persons whose participation ended on a lower level?

$$E\,(x_5 \mid x_i, y_i = 0) \neq E\,(x_5 \mid x_i) \tag{3}$$

Thus, the questions refer to: (1) the relationship between the dropout indicator and previously measured variables; (2) the differences between persons continuing participation and dropouts on second moments (i.e., variances and covariances) of previously observed features; and (3) the relationship between features assessed on the highest participation level and features of previous levels, including the dropout indicator.

If differences between the groups of continuing participants and dropouts on the analyzed variables become apparent using the adequate statistical procedures, this would indicate that sample loss was associated with selectivity whose extent can be represented by statistical reference values. However, the reverse does not necessarily hold: If no differences are found, this *cannot* be taken as evidence that there is no sample selectivity. For instance, we cannot exclude the possibility that some features relevant to BASE that predict sample attrition were not assessed in the first place, and therefore could not be analyzed. This problem is particularly important for the analysis of sample loss at the step from participation level 1 to level 2 for which only the variables measured on the first level are available. In addition, the applied procedures do not capture all forms of sample selectivity possible in principle, but are limited to means, frequency distributions, variances, and covariances. However, as the analyses of BASE concentrate on these aspects of the data anyway, this restriction to first and second moments seems to be acceptable.

3.3.2   Selectivity Depending on Age Group and Gender

The variables age and gender are of special significance in BASE. On the one hand, they define the design of BASE as stratification variables. On the other hand, the description and explanation of differences by age and gender across all domains are among the study's central goals.

The validity of statements on age gradients and gender differences would be impaired if the magnitude of selectivity effects was a function of age or gender. In other words, if the selectivity of women differed from that of men, or if selectivity decreased or increased with age, the age and gender differences observed on the level of the Intensive Protocol would at least partially reflect this differential selectivity.

Estimates of prevalence rates represent a good example for this problem. It is possible that selectivity effects lead to a general underestimation of the prevalence of a disease. An examination of selectivity for each age group could show that the degree of underestimation is quite low at younger ages but becomes considerable at higher ages. In this case, the results of selectivity analysis would not only influence the interpretation of the observed prevalence rate, but also the interpretation of the age trend.

Therefore, possible consequences of sample attrition need to be examined for each age group, and for men and women separately. Ideally, one would analyze selectivity for each of the 12 cells of the study design (i.e., for men and women in each of the six age groups) in order to capture *interactions* between age and gender on selectivity, too. However, such analyses are associated with a marked reduction of statistical power due to relatively low numbers in each of the 12 cells. Mainly for this reason, we restrict our analyses to a comparison of selectivity between men and women across all age groups, and between the six age groups (70–74, 75–79, 80–84, 85–89, 90–94, 95+) across both sexes.

Statements concerning means, distributions, and other statistical reference values can either refer directly to the observed sample or, by weighting, can take into account that there are more women than men, and more younger than older persons among West Berliners aged 70 and above as compared with a sample stratified by age and gender. If not stated otherwise, all results presented here refer to *unweighted* data.

4   Results

In the following, we report the results of the selectivity analyses under three aspects: (1) means and distributions; (2) variability and covariance patterns; (3) cumulative selectivity effects on Intensive Protocol constructs. Because of its particular importance we begin with a separate consideration of mortality.

4.1   One-Year Mortality and Participation Depth

In a study of old age, the associations between participation likelihood and mortality are especially relevant. Figure 2.2 shows one-year mortality as a function of participation depth. Whereas one-year mortality is over 13% in the verified parent sample (participation level 1), it is under 6% in the Intensive Protocol sample (level 5).

Direct interpretation of these numbers is difficult because the mortality rate of persons who reached higher participation levels has to be lower than for those on lower lev-

els anyway, due to the longer period of time necessary for the collection of more detailed data. Thus, it would be possible to find differences in one-year mortality even if persons on various participation levels did not differ in their mortality risk. Therefore, it is necessary to compare empirically observed values with estimates that consider different examination durations under the assumption that the mortality risk is independent of participation level.

As illustrated by the comparison of observed and estimated values (see Fig. 2.2), *one-year mortality remains lower for persons reaching participation level 5 even if longer participation durations are taken into account*. The difference between the one-year mortality observed in the Intensive Protocol sample and the mortality expected due to the longer participation period is 3%. Accordingly, the mortality rates on lower participation levels are slightly higher than expected from participation durations.[5]

In summary, the Intensive Protocol sample's one-year mortality is clearly reduced in comparison to the parent sample. Considering the high mortality rate among older adults, this result is to be expected due to the duration of the Intensive Protocol (median = 133 days). Statistical analysis shows that most of the observed reduction of one-year mortality in the Intensive Protocol sample ($N = 516$) versus the parent sample ($N = 1,908$) is indeed due to the time taken for data collection. However, participants in the Intensive Protocol still had a lower mortality risk than persons in the parent sample. This finding is consistent with other studies' results (Cooney et al., 1988; Manton & Woodbury, 1983; Powell et al., 1990; Siegler & Botwinick, 1979). A more detailed analysis and discussion of this finding would require the identification of variables related to both participation depth and mortality (cf. P. B. Baltes et al., 1971; Kruse, Lindenberger, & Baltes, 1993; Maier & Smith, in press; Manton & Woodbury, 1983) and are beyond the rationale of the present investigation.
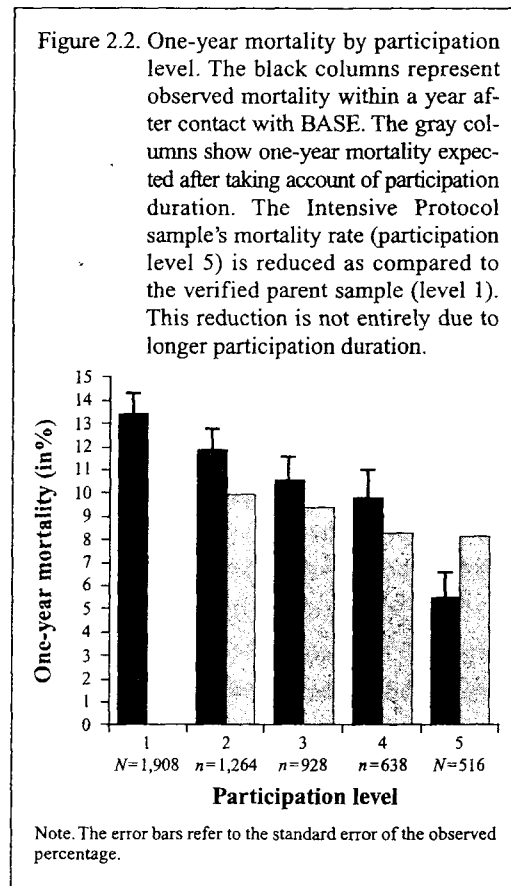
4.2   Means and Frequency Distributions

Logistic regressions were used to analyze the relationship between sample attrition and previously measured characteristics. The regression coefficient exponentials can be interpreted as "odds ratios." In the following, we briefly introduce this term to facilitate the understanding of our results.

For two-point (dichotomous) independent variables, an odds ratio can be interpreted directly as the probability quotient for manifestations of these variables. Thus, an odds ratio of 2.0 on the gender variable with 0 for men and 1 for women indicates that the dropout likelihood is twice as large for women as for men. For continuous, independent

---

[5] The estimates of one-year mortality taking into account differential participation duration shown in Figure 2.2 were calculated as follows. For every person who at least reached participation level 2, the duration of participation in days (s) was subtracted from 365 to determine the period of time during which death could have occurred. Given the one-year mortality rate (m) for the parent sample ($m = .1347$) and under the simplifying assumption of a constant mortality risk, the survival probability for the remaining days of the year (Pr) can be expressed in the following equation:

$$Pr = (1 - m)^{(365 - s)/365}$$

By subtracting this survival probability from one, one obtains the required mortality likelihood corrected for duration of participation. The expected values shown in Figure 2.2 are based on the mean of these probabilities.

Figure 2.2. One-year mortality by participation level. The black columns represent observed mortality within a year after contact with BASE. The gray columns show one-year mortality expected after taking account of participation duration. The Intensive Protocol sample's mortality rate (participation level 5) is reduced as compared to the verified parent sample (level 1). This reduction is not entirely due to longer participation duration.

Note. The error bars refer to the standard error of the observed percentage.

variables, the probability quotients expressed by odds ratios are scaled in the units of the independent variables. For example, if age is measured in years and the odds ratio is 1.1, the likelihood of dropout before the next participation level is 1.1 times higher for 71-year-olds than for 70-year-olds, and $1.1^{10} = 2.59$ times higher for 80-year-olds than for 70-year-olds.[6]

4.2.1   Contrast 1: Participation Level 1 Only (n = 644) versus Higher Participation Levels (n = 1,264)

The variables available on level 1 (age, gender, marital status, Social Index of Living Area, place of residence, and one-year mortality) were entered into the model as covariates. The variables age and Social Index of Living Area are continuous, whereas gender, place of residence, and one-year mortality are dichotomous variables with the manifestations 0 and 1. Marital status was divided into four dummy variables, with the group of

[6] Note that results of logistic regression coincide with those of discriminant analysis if the independent variables are more or less normally distributed (cf. Haggstrom, 1983).

the married as the reference category. We calculated the coefficients, their standard deviations (SD), p-scores, odds ratios, and pseudo-R^2.

With respect to this first contrast, the explanatory power of the total model was remarkably low, with a pseudo-R^2 of .014. Selectivity effects on the variables available on this participation level thus only explain a very small part of the attrition process. Significant differences between persons continuing to participate and dropouts were only found on the variables gender and one-year mortality. The likelihood of dropping out of the study was 1.68 times larger for persons who died within a year than for those who survived. For women, the dropout probability was 1.65 times larger than for men. The latter means that, initially, a larger number of addresses needed to be drawn for women to end up with equal numbers of men and women in the Intensive Protocol sample. There was no systematic association between participation likelihood and marital status, or place of residence on this participation level. Surprisingly, the same was true for age.

4.2.2 Contrast 2: Up to Participation Level 2 Only (n = 336) versus Higher Participation Levels (n = 928)

In addition to the variables listed in Section 4.2.1, the following variables could be considered on level 2: observer-rated ADL (Activities of Daily Living), dementia, and sensory impairment; education; and the subjective variables life satisfaction, physical health, and mental health (cf. Table 2.1).

As is to be expected, the inclusion of further variables resulted in greater explanatory power of the model than that on the first level (pseudo-R^2 = .19). This was mainly due to the Social Index of Living Area, the observational dementia rating, and subjective mental health. The higher the Social Index of Living Area (i.e., the "better" the residential area), the lower the dropout likelihood was (odds ratio = .73, $p < .01$). The probability of dropout also decreased with better subjective mental health (odds ratio = .70, $p < .01$). As for the dementia rating, where higher scores indicate a greater likelihood of dementia, dropout became more likely with higher scores (odds ratio = 1.09, $p < .01$). Persons with dementia symptoms were therefore more likely not to continue participation in the study. In addition, a larger probability of dropout for women was again found on this level (but with a greater standard deviation; odds ratio = 1.85, $p < .05$). A higher education level reduced the probability of dropout (odds ratio = .64, $p < .05$). For the remaining variables, no additional significant selectivity effects were found.

4.2.3 Contrast 3: Up to Participation Level 3 Only (n = 290) versus Higher Participation Levels (n = 638)

Core variables from the Intake Assessment were added in this model (see Table 2.1). In comparison to the model on contrast 2, this model's explanatory power is considerably lower (pseudo-R^2 = .05). Accordingly, the likelihood of ending participation after the Intake Assessment was not strongly related to the examined variables. On this level, place of residence and life satisfaction had an influence. Community-dwelling persons were nearly twice as likely to end participation on this level than the institutionalized (odds ratio = 0.49, $p < .05$). As to be expected, the dropout likelihood decreased with higher scores on the life satisfaction scale (odds ratio = 0.70, $p < .05$). None of the other variables examined in this model made significant additional contributions to the probability of further participation.

4.2.4   Contrast 4: Dropout During the Intensive Protocol (n = 122) versus Highest Participation Level (N = 516).

In this model, no additional variables were examined. The pseudo-R^2 of .16 indicates that this model describes dropout much better than the previous one. Here, one-year mortality had the largest effect (odds ratio = 7.07, $p < .01$). As explained above, to a large extent, this is due to the duration of the Intensive Protocol. In addition, the Social Index of Living Area was found to be relevant (odds ratio = 0.80, $p < .05$). Persons from "better" residential areas were less likely to drop out on this final level. Effects of the Body Mass Index and the Digit Letter test were also significant. Those with higher intellectual abilities (as measured by the Digit Letter test) were less likely to drop out during the Intensive Protocol (odds ratio = 0.99, $p < .05$). The same holds for the Body Mass Index,[7] an indicator of physical health: Dropout probability decreased with higher scores (odds ratio = 0.17, $p < .05$). No systematic associations between dropout likelihood during the Intensive Protocol and the other examined variables were found.

4.2.5   Summary of the Analyses of Means and Frequency Distributions

Relations between participation likelihood and observed variables were rather weak. The most prominent selectivity effect concerned one-year mortality. Dropout in the course of the Intensive Protocol is determined by this feature to a considerable degree and can be linked in part, but not completely, to the duration of data collection. Significant effects of the Social Index of Living Area on participation likelihood were observed on the second and fourth participation level. Accordingly, persons from "better" areas are slightly over-represented in the final Intensive Protocol sample. Together with the positive association between education and participation likelihood, this indicates a slight overrepresentation of the higher social classes. However, this education effect is only significant on the second participation level and has a high standard deviation.

On the third level, neither education nor income had an effect on participation likelihood, thus contradicting the hypothesis of clear selectivity by social class. The effects of the dementia rating score and of subjective mental health on the second level, of life satisfaction on the third, and of the Digit Letter test on the fourth participation level indicate an overrepresentation of healthier and more satisfied people, and of those with higher intellectual performance.

However, it must be added that many of the analyzed variables were not significantly related to participation probability. In particular, of the health measures, only the Body Mass Index had an influence on participation likelihood, so that one cannot assume a general overrepresentation of healthy individuals. Instead, a higher participation likelihood for persons living in institutional settings was found on level 3, thus counteracting the positive selectivity for good health. Surprisingly, there was no age effect on selectiv-

---

[7] The relationship between Body Mass Index and indicators of risk is known to be curvilinear in the general population (Andres, 1985); that is, both very high and very low values on the Body Mass Index are associated with higher mortality risks. In the present sample of old and very old people, however, individuals with very high Body Mass Index scores (e.g., very obese persons) were virtually absent (Andres, 1985; Borchelt & Steinhagen-Thiessen, 1992). Therefore, the selection effect on the Body Mass Index in the positive direction helps to explain the observed reduction of mortality risk among continuing participants.

ity independent of the variables included in the models (but see Section 5 for an explanation of this finding). The gender effects on the first and second participation level correspond to the higher dropout likelihood of women described in the literature. However, because of the stratification of the Intensive Protocol sample by gender, this does not influence the sample's composition.

4.3   Variability and Correlational Patterns

A comparison of variances and covariances can give further information on sample selectivity. It is possible that logistic regressions indicate that persons with especially low or high scores on certain variables are less likely to continue to participate in the study. However, this selectivity effect based on means does not necessarily imply that the two groups differ in their variability.

Another possibility is that consequences of sample attrition are more pronounced in variances and covariances than in means. If, for example, particularly healthy and particularly ill persons were, for different reasons, less likely to reach the next participation level than persons of average health, dropouts would not necessarily differ, on average, from the original sample on health variables. A selectivity effect like this can hardly be proved using logistic regression (i.e., the logit-model would be misspecified). However, the inconsistency of the attrition process should lead to a decrease of variance on the health variables, with the consequence that the variances of dropouts and those continuing participation would need to be estimated by two different parameters, rather than one common parameter. Hence, the comparison of variance-covariance matrices allows the examination of sample selectivity in terms of variances and covariances.[8]

Comparing variances and covariances, we again tested whether the group of dropouts differs from the group of those continuing participation. Three contrasts were defined for this purpose. In the *first contrast*, we examined whether persons on whom we only have information from level 1 ($n = 644$) differ from persons on whom more information is available, at least from level 2 ($n = 1,264$). The *second contrast* captures differences between persons on whom we have information from level 2, but not from level 3 ($n = 336$), and those for whom at least a complete Intake Assessment is available (level 3 and above; $n = 928$). Finally, the *third contrast* compares persons with a complete Intake Assessment, but no more ($n = 412$), with persons who completed the entire protocol ($N = 516$). Because of the low number of cases ($n = 122$), no separate contrast was defined for the participants who began the Intensive Protocol but did not complete it.

For the sake of clarity, we dispensed with a complete list of statistical reference values. Instead, group differences on single variables were only reported when their significance reached an α-level of $p = .01$. Dichotomous variables such as marital status dummies were not considered, because, in these cases, variance differences are statistically dependent on previously tested differences in means. Also, some variables were rescaled to yield well-conditioned variance-covariance matrices. For instance, the Social Index of

---

[8] The comparison of variance-covariance matrices was carried out using the statistics program EQS which was developed for structural equation models (Bentler, 1989). The main advantage of this program in comparison to standard procedures (e.g., the Box-M-test in SPSSX) is the option of carrying out tests of significance for single matrix elements or groups of elements selected by the user. A printout of the applied programs can be requested from the first author.

Living Area was divided by 100. The Comparative Fit Index (CFI; cf. Bentler, 1989) was used as an overall measure of fit of the models in which the variances and covariances of the compared groups are constrained to be equal. Generally, a CFI above 0.95 is taken as indicating good model fit.

4.3.1   Contrast 1: Participation Level 1 Only (n = 644) versus Higher Participation Levels (n = 1,264).

Here, only-differences of variances and covariances in age and the Social Index of Living Area could be tested (the only nondichotomous variables on this level). For this purpose, a model was specified in which variances and covariances for both groups were each estimated with the *same* parameter. Then we examined whether the estimation with two *different* parameters significantly improved the model's fit. This was not the case ($\chi^2(3) = 5.25$, $p = .15$). The fit of the model with across-group equality constraints was good (CFI = 1.000).

4.3.2   Contrast 2: Participation Level 2 Only (n = 336) versus Higher Participation Levels (n = 928)

Variance and covariance differences were tested for age, gender, education (the rough index in the Short Initial Assessment), life satisfaction (item), and subjective physical and mental health. No significant differences between the two groups were found. Again, the fit of the constrained model was good (CFI = 0.979).

4.3.3   Contrast 3: Participation Level 3 Only (n = 412) versus Highest Participation Level (N = 516)

In addition to the variables mentioned earlier, subjective walking distance, equivalent income, the number of close companions, ADL score, Body Mass Index, nonagitation, satisfaction with aging, life satisfaction (scale), Short Mini Mental State Examination (SMMS), Digit Letter test, and the depressivity score were considered (see Table 2.1). The overall test across all 17 variances was statistically significant ($\chi^2(17) = 62.13$, $p < .01$). Inspection of individual variances showed that this was due to two parameters. First, the variance in the Social Index of Living Area was lower in the Intensive Protocol sample than in the group of persons who only reached participation level 3 ($SD = 1.07$ vs. $SD = 1.15$, $\chi^2(1) = 26.24$, $p < .001$). Second, the variance in income was *higher* in the Intensive Protocol sample than in the comparison group ($SD = 1,037$ vs. $SD = 898$, $\chi^2(1) = 10.23$, $p < .001$).

The overall test for covariances was also significant ($\chi^2(136) = 204.45$, $p < .01$). Of the 136 individual comparisons 3 proved to be significant: the relationships between subjective physical health and nonagitation ($\chi^2(1) = 11.38$, $p < .01$), between subjective mental health and life satisfaction (scale; $\chi^2(1) = 7.36$, $p < .01$), and between subjective walking distance and depressivity score ($\chi^2(1) = 9.73$, $p < .01$). In all three cases, the covariance relationship was stronger in the Intensive Protocol sample than in the group of persons who only reached level 3.

Thus, in this contrast, there were indications of a significant overall difference in variance-covariance relations between dropouts and continuers. At the level of individual comparisons, however, only 5 of 153 resulted in significant differences between the two groups. Therefore, the impression of a relatively high degree of similarity in vari-

ances and covariances of both groups prevails. This is again underlined by the good fit of the model with across-group equality constraints (CFI = 0.983).

4.4   Cumulative Selectivity Effects on Intensive Protocol Constructs

The formulae on selectivity developed by Pearson (1903) and Lawley (1943) allow the estimation, using linear regressions, of statistical reference values (means, variances, and covariances) that take account of observed selectivity (Meredith, 1964, 1993; Muthén, Kaplan, & Hollis, 1987; Smith, Holt, & Smith, 1989). Independent (or selection) variables are distinguished from dependent variables (i.e., variables on which only persons continuing participation have scores). Means, variances, and covariances of the parent sample on the dependent variables are estimated on the basis of the linear relationships between selection variables and dependent variables and of the differences in selection variable means between the parent and the selected sample.

The Pearson-Lawley method uses the results presented above on the relationships between expected values (means, distributions), variances, and covariances, with the understanding that the regressions of the dependent variables on selection variables are linear and that the conditional variances are constant (homoscedasticity). Under these assumptions, the method allows a direct estimation of selectivity effects on the Intensive Protocol constructs central to BASE. Within the framework of the linear model, this projection makes optimal use of all of the available information (cf. Meredith, 1964).
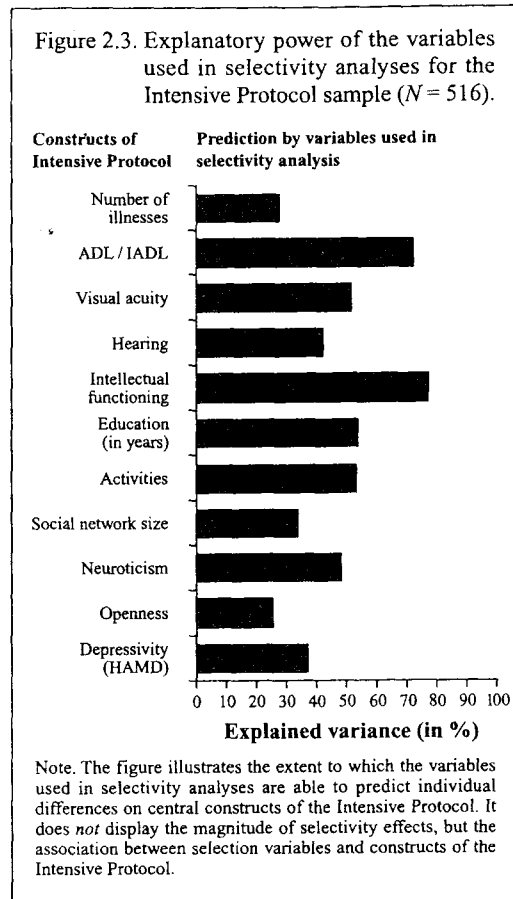
Aitkin (1934) and Lawley (1943) showed that these formulae can be applied *repeatedly*. For our selectivity analyses this means that the variables available on participation level 1 can be used to estimate the parent sample's manifestations on level 2 variables. Then, the variables of level 1 *and* 2 (i.e., the *observed* values on level 1, and the *estimated* values on level 2) serve as selection variables, and the variables available on level 3 become the dependent variables. Finally, the variables on the first three levels are the selection variables by which the manifestations of the parent sample on the Intensive Protocol constructs can be estimated.[9]

When the results obtained with this method are interpreted, the following rule, related to the basic paradox of selectivity analysis mentioned in Section 2.1, needs to be kept in mind: The more closely features on lower participation levels (i.e., the selection variables) are associated with variables on the following level (i.e., the independent variables), the more meaningful the calculated estimates become. Dependent variables for which there are no "precursors" on previous participation levels cannot be corrected and inevitably maintain the means that were observed in the selected subsample.

Therefore, we first examined how well the selection variables to be considered explained the variance in the chosen Intensive Protocol constructs. This analysis was based on the Intensive Protocol sample ($N = 516$) and, accordingly, does not involve any kind of estimation procedure.

As shown in Figure 2.3, the proportion of explained variance in the Intensive Protocol constructs was generally quite high. The lowest value (28% explained variance) was found for the number of moderate to severe illnesses, and the highest was observed for

---

[9]The statistical program Mx (Neale, 1991) was used for the calculation of the estimates using the Pearson-Lawley formulae. The printout of the program can be requested from the first author.

Figure 2.3. Explanatory power of the variables
used in selectivity analyses for the
Intensive Protocol sample ($N = 516$).

| Constructs of Intensive Protocol | Prediction by variables used in selectivity analysis |

Explained variance (in %)

Note. The figure illustrates the extent to which the variables
used in selectivity analyses are able to predict individual
differences on central constructs of the Intensive Protocol. It
does *not* display the magnitude of selectivity effects, but the
association between selection variables and constructs of the
Intensive Protocol.

general intelligence (78%). An inspection of correlations and semipartial regression co-
efficients (which are not presented here in detail) showed that the variables deemed to
function as "precursors" were indeed the ones that mainly contributed toward explaining
variance in the constructs assigned to them. For instance, the observational ratings of de-
mentia, sensory functioning, and ADL on participation level 2 played an important role
in explaining variance in the corresponding Intensive Protocol constructs. Thus, insofar
as there is selectivity, the predictive power of the selection variables is definitely large
enough to detect it.

4.4.1   Effects of Selectivity on the Entire Sample

Figure 2.4 shows the discrepancies between the means estimated for the parent sample
and the means observed in the Intensive Protocol. The normed standard deviation E_{SD} is
the measure used for this comparison (estimated mean of the parent sample minus ob-
served mean of the selected sample, divided by the parent sample's estimated standard
deviation; cf. Hedges & Olkin, 1985). It indicates the distance between the mean of the

Figure 2.4. Selectivity effects in BASE. Discrepancies between the observed means of central Intensive Protocol constructs and those estimated using the Pearson-Lawley formulae to correct for selectivity. The sizes of the discrepancies are expressed in normed standard deviations and reflect the degree of observed cumulative (i.e., across-level) selectivity effects.
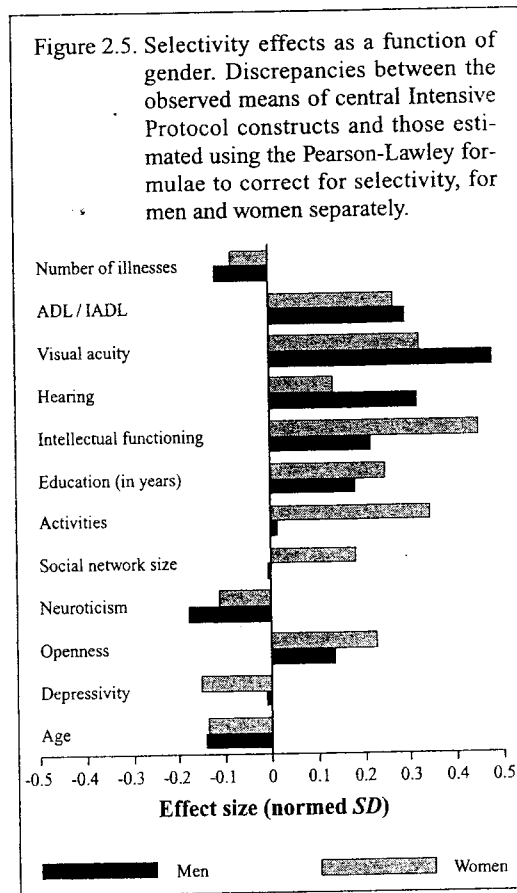

Effect size (normed *SD*)
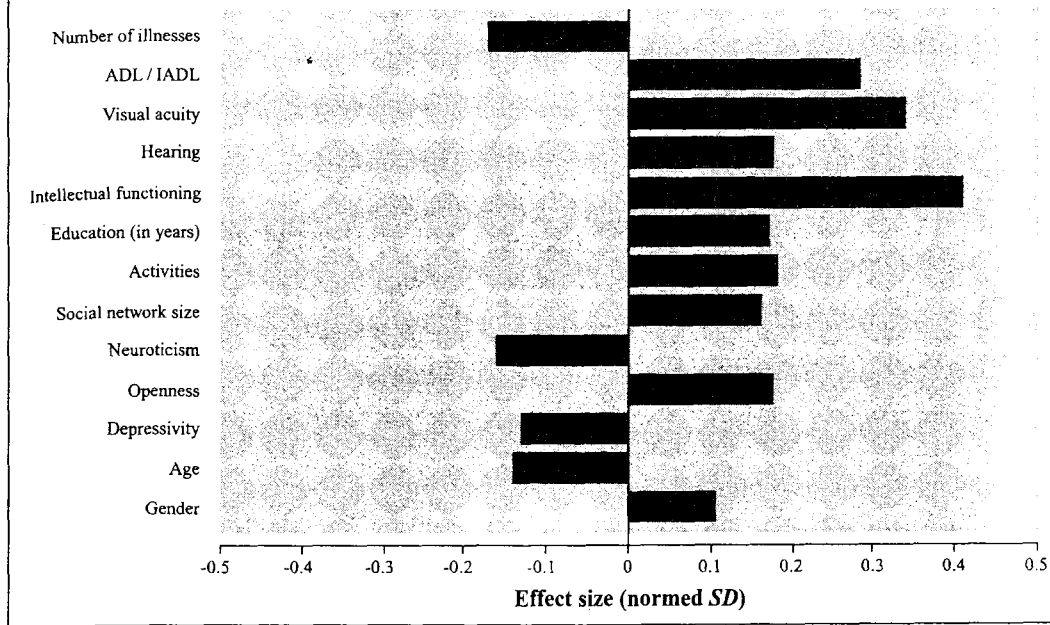
selected sample and the mean of the parent sample. Thus, $E_{SD} = 0.5$ would signify that the selected sample's mean is half a standard deviation above the parent sample's mean. The variables age and gender (0 = men, 1 = women) are also listed because some of the effects could be linked to the design variables. For these two variables, effect sizes could be directly calculated because these features are known for every participation level.

The pattern of effects confirms the picture reported in the literature: Persons with better intellectual functioning, better vision and hearing, more self-reliance in everyday activities (ADL/IADL), less illnesses, and a higher education level were more likely to proceed through the entire Intensive Protocol. The findings on less frequently examined domains such as social network size, number of activities, neuroticism, openness, and depressivity also fit into this general pattern.

If one assumes a normal distribution, an effect size of .14 (the smallest effect size for depressivity) indicates that 56% of the observations in the Intensive Protocol – instead of chance 50% – lie above the mean of the parent sample. The effect size of .42 (the largest effect size for intellectual functioning) signifies a proportion of 66%. The effect sizes therefore represent relatively small deviations from the values estimated for the parent sample.

The estimated prevalence rates for the clinical diagnoses of dementia and depression are not shown in Figure 2.4. A prevalence estimate of 30% was made for dementia –

Figure 2.5. Selectivity effects as a function of gender. Discrepancies between the observed means of central Intensive Protocol constructs and those estimated using the Pearson-Lawley formulae to correct for selectivity, for men and women separately.
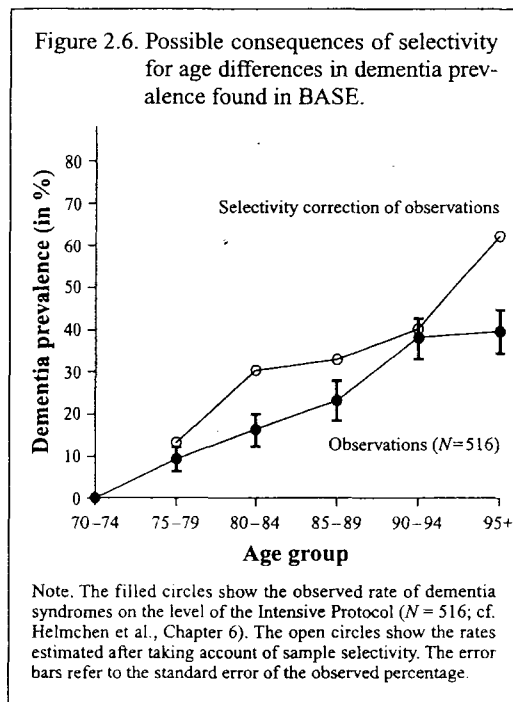
considerably higher than the observed 21%. For depression, there was little difference between the estimated prevalence and the observed prevalence (27% estimated for the parent sample vs. 26% observed).

4.4.2 Effects of Selectivity on Gender Differences

Figure 2.5 refers to possible gender differences in selectivity. The clearest differences were found on the constructs social activities (men: .02, women: .35), intellectual functioning (men: .22, women: .45), and social network size (men: -.01, women: .19). With respect to these three variables, women in the Intensive Protocol were more positively selected than men.

4.4.3 Effects of Selectivity on Age Differences

An inspection of the effect sizes calculated separately for the six age groups showed little selectivity differences as a function of age. The clinical diagnosis of dementia was a notable exception (cf. Helmchen et al., Chapter 6). Figure 2.6 shows the observed and estimated dementia prevalence rates. The observed values suggest that prevalence does

Figure 2.6. Possible consequences of selectivity for age differences in dementia prevalence found in BASE.

Note. The filled circles show the observed rate of dementia syndromes on the level of the Intensive Protocol ($N = 516$; cf. Helmchen et al., Chapter 6). The open circles show the rates estimated after taking account of sample selectivity. The error bars refer to the standard error of the observed percentage.

not increase any further in the two highest age groups. In contrast, the *estimated* values support the hypothesis that dementia prevalence continues to rise, even beyond the age of 90. In this case, taking account of selectivity leads to a quite different conclusion. The examination of the regression coefficients revealed that the observer rating of dementia on participation level 2, as well as the SMMS and the Digit Letter test on level 3, made the largest contributions to the prediction of these estimates.

5   Discussion

Before we summarize and comment on our results, we expressly want to point out the limits of selectivity analyses. Selectivity analyses show the extent of observed selectivity, but not the extent of selectivity that is possible in principle. They try to relate *available* data to each other and to make optimal use of available information, but obviously cannot deal with issues for which additional information is necessary. This necessary deficit is especially relevant for the step from the first ($N = 1,908$) to the second participation level ($n = 1,264$) which is documented by only a few variables. The question of whether the availability of more variables on level 1 would have resulted in the observation of more selectivity at this step (and thus overall) cannot be answered conclusively. With this limitation in mind, the results of the selectivity analyses can be summarized as follows:

First, the selectivity analyses indicate that neither variances nor covariance relations in the Intensive Protocol sample (N = 516) are very different from those of the samples

on lower participation levels. With one exception – a slight reduction of the variance in the Social Index of Living Area – there were no signs of variance reduction. Thus, the manifold forms of aging and social situations, as well as the structure of associations among variables, appear to be maintained in the Intensive Protocol sample. Concerning selectivity, it seems permissible to generalize statements about associations among variables observed in the Intensive Protocol sample to the parent sample. This result is particularly important for systemic (e.g., multidisciplinary) analyses examining relationships between different domains of functioning.

Second, in comparison with variances and covariances, selectivity effects in means and frequency distributions were more pronounced. Judging by the number of variables analyzed, the number of significant effects was rather low. Across all participation levels, there were indications of positive selection by social class, subjective mental health, and life satisfaction, as well as intelligence and dementia. There were no signs of selectivity with respect to health measures, with the exception of Body Mass Index on the last level. The higher participation likelihood observed for individuals in institutional settings suggests that different selectivity effects may have worked in opposite directions. Possibly, the higher participation of this group partially evens out the positive selection effect for health and intelligence.

Surprisingly, no strong links between age and participation likelihood were found on any participation level. This result seems to contradict the varying pattern of sample loss by age group shown in Figure 2.1. However, one needs to remember that only linear age effects can be captured by the reported logistic regressions (i.e., either a linear decrease or increase of sample attrition with age). More detailed analyses of the data in Figure 2.1 indicate *nonlinear* relationships (Gilberg & Pötter, 1994). For example, sample loss in the age group of the 85- to 89-year-olds was larger than in the two adjacent groups.

More clear-cut selection effects were observed by gender. However, because of the study's stratified design, women's lower participation likelihood had no influence on the proportion of men and women in the Intensive Protocol sample.

Third, the participants in the BASE Intensive Protocol sample have a lower mortality rate in comparison to the parent sample (see Fig. 2.2). This result is largely, but not completely, due to the duration of the Intensive Protocol. Further interpretation of this mortality-associated sample selectivity requires the identification of measures that predict both participation depth and death rates. In the examined age range, of course, the population composition already alters so much by age that the assumption of a uniform parent sample is not really tenable. In Germany, 81% of women and 66% of men reach age 70, but only about 18% of women and 7% of men reach age 90 (Statistisches Bundesamt [Federal Statistical Office], 1993). It is well known that individual differences in longevity are correlated with features such as health, intelligence, and well-being (Siegler & Botwinick, 1979; for first analyses of the BASE data set regarding this issue, cf. Maier & Smith, in press). Thus, mortality is a selective process. Inevitably, age-related analyses of the BASE data set confound the aging process of the survivors with the selectivity of mortality (cf. Keiding, 1991; Mulder, 1993).

Fourth, the cumulative analysis of selectivity and its projection onto Intensive Protocol constructs using the Pearson-Lawley formulae shows that the Intensive Protocol sample is positively selected in all considered domains. However, the extent of these selectivity effects never exceeded half a standard deviation, and varied only slightly by age. For the interpretation of results presented in other chapters, this means that the val-

ues found empirically on the level of the Intensive Protocol tend to show old age and aging in a slightly too positive light.

An important exception from the general rule of rather low and age-stable selectivity effects is dementia prevalence. In this case, there was quite a substantial discrepancy between the rate observed in the Intensive Protocol sample (21%) and the rate estimated using the Pearson-Lawley formulae (30%). Clearly, sample attrition led to an underestimation of dementia prevalence in the parent sample, especially for the group aged 95 and above (40% vs. 62%).

Unfortunately, it is nearly impossible to compare the extent of selectivity in BASE directly with results from other studies. First, a comparison is impeded by many differences in methodology and content. Second, mentions of selectivity are often so fragmentary that one can only speculate on effect sizes. In general, the impression prevails that the extent of selectivity effects in BASE is more or less within the usual range. In view of the considerable degree of sample attrition (see Fig. 2.1) and of BASE participants' particularly old age, this is certainly a satisfactory result.

References

Aitkin, A. C. (1934). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society, 4*, 106–110.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit and probit models*. Beverly Hills, CA: Sage.

Andres, R. (1985). Mortality and obesity: The rationale for age specific height-weight tables. In R. Andres, E. L. Bierman, & W. R. Hazzard (Eds.), *Principles of geriatric medicine* (pp. 311–318). New York: McGraw-Hill.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1988). *Life-span developmental psychology: Introduction to research methods*. Hillsdale, NJ: Erlbaum.

Baltes, P. B., Schaie, K. W., & Nardi, A. H. (1971). Age and experimental mortality in a seven-year longitudinal study of cognitive behavior. *Developmental Psychology, 5*, 18–26.

Bentler, P. M. (1989). *EQS: Structural equations manual (Version 3.0)*. Los Angeles: BMDP Statistical Software.

Borchelt, M., & Steinhagen-Thiessen, E. (1992). Physical performance and sensory functions as determinants of independence in activities of daily living in the old and the very old. *Annals of the New York Academy of Sciences, 673*, 350–361.

Cassel, C. M., Särndal, C. E., & Wretman, J. H. (1977). *Foundations of inference in survey sampling*. New York: Wiley.

Cooney, T. M., Schaie, K. W., & Willis, S. K. (1988). The relationships between prior functioning on cognitive and personality dimensions and subject attrition in longitudinal research. *Journal of Gerontology: Psychological Sciences, 43*, P12–P17.

Dalenius, T. (1988). A first course in survey sampling. In P. Krishnaiah & C. R. Rao (Eds.), *Handbook of statistics: Vol. 6. Sampling* (pp. 15–46). Oxford: North-Holland.

DeMaio, T. J. (1980). Refusals: Who, where and why. *Public Opinion Quarterly, 44*, 223–233.

Esser, H., Grohmann, H., Müller, W., & Schäffer, K.-A. (1989). *Mikrozensus im Wandel: Untersuchungen und Empfehlungen zur inhaltlichen und methodischen Gestaltung* (Schriftenreihe Forum der Bundesstatistik, Vol. 11). Stuttgart: Metzler-Poeschel.

Eye, A. von. (1989). Zero-missing non-existing data: Missing data problems in longitudinal research and categorical data solutions. In M. Brambring, F. Lösel, & H. Skowronek

(Eds.), *Children at risk: Assessment, longitudinal research, and intervention* (pp. 336–355). Berlin: de Gruyter.

Gilberg, R., & Pötter, U. (1994). *Poststratifizierungsgewichte für die BASE-Studie.* Unpublished manuscript, Max Planck Institute for Human Development, Berlin.

Goudy, W. J. (1976). Nonresponse effects on relationships between variables. *Public Opinion Quarterly, 40,* 360–369.

Haggstrom, G. W. (1983). Logistic regression and discriminant analysis by ordinary least squares. *Journal of Business and Economic Statistics, 1,* 229–238.

Hawkins, D. F. (1975). Estimation of nonresponse bias. *Sociological Methods and Research, 3,* 461–488.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hertzog, C., Schaie, K. W., & Gribbin, K. (1978). Cardiovascular disease and changes in intellectual functioning from middle to old age. *Journal of Gerontology, 33,* 872–883.

Herzog, A. R., & Rodgers, W. L. (1988). Age and response rates to interview sample surveys. *Journal of Gerontology: Social Sciences, 43,* S200–S205.

Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective. *Journal of the Royal Statistical Society, Series A, 154,* 371–412.

Kessler, R. C., Little, R. J., & Groves, R. M. (1995). Advances in strategies for minimizing and adjusting for survey nonresponse. *Epidemiologic Reviews, 17,* 192–204.

Kruse, A., Lindenberger, U., & Baltes, P. B. (1993). Longitudinal research on human aging: The power of combining real-time, microgenetic, and simulation approaches. In D. Magnusson & P. Casaer (Eds.), *Longitudinal research on individual development: Present status and future perspectives* (pp. 153–193). Cambridge: Cambridge University Press.

Kruskal, W., & Mosteller, F. (1979a). Representative sampling: I. Non-scientific literature. *International Statistical Review, 47,* 13–24.

Kruskal, W., & Mosteller, F. (1979b). Representative sampling: II. Scientific literature, excluding statistics. *International Statistical Review, 47,* 111–127.

Kruskal, W., & Mosteller, F. (1979c). Representative sampling: III. The current statistical literature. *International Statistical Review, 47,* 245–265.

Kühnel, S. M., Jagodzinski, W., & Terwey, M. (1989). Teilnehmen oder boykottieren: Ein Anwendungsbeispiel der binären logistischen Regression mit SPSSX. *ZA-Information, 25,* 44–75.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62,* 28–30.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal (the American Statistical Association, 90,* 1112–1121.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Lowe, F. E., & McCormick, T. C. (1955). Some survey sampling biases. *Public Opinion Quarterly, 19,* 303–315.

Maier, H., & Smith, J. (in press). Psychological predictors of mortality in old age. *Journal of Gerontology: Psychological Sciences.*

Manton, K. G., & Woodbury, M. A. (1983). A mathematical model of the physiological dynamic of aging and correlated mortality selection: II. Application to Duke Longitudinal Study. *Journal of Gerontology, 38,* 406–413.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29,* 409–454.

McArdle, J. J., & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 276–298). Washington, DC: American Psychological Association.

McArdle, J. J., Hamagami, F., Elias, M. F., & Robbins, M. A. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental Aging Research, 17,* 29–51.

Meinlschmidt, G., Imme, U., & Kramer, R. (1990). *Sozialstrukturatlas Berlin (West): Eine statistisch-methodische Analyse mit Hilfe der Faktorenanalyse.* Berlin: Senatsverwaltung für Gesundheit und Soziales.

Mercer, J. R., & Butler, E. W. (1967/68). Disengagement of the aged population and response differentials in survey research. *Social Forces, 46,* 89–96.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29,* 177–185.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525–543.

Mulder, P. G. H. (1993). The simultaneous processes of ageing and mortality. *Statistica Nederlandica, 47,* 253–267.

Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not completely missing at random. *Psychometrika, 52,* 431–462.

Neale, M. C. (1991). *Mx: Statistical modelling.* Department of Human Genetics, Box 3MCV, Richmond, VA 23298.

Norris, F. H. (1985). Characteristics of nonrespondents over five waves of a panel study. *Journal of Gerontology, 40,* 627–636.

Nuthmann, R., & Wahl, H.-W. (1996). Methodische Aspekte der Erhebungen der Berliner Altersstudie. In K. U. Mayer & P. B. Baltes (Eds.), *Die Berliner Altersstudie* (pp. 55–83). Berlin: Akademie Verlag.

Nuthmann, R., & Wahl, H.-W. (1997). *Methodological aspects of the Berlin Aging Study* (Technical report). Berlin: Max Planck Institute for Human Development.

Oh, H. L., & Scheuren, F. J. (1983). Weighting adjustment for unit non-response. In W. G. Madow, I. Olkin, & D. B. Rubin (Eds.), *Incomplete data in sample surveys: Vol. 2. Theory and bibliographies* (pp. 143–184). New York: Academic Press.

Panel on Incomplete Data. (1983). Part I: Report. In W. G. Madow, I. Olkin, & D. B. Rubin (Eds.), *Incomplete data in sample surveys: Vol. 1. Report and case studies* (pp. 3–106). New York: Academic Press.

Pearson, K. (1903). Mathematical contributions to the theory of evolution: XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London (Series A), 200,* 1–66.

Powell, D. A., Furchtgott, E., Henderson, M., Prescott, L., Mitchell, A., Hartis, P., Valentine, J. D., & Milligan, W. L. (1990). Some determinants of attrition in prospective studies on aging. *Experimental Aging Research, 16,* 17–24.

Powers, E. A., & Bultena, G. L. (1972). Characteristics of deceased dropouts in longitudinal research. *Journal of Gerontology, 27,* 350–353.

Rendtel, U., & Pötter, U. (1992). *Über Sinn und Unsinn von Repräsentativitätsstudien* (Discussion paper No. 61). Berlin: Deutsches Institut für Wirtschaftsforschung.

Rudinger, G., & Wood, P. K. (1990). N's, times and number of variables in longitudinal research. In D. Magnusson & L. R. Bergman (Eds.), *Data quality in longitudinal research* (pp. 157–180). Cambridge: Cambridge University Press.

Schaie, K. W. (1983). The Seattle Longitudinal Study: A twenty-one year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford.

Schaie, K. W., Labouvie, G. V., & Barrett, T. J. (1973). Selective attrition effects in a fourteen-year study of adult intelligence. *Journal of Gerontology, 28,* 328–334.

Shock, N. W., Greulich, R. C., Costa, P. T., Jr., Andres, R., Lakatta, E. G., Arenberg, D., & Tobin, J. D. (1984). *Normal human aging: The Baltimore Longitudinal Study on Aging* (NIH Publication No. 84-2450). Washington, DC: Government Printing Office.

Siegler, I. C., & Botwinick, J. (1979). A long-term longitudinal study of intellectual ability of older adults: The matter of selective subject attrition. *Journal of Gerontology, 34,* 242–245.

Smith, C. J., Holt, D., & Smith, T. M. F. (1989). *Analysis of complex surveys.* New York: Wiley.

Statistisches Bundesamt. (1993). *Statistisches Jahrbuch 1993.* Stuttgart: Metzler-Poeschel.

Streib, G. F. (1966). Participants and dropouts in a longitudinal study. *Journal of Gerontology, 21,* 200–209.

Tennstedt, S. L., Dettling, U., & McKinlay, J. B. (1992). Refusal rates in a longitudinal study of older people: Implications for field methods. *Journal of Gerontology: Social Sciences, 47,* S313–S318.

Weaver, C. N., Holmes, S. L., & Glenn, N. D. (1975). Some characteristics of inaccessible respondents in a telephone survey. *Journal of Applied Psychology, 60,* 260–262.

The Berlin Aging Study
Aging from 70 to 100

Edited by

PAUL B. BALTES

KARL ULRICH MAYER

A Research Project of the Berlin-Brandenburg Academy of Sciences

Coordinating Editor: Julia Delius

CAMBRIDGE *2001*
UNIVERSITY PRESS

Berlin-Brandenburg Academy of Sciences

Study Group: Aging and Societal Development

Paul B. Baltes (chair)
Georg Elwert
Angela Friederici
Wolfgang Gerok
Hanfried Helmchen
Karl Ulrich Mayer (deputy chair)
Jürgen Mittelstraß
Elisabeth Steinhagen-Thiessen
Karl Mathias Neher (BASE coordination)
Ursula M. Staudinger (coordination of the study group)

Steering Committee of the Berlin Aging Study (BASE)

Paul B. Baltes (co-chair)
Markus Borchelt
Hanfried Helmchen
Michael Linden
Karl Ulrich Mayer (co-chair)
Jacqui Smith
Elisabeth Steinhagen-Thiessen
Michael Wagner
Karl Mathias Neher (coordination)