Attrition in the New Beneficiary Survey and Followup, and Its Correlates

by Kate Antonovics, Robert Haveman, Karen Holden, and Barbara Wolfe

In this article we explore the extent of and reasons for attrition in the New Beneficiary Survey (NBS) between the first interview in 1982 and the followup interview in 1991. We examine a variety of potential determinants of attrition, separating the probability of attrition due to death from a refusal to be interviewed. Because the NBS sample is drawn from and linked to Social Security administrative records, information on mortality as a cause of attrition is exact. Hence, we are able to examine differences in the patterns and predictors of attrition due to these two causes of attrition and differences between attrition among retired and disabled workers.

* Kate Antonovics is with the Department of Economics, University of California, San Diego. The authors listed below are all within the University of Wisconsin-Madison. Robert Haveman is with the Department of Economics, La Follette Institute of Public Affairs; Karen Holden is with the Department of Consumer Science, La Follette Institute of Public Affairs; and Barbara Wolfe is with the Department of Economics, Preventive Medicine. *(Manuscript received October 1999; submitted for external review November 1999; revise and resubmit recommended January 2000; revision received February 2000; paper accepted March 2000.)*

Panel surveys interview the same individuals more than once over a period of time. Attrition from the survey occurs when those individuals die, refuse to be interviewed again, or, for some other reason, cannot be contacted. If the original sample was representative of a specific population, then survey analysis may provide misleading conclusions about changes in population characteristics over time if these individuals leave the sample in a nonrandom way. Therefore, it is important to identify the characteristics of individuals who leave the survey for various reasons.

This article explores the extent of and reasons for attrition in the New Beneficiary Survey (NBS) between the first interview in 1982 and the followup interview in 1991. Presented is a comparison of the characteristics of survivors (the reinterviewed sample) with attriters (those in the sample not reinterviewed) from the retired-worker and disabled-worker samples. The article explores a variety of potential determinants of attrition to the probability of attrition. These determinants are examined alone and in a multivariate framework. The NBS sample population is drawn from and linked to Social Security Administrative records, which have exact matched data on mortality as a cause of attrition. These data do not depend on survey-reported reasons for attrition; hence, it allows the examination of the differences in the patterns and predictors of attrition due to death and due to other reasons, primarily, the refusal to be interviewed. Attrition due to death must be identified precisely because misidentification of death as refusal to be interviewed may lead researchers to infer more selective attrition than might be the case.

Different patterns of attrition are evident in the comparison of attrition levels and the determinants of attrition for the retired and disabled samples, both composed of persons with relatively high mortality risk. In particular, individuals' health, health insurance coverage, and level of education have different impacts on their likelihood of attrition. In general, it appears that refusal to be interviewed is more evenly spread across populations and characteristics than is death. The analysis shows that attrition due to death and attrition due to refusal are quite different processes, even though health conditions play a role in both processes. The results suggest that because attrition patterns (including death) may be quite different across population samples, sample-specific attrition patterns must be analyzed over the lifetime of any panel study. Long-term studies of panel attrition are necessary to provide researchers analyzing the data with information on potential biases due to nonrandom attrition.

I. Introduction

The ability to generalize from survey-based estimates of the determinants of economic behavior requires representative samples from a larger population whose behavior is at issue. When cross-section samples are not random or representative, use of well-established weighting techniques may adjust for sampling variation across population groups and differential response rates. However, a more difficult problem arises in dealing with panel data when over time there

is nonrandom attrition from the sample across population groups (Fitzgerald, Gottschalk, and Moffitt 1998). This problem is especially severe if the probability of attrition is related to that aspect of behavior whose determinants are being analyzed (Zabel 1998). For example, persons with disabilities in a representative first-year sample are more likely to attrite from a longitudinal database (because of death, institutionalization, or refusal to answer) than similarly aged nondisabled people; hence, researchers seeking estimates of time-related behavioral responses using data from waves subsequent to the first interview must attend to this selective attrition.[1]

Nonrandom attrition may also be a problem when the population of interest is composed of the observations from the original sample who have not died (survivors). While the sample of survivors will accurately represent the population of survivors if attrition from the sample is due only to death, sample attrition for reasons other than death may lead to a sample of reinterviewed observations that is not representative of the population of survivors.[2] Analytical challenges exist if, as with many data sets, deaths cannot be distinguished from other reasons for attrition. Indeed, even if one is interested only in survivors, attrition may create potential bias. For example, estimates of the effect of benefit changes on the well-being of survivors may be biased if the benefit change nonrandomly affects the mortality probability of recipients, and hence the composition of the sample of survivors. In this case, analysis based only on actual survivors would fail to accurately assess the full impact of a benefit change.

In this article, we explore the extent of and reasons for attrition in the New Beneficiary Survey (NBS) between the first interview in 1982 and the followup interview in 1991. We begin by comparing the characteristics of survivors (the reinterviewed sample) with attriters, those who were not reinterviewed. Next, we relate a variety of potential determinants of attrition to the probability of attrition, and we examine these determinants both individually and in a multivariate framework. Because the NBS sample (described below) is drawn from and linked to Social Security administrative records, information on mortality as a cause of attrition is exact, rather than depending, as in most other surveys, on the identification of a death by interviewer or other nonadministrative sources. Hence, we are able to examine differences in the patterns and predictors of attrition due to death and due to other reasons (primarily, refusal to be interviewed).

Our analysis compares differences in both attrition levels and the determinants of attrition between two population groups with relatively high mortality risk: retired persons and disabled persons. We assume that attrition is an absorbing state in that neither individuals nor their proxy respondents can return to the survey once they exit. Although both groups face relatively high risks of death (the retired because of their advanced age and the disabled because of adverse health conditions), there may be quite different relationships between individual characteristics (including age and health conditions) and attrition due to death and other causes between the two groups. Hence, extrapolating information about the correlates of death (or attrition from other causes) from an analysis of one group to the other group may lead to erroneous conclusions.

II. Data

As indicated, our data are from the 1982 Social Security NBS, a survey of individuals who received their first Social Security payment between June 1980 and June 1981 (Ycas 1992). The NBS contains samples of all new SSA beneficiaries, including retired workers, disabled workers, spouses, surviving spouses, divorced spouses, and surviving divorced spouses. As a sample of new beneficiaries, the NBS is not representative of the U.S. population or even of all Social Security beneficiaries. The sample was first interviewed in 1982, and surviving respondents and surviving spouses were resurveyed in the 1991 New Beneficiary Followup (NBF) survey. Social Security earnings and master beneficiary records, with information on deaths from Social Security Administration files, were linked to individuals in the sample.

The NBS sample selected for this analysis is composed of the disability beneficiaries who were aged 20–64 when they first applied for benefits, and of retired workers who were 62–72 at that time.[3] In the 1982 samples, there were 9,065 retired and 5,167 disabled workers. At the 1991 reinterview, 31 percent of the retired workers and 39 percent of the disabled workers had attrited from the sample, with the majority (21.7 percent of the retired workers and 30.8 percent of the disabled workers) leaving because of death.

III. Attrition Patterns Among Socioeconomic Groups

Table 1 compares the demographic characteristics of the 1982 samples of retired workers and disabled workers with those of the observations that have attrited from each sample between 1982 and 1991, and those that remain in the 1991 samples. The fourth column presents the results from a Kolmogorov-Smirnov test designed to reveal statistically significant differences between the characteristics of the observations who remain in the sample and those who attrited from it.[4] The last column of each panel shows the ratio of the percentage of observations with a particular characteristic in the 1991 sample to the percentage with that characteristic in 1982; numbers in excess of unity indicate a rate of attrition below the average of the 1982 sample.

The higher rate of attrition for the disabled sample (39 percent) than for the retired sample (31 percent) is due entirely to higher mortality rates among the former. Clearly, the adverse health conditions that qualify disabled individuals for Social Security Disability Insurance (DI) benefits and the higher percentage of males in the sample override the younger average age of the disabled sample in explaining the higher prevalence of mortality in this group.

Table 1.—Attrition in the retired- and disabled-worker samples, by demographic characteristics

Characteristic	1982 sample	Attriters	1991 sample	Kolmogorov-Smirnov p-value	1991/1982
Retired sample					
Total number	9,065	2,840	6,225		0.69
Death		1,966			
Percent due to death		69%			
Race:					
White	89%	89%	90%	1.00	1.00
Black	9%	10%	9%	1.00	.97
Other	2%	2%	2%	1.00	1.00
Gender:					
Male	58%	65%	55%	.00	.95
Female	42%	35%	45%		1.08
Age in 1982:					
62–64	32%	29%	33%	.00	1.04
65–69	61%	61%	60%	1.00	1.00
70–72	4%	5%	4%	.96	.91
73–75	3%	4%	2%	.51	.81
Marital status in 1982:					
Married	74%	71%	75%	.02	1.01
Single	26%	29%	25%		.96
Income in 1982:					
$9,999 or less	9%	11%	9%	.55	1.00
$10,000–$19,999	25%	26%	24%	.63	.96
$20,000–$29,999	27%	28%	27%	.60	1.00
$30,000–$39,999	17%	16%	18%	.58	1.06
$40,000–$49,999	8%	8%	9%	.80	1.12
$50,000–$59,999	5%	4%	5%	1.00	1.00
$60,000 or more	8%	8%	9%	.92	1.12
Disabled sample					
Total number	5,167	2,009	3,158		.61
Death		1,592			
Percent due to death		79%			
Race:					
White	81%	81%	81%	1.00	1.00
Black	16%	16%	16%	1.00	1.00
Other	3%	3%	3%	1.00	1.00
Gender:					
Male	69%	74%	67%	.00	.97
Female	31%	26%	33%		1.06
Age in 1982:					
20–35	12%	9%	14%	.00	1.17
36–45	10%	7%	12%	.00	1.20
46–55	23%	20%	24%	.02	1.04
56–65	55%	64%	49%	.00	.89
Marital status in 1982:					
Married	66%	67%	65%	.72	.98
Single	34%	33%	35%		1.03
Income in 1982:					
$9,999 or less	27%	26%	29%	.28	1.07
$10,000–$19,999	31%	33%	31%	.66	1.00
$20,000–$29,999	20%	21%	19%	.79	.95
$30,000–$39,999	12%	11%	12%	1.00	1.00
$40,000–$49,999	5%	5%	5%	1.00	1.00
$50,000–$59,999	2%	2%	2%	1.00	1.00
$60,000 or more	2%	2%	2%	1.00	1.00

A few important differences in the socioeconomic characteristics of the three groups (the entire 1982 sample, the attriters, and the surviving 1991 sample) are worth noting. As indicated in the final column, older individuals have a higher attrition rate than younger people for both the disabled and retired samples. The column 4 p-value for the youngest age groups (those aged 62–64 in 1982 for the retired sample; those aged 20–35 for the disabled sample) indicates that the proportion of the sample of survivors in the youngest age group is significantly greater than the proportion of attriters in that age group for both the retired and disabled samples.[5] For the disabled workers, the p-value indicates statistically significantly greater attrition rates for the oldest age group and smaller attrition rates for the remaining groups. Similarly (and consistent with mortality risks), males account for a higher proportion of the attriters than of the survivors. A somewhat higher percentage of unmarried than married retired workers attrited from the sample; although the difference in attrition rates between married and unmarried retired workers is statistically significant, it is quantitatively small. This difference is not observed for the disabled. Statistically significant attrition differences are not observed for the other characteristics included in the table.

IV. Multivariate Correlates of Attrition

While the comparisons in table 1 suggest a number of differences in the characteristics of the 1991 survivors and those who attrited from the 1982 samples, they fail to reveal the independent relationship of any individual characteristic to the probability of attrition, holding constant other observed characteristics. In the following sections, we estimate bivariate probit models describing the relationship of the socioeconomic characteristics of the 1982 samples of both retired and disabled workers to the probability of leaving the sample for any reason, attriting due to death, and (conditional on not dying) attriting for any nondeath reason.[6]

Correlates of Overall Attrition

Table 2 presents the results of a probit model revealing the relationship of a set of individual background and socioeconomic characteristics to the probability of attriting from the sample. Coefficients, together with an indicator of their statistical significance, are shown in the first column of each panel. The second column of each panel presents the simulated changes in the predicted probability of attrition associated with changes in that specific independent variable, holding all other variables at their mean values.[7] The included variables were chosen because of their expected causal relationship to the probability of attrition (for example, age, gender, health status) or general interest in the socioeconomic correlates of attrition (for example, race, marital status, and region). We expect that various health conditions will have differential effects on the probability of attrition (primarily because of their expected relationship to the probability of dying), and hence include the full set of them in the analyses.[8] We also include a variable describing the number of health conditions in order to better understand the relationship between the severity of health conditions and the probability of attrition, especially dying. For the dependent variables, a 1982 observation is assigned a value of one if that person attrited from the sample, and a zero if that person remained in the sample until 1991.

The patterns among the demographic groups shown in table 1 are largely confirmed by the probit results for both samples.[9] Gender has a large substantive impact on attrition. Holding all other factors constant, both retired and disabled men are approximately 10 percent more likely to attrite than are retired or disabled women. After controlling for other factors, there is not a statistically significant difference between whites and nonwhites in the probability of attrition for either sample. Age, however, is related to the probability of attrition and is quantitatively important. Among the retired, an additional year of age adds 1.5 percentage points to the probability of attrition; among the disabled, the impact is about one-half of this amount. Somewhat curiously, while the number of years of education of retired workers is negatively associated with the probability of attrition, the relationship is positive for the sample of disabled workers. In neither sample is the quantitative magnitude of education differences large; an additional year of schooling decreases the probability of attrition by 0.5 points for retired workers, but increases it by 0.9 points for disabled workers.

Both being married and increasing the number of minor children in the household by one increase the probability of attriting from the sample by more than 6 points and 4 points, respectively, for retired workers. Being married has a smaller effect on the probability of attrition of disabled workers. The number of minor children in the household has no significant relationship to the probability of attriting for the disabled group. When other factors are controlled for, income is not statistically significantly related to the probability of attriting for either group.

Table 2 also includes estimates of the relationship between each of a number of health-related characteristics and the probability of attrition. These include: (1) having private health insurance,[10] (2) having a specified health condition, (3) having worked in a relatively hazardous job (as measured by the fatality rate of the longest or last job), and (4) the time period since the end of the person's last work spell prior to the 1982 interview. The last two variables are included in an attempt to measure work-related health conditions that are not explicitly identified in the NBS interview. The job fatality rate is included as an attempt to capture the severity of work-associated disabilities that would have led to retirement or disability recipiency.[11] The time period variable is included to control for the existence of a health problem that might lead to a period of nonwork prior to benefit application or receipt.

For both the retired and the disabled groups, the number of health conditions is positively related to the probability of attriting from the sample, with one more condition increasing the chances of attrition by 6 percentage points for the retired-worker sample and 2 percentage points for the disabled-worker

sample. Retired workers who report having cancer in 1982 have an 11 percentage point greater probability of attriting from the sample, while disabled workers have a 26 percentage point greater probability of attriting.[12] The variable describing the extent of hazardous working conditions on the longest (or last) job has no apparent relationship to the probability of attrition for either sample of new beneficiaries. The length of time since the end of the person's last work spell prior to receipt of benefits is negatively associated with attrition among retired workers but is not significantly related to the probability of attrition among disabled workers.

Finally, living in a metropolitan area (in 1982) is positively associated with a higher probability of attrition; the magnitude of effect is nearly 6 percentage points for retired workers and 5 percentage points for disabled workers. Living in the northeast and west is positively associated with attriting for the disabled sample.

The overall pattern of relationships between these character-

Table 2.—Probit estimates of probability of attriting from sample for any reason

Characteristic	Retired workers		Disabled workers	
	Coefficient	Change in probability of leaving sample [1]	Coefficient	Change in probability of leaving sample [1]
Constant	-3.2303 **		-1.8861 **	
Male (v. female)	.2871 **	0.0988	.2725 **	0.1021
Nonwhite (v. white)	-.0359	-.0125	.0513	.0197
Years of education	-.0151	-.0053	.0235 **	.0090
Age in 1982	.0417 **	.0146	.0194 **	.0074
Married (v. nonmarried)	-.1695 **	-.0604	-.0827 *	-.0317
Income (thousands of dollars)	-.0003	-.0001	-.0017	-.0007
Number of minors in household	-.1253 **	-.0438	-.0136	-.0052
Private health insurance (v. no private health insurance)	-.1067 **	-.0379	-.0643	-.0246
Number of health problems	.1728 **	.0604	.0536 *	.0205
Health condition (v. no health condition):				
Cancer (malignant tumor)	.3012 **	.1117	.5513 **	.2168
Any heart problem	-.1028 **	-.0357	-.0107	-.0041
Blindness or eye trouble	-.1346 **	-.0457	-.1057 *	-.0401
Eye or retina condition	-.1882 **	-.0632	.0385	.0148
Trouble hearing with aid	-.2753 **	-.0908	-.2088 **	-.0780
Missing arm, leg, hand	.1531	.0554	.3513 **	.1383
Bone or muscle condition	-.2882 **	.1003	-.1642 **	-.0631
Stiffness or deformity	-.1254 **	-.0428	-.2373 **	-.0903
Nervous system condition	-.0166	-.0058	-.0443	-.0169
Other paralysis	.1446	.0522	.1082	.0419
Lung or respiratory condition	.0367	.0129	.2799 **	.1083
Digestive system condition	-.0326	-.0114	.0562	.0215
Fatality rate (longest occupation or last job worked)	.0019	.0007	.0012	.0004
Time since last employment (v. 1–2 years):				
Less than 1 year		-.0228	-.0865	-.0326
3–5 years	.0526	.0186	.0626	.0240
6–10 years	.0872	.0311	.0992	.0384
11 or more years	.0367	.0129	.4288	.1692
SMSA residence (v. non-SMSA residence)	.1706 **	.0584	.1431 **	.0542
Region of residence (v. North Central):				
South	-.0482	-.0168	.0036	.0014
Northeast	.0438	.0154	.1036 *	.0399
West	.0150	.0052	.1256 **	.0485
Number of observations	8,697		4,797	
Log likelihood ratio	-5180.76		-3003.35	

*Statistically significant at the 0.10 level.

**Statistically significant at the 0.05 level.

[1]Change is from 0 to 1 for dummy variables. For continuous variables the change is one unit from the mean. Probabilities are evaluated at the means for all other variables.

istics and attriting from the sample appears similar for the retired-worker and disabled-worker samples. For both groups, males, older individuals, those in poor health, and those who live in urban areas are more likely to attrite. Nevertheless, a chi-square test of whether or not the model is the same for both groups is rejected at the 1-percent level of significance.[13]

Correlates of Attrition Due to Mortality

In this section we study the correlates of attrition from the NBS over the 1982–91 period due to the death of the respondent. Because we are able to rely only on information contained in the NBS, we cannot analyze the relationship between observed mortality and the reason that a beneficiary applies for benefits (for example, the specific disabling condition that qualified the disabled for benefits).[14] However, the NBS does precisely measure mortality, hence allowing examination of the relationship of personal and employment characteristics to sample attrition due to mortality and a comparison of these effects with the effects of these characteristics on attrition due to all other reasons (primarily, refusal to be interviewed).

Table 3 presents results on the probability of attriting due to death. Because mortality is the dominant reason for attrition (table 1), the results in table 3 are similar to those in table 2. We emphasize the few important differences between them. Not surprisingly, age and gender play a somewhat larger role in attrition due to death. The probability that a retired male will die over the 1982–91 period is nearly 12 points greater than that for retired women; among the disabled the probability is 10 points higher for males. Each additional year of age raises the probability of attrition due to death by nearly 2 points for the retired sample, but only about 1 point for the disabled. As with the table 2 results, years of education have different effects for the retired and disabled samples; again the quantitative effects are very small in both cases.[15]

For both the retired and disabled samples, income is negatively related to the probability of mortality, but only for retired workers is it statistically significant. In both cases, the quantitative effect is very small. As with the table 2 results, being married and the number of minors in the household are negatively related to the probability of death among both the retired and disabled groups, suggesting the importance of family support and resources in maintaining health status; for the disabled sample, only being married is statistically significant. Having private insurance is negatively related to the probability of mortality for both groups, but is statistically significant only for the retired sample.

The number of health problems is positively associated with the probability of death for both of the groups, but is statistically significant for only the retired sample. For this group, each additional health condition is associated with an increased probability of death of nearly 6 percentage points.

Several of the coefficients on the health conditions are statistically significant and quantitatively large. The patterns differ somewhat between the two groups. Particularly large effects are indicated for potentially life-threatening illnesses such as cancer and lung/respiratory/digestive conditions. These effects are particularly large for the disabled sample, consistent with eligibility criteria for DI that include having severe and life-threatening health disorders. The fatality rate on the last job is positively related to the probability of mortality for the retired sample, but the effect is quantitatively small. As with the results in table 2, both retired and disabled workers who have recently worked have a significantly lower probability of leaving the sample because of death, and for the disabled, the effect is quantitatively large (over 10 percentage points), suggesting that this variable may proxy for the severity of health conditions not otherwise measured in the data. A chi-square test of whether or not the same model applies to both samples is strongly rejected.[17]

Correlates of Attrition Due to Nonmortality Reasons

Apart from mortality, "refusal to be interviewed" is the primary reason for attrition from the NBS sample. The relationships of personal and employment characteristics to this nonresponse source of attrition are shown in table 4. The estimates in table 4 are from a probit model fit to those observations who are still living in 1991, as this reason for attrition is relevant only for the surviving members of the 1982 sample of new beneficiaries.

There are relatively few statistically significant relationships revealed in table 4, suggesting that refusal to be interviewed is a more random reason for attrition than is mortality. While older benefit recipients have a higher probability of attrition and mortality (tables 2 and 3), these characteristics are negatively related to the nonresponse reason for attrition, though only for the disabled sample is the relationship statistically significant. For the disabled group, the probability of attriting due to nonresponse is positively and significantly related to the number of health problems. Interestingly, while geographic variables had no effect on the probability of mortality, living in an urban area is positively and significantly associated with the probability of attrition due to nonresponse, as is living in the northeast or west. The urban effect may be due to greater mobility of urban residents; the regional effects are unexplained, but they may serve as indicators of different practices used to determine eligibility for receipt of disability benefits among Social Security Administration regional offices. Again, the test of whether the same model applies to both samples is rejected at the 1-percent level.[18]

V. Conclusions

We have compared patterns of attrition between two groups of Social Security beneficiaries—retired workers and disabled workers—and have explored the effects of personal, health, and employment conditions on the probability of attrition due to mortality and due to other reasons (primarily, unwillingness to respond to the interview). Mortality is precisely measured in the NBS data, enabling a reliable comparison of reasons for attrition

due to mortality and attrition due to other reasons. Also, the unusually large sample of severely disabled persons allows us to examine how initial health conditions of persons in the sample are differentially related to attrition due to death and to other reasons.

Users of panel data need to understand the pattern and reasons for attrition since selective attrition may lead to misleading conclusions about changes in population characteristics over time. Such an exploration is especially important when using data from somewhat unique samples that may not conform to well-known patterns of attrition from representative population samples. The NBS is not representative of either the national population or even of all Social Security beneficiaries and therefore, not surprisingly, exhibits attrition patterns that are shaped in part by the nature of its unique sampling design.

This exploration of attrition in the NBS is of particular import to researchers who wish to exploit the longitudinal nature of the NBS sample in exploring time-related behavioral patterns.

Table 3.—Probit estimates of probability of attriting from sample due to death

Characteristic	Retired workers		Disabled workers	
	Coefficient	Change in probability of leaving sample [1]	Coefficient	Change in probability of leaving sample [1]
Constant	-5.1837 **		-2.7461 **	
Male (v. female)	.4297 **	0.1164	.3079 **	0.1023
Nonwhite (v. white)	.0085	.0024	.1077 *	.0378
Years of education	-.0145 **	-.0041	.0252 **	.0087
Age in 1982	.0659 **	.0185	.0313 **	.0108
Married (v. nonmarried)	-.1819 **	-.0528	-.0924 *	-.0321
Income (thousands of dollars)	-.0018 **	-.0005	-.0009	-.0003
Number of minors in household	-.0779 **	-.0218	-.0060	-.0021
Private health insurance (v. no private health insurance)	-.0870 **	-.0249	-.0281	-.0097
Number of health problems	.2096 **	.0587	.0203	.0070
Health condition (v. no health condition):				
Cancer (malignant tumor)	.3468 **	.1096	.6997 **	.2655
Any heart problem	-.0697	-.0194	.0852	.0292
Blindness or eye trouble	-.1297 *	-.0347	.0018	.0006
Eye or retina condition	-.2710 **	-.0692	-.0147	-.0051
Trouble hearing with aid	-.3497 **	-.0873	.1849 **	-.0616
Missing arm, leg, hand	.2322	.0711	.3256 **	.1201
Bone or muscle condition	-.3183 **	-.0889	-.1953 **	-.0682
Stiffness or deformity	-.1478 **	-.0397	-.2579 **	-.0883
Nervous system condition	-.1039	-.0279	-.0500	-.0170
Other paralysis	.2272	.0694	.1721 **	.0614
Lung or respiratory condition	.0998 *	.0289	.3168 **	.1126
Digestive system condition	.0008	.0002	.1648 **	.0574
Fatality rate (longest occupation or last job worked)	.0025 *	.0007	.0010	.0003
Time since last employment (v. 1–2 years):				
Less than 1 year	-.0723 **	-.0201	.3362 **	-.1055
3–5 years	.1102 *	.0321	.0769 *	.0267
6–10 years	.0796	.0230	.0534	.0186
11 or more years	.0863	.0250	.2963	.1090
SMSA residence (v. non-SMSA residence)	.0531	.0147	.0490	.0168
Region of residence (v. North Central):				
South	-.0181	-.0051	.0294	.0101
Northeast	-.0098	-.0027	.0490	.0170
West	-.0561	-.0155	-.0010	-.0003
Number of observations	8,697		4,797	
Log likelihood ratio	-4227.78		-2685.59	

*Statistically significant at the 0.10 level.
**Statistically significant at the 0.05 level.
[1] Change is from 0 to 1 for dummy variables. For continuous variables the change is one unit from the mean. Probabilities are evaluated at the means for all other variables.

Assuming determinants of attrition inferred from samples that include a large percentage of nonretired or nondisabled is not appropriate to this sample, whose initial status depends on age and health conditions that are related to attrition. But our analysis also has implications for the exploration of attrition in other data sets. This analysis shows that attrition due to death and attrition due to refusal are quite different processes, even though health conditions play a role in both processes. Age affects attrition differently, with greater age increasing the chances of mortality but reducing the chances of refusal. Interestingly, the health conditions that increase the chances of death appear to have no effect on the probability of refusal. In

Table 4.—Probit estimates of probability of attriting from sample for reason other than death

Characteristic	Retired workers		Disabled workers	
	Coefficient	Change in probability of leaving sample [1]	Coefficient	Change in probability of leaving sample [1]
Constant	-0.4339		-1.3172 **	
Male (v. female)	.0093	-0.0018	.0745	0.0129
Nonwhite (v. white)	-.1203	-.0220	-.0822	-.0140
Years of education	-.0073	-.0014	.0097	.0017
Age in 1982	-.0110	-.0021	-.0072 **	-.0013
Married (v. nonmarried)	-.0891 *	-.0177	-.0229	-.0040
Income (thousands of dollars)	.0009 *	.0002	-.0028	-.0005
Number of minors in household	-.1748 **	-.0339	-.0165	-.0029
Private health insurance (v. no private health insurance)	-.0861 *	-.0172	-.1041	-.0184
Number of health problems	.0239	.0046	.1159 **	.0204
Health condition (v. no health condition):				
Cancer (malignant tumor)	.1280	.0266	-.1320	-.0215
Any heart problem	-.0910	-.0174	-.2190 **	-.0394
Blindness or eye trouble	-.0803	-.0150	-.3133 **	-.0503
Eye or retina condition	.0219	.0043	.1305	.0243
Trouble hearing with aid	-.0331	-.0063	-.1708 *	-.0281
Missing arm, leg, hand	-.2322	-.0390	.2330	.0470
Bone or muscle condition	-.0984	-.0190	-.0201	-.0035
Stiffness or deformity	-.0136	-.0026	-.0914	-.0161
Nervous system condition	.1603	.0340	-.0166	-.0029
Other paralysis	-.1672	-.0293	-.0990	-.0165
Lung or respiratory condition	-.1137	-.0209	.0540	.0097
Digestive system condition	-.0450	-.0086	-.2074 **	-.0350
Fatality rate (longest occupation or last job worked)	.0003	.0001	.0016	.0003
Time since last employment (v. 1–2 years):				
Less than 1 year	-.0269	-.0052	.2122 *	.0417
3–5 years	-.0630	-.0118	.0021	.0004
6–10 years	.0581	.0116	.1432	.0273
11 or more years	-.0551	-.0104	.5897	.1442
SMSA residence (v. non–SMSA residence)	.3226 **	.0572	.3155 **	.0516
Region of residence (v. North Central):				
South	-.0847	-.0161	-.0590	-.0103
Northeast	.1037 *	.0208	.1939 **	.0367
West	.1241 **	.0253	.3357 **	.0680
Number of observations	6,817		3,280	
Log likelihood ratio	-2450.08		-1087.64	

*Statistically significant at the 0.10 level.

**Statistically significant at the 0.05 level.

[1] Change is from 0 to 1 for dummy variables. For continuous variables the change is one unit from the mean. Probabilities are evaluated at the means for all other variables.

general, it appears that refusal to be interviewed is more evenly spread across populations and characteristics than is death. Thus, it is important that panel studies identify attrition due to death precisely. Misidentification of death as refusal to be interviewed would mislead researchers into inferring more selective attrition than might be the case.

The exploration of attrition among retired and disabled workers points to the importance of understanding initial sample characteristics when examining attrition patterns. Attrition patterns are different for these two groups. For example, more years of education reduces attrition among retired workers but increases it among disabled workers. We suggested earlier that this may be because, among disabled workers, more highly educated workers work at jobs at which they can remain until their disabling condition is further advanced. Private health insurance reduces attrition among retired workers, perhaps because these workers are somewhat economically better off, but has no effect among disabled workers, perhaps because this group are all selected into that status by a severe health condition. Likewise, health conditions that are more likely to be fatal have a larger impact on attrition among the disabled than among retired workers. As the disabled age, however, it is not clear if these differences will continue. The implication is that attrition studies should continue over the lifetime of panels in order to understand the attrition and selectivity issues better.

This study joins a prior study on attrition from the NBS. Iams and McCoy (1991) studied mortality in the NBS retired-worker sample, finding results that complement ours. While they concluded that their model of survival status "confirms many of the relationships reported in the literature pertaining to survival," they went further and suggested that health measures could be used to forecast future mortality probabilities for both retired and disabled workers. Our results, however, suggest that measured health effects on mortality may reflect the process by which individuals gain beneficiary status to the DI program, rather than independent effects; caution is required in using results measured over one group of recipients to infer mortality patterns for other groups. The difference in factors explaining attrition for those in the retired and disabled samples suggests further caution in attempting to predict "survival status" among different socioeconomic or beneficiary status groups.

Notes

Acknowledgments: We would like to acknowledge the research assistance of Melissa Mullikin in this project, as well as the excellent technical typing of Dawn Duren. Financial assistance from the Social Security Administration, the La Follette Institute of Public Affairs and the Institute for Research on Poverty, both of the University of Wisconsin-Madison, supported this research.

[1] This problem is particularly acute in the case of attrition from medical experiments, where attrition may be related to the outcome under study. See Philipson and Hedges (1998). In general, the problem exists whenever the smaller, post-attrition population is viewed as representative of the pre-attrition population.

[2] Note that the reasons for attrition may not be independent. For example, terminally ill individuals may be institutionalized or otherwise unable or unwilling to respond. Hence, even if the investigator has accurate records of attrition due to death, the appropriate treatment of observations who are likely to die shortly remains. Similarly, a surviving spouse may be unwilling to respond to an interview shortly after the death of the other spouse. The first example raises the question of whether the terminally ill observation should be treated as a deceased person or as an attriter with missing information. In the second example, the nonresponse may be temporary (with some information available from subsequent interviews) or it may be permanent.

[3] We excluded retired workers who came on the rolls after age 72. At the time the sample was selected, age 72 was the age at which the earnings test ended. Persons coming on the rolls after age 72 were a unique group in that they postponed retirement without ostensible gain to doing so. Examination of their characteristics suggests that these people have long work careers and remain relatively high earners.

[4] The Kolmogorov-Smirnov two-sample test makes no assumptions about underlying distributions of the samples across the variable categories and hence is based on less restrictive distributional assumptions than parametric tests such as a t-statistic. Column 4 shows the p-value of the test.

[5] The ages in the table are from the 1982 interview. This coincides with the date at which the socioeconomic characteristics were reported.

[6] A bivariate probit technique is employed rather than a multinomial logit model (with the reasons for attrition as the absorbing categories). The latter model requires the assumption that inclusion or exclusion of any option does not affect selection among the other options (the so-called "independence of irrelevant alternatives" assumption), which assumption would appear to be violated given that one of the states is death and only those who do not die have the ability to respond to the interview or refuse. Apart from this, estimated coefficients in the multinomial logit model describe effects relative to those of the excluded option, making interpretation difficult.

[7] For discrete variables, we show the change in the probability associated with moving from a value of zero to one. For continuous variables, the change in the probability associated with a unit change measured at the mean of the independent variables (for example, 1 year in the case of age) is simulated.

[8] Each observation is assigned a value of one if they report the presence of a listed condition.

[9] About 4 percent of the retired workers and 6 percent of disabled workers included in table 1 have at least one missing value on the set of independent variables. Because the missing values are primarily across the health status variables, which are key to our analysis, these cases are not included in tables 2–4.

[10] We use private health insurance coverage, as both beneficiary groups are eligible for Medicare for most of the interval, 1982–91. DI beneficiaries are eligible for Medicare 2 years after becoming a beneficiary; retired-worker beneficiaries become eligible at age 65.

[11] The data on occupational fatality rates were drawn from the U.S. Department of Labor, Bureau of Labor Statistics (1994). See Appendix A.

[12] Reported coefficients measure the association of the reported

condition relative to not having that reported condition and the probability of attrition. Individuals can report multiple conditions.

[13] The test statistic is 147.85, which is distributed chi-square with 30 degrees of freedom.

[14] The NBS does not contain information on the nature and severity of health conditions that qualified the disabled sample for benefits; only age and earnings matter in regard to eligibility for retired-worker benefits.

[15] One possible explanation of the positive effect of education for the disabled sample is that the employment arrangements of more educated disabled people may allow them to delay benefit receipt until their health conditions are severe and life threatening; education may proxy for this unobserved severity of health problems among the disabled. While we control for health conditions in the analysis, these variables do not indicate the severity or life-threatening character of the condition.

[16] While the overall criteria for eligibility for DI benefits is inability to engage in substantial gainful activity, the presence of severe health conditions that are clearly life threatening would qualify one for DI.

[17] The chi-square test statistic is 256.78, with 30 degrees of freedom; it is significant at the 1-percent level.

[18] The test statistic in this case is 49.09, smaller than that for attrition due to death but still statistically significant at the 2-percent level.

References

Iams, Howard M. and John L. McCoy. 1991. "Predictors of Mortality Among Newly Retired Workers." *Social Security Bulletin,* Vol. 54, No. 3, pp. 2–10.

Fitzgerald, John; Peter Gottschalk; and Robert Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources,* Vol. 33, No. 2, pp. 251–299.

Philipson, Tomas and Larry Hedges. 1998. "Subject Evaluation in Social Experiments." *Econometrica* , Vol. 66 (March), pp. 381–408.

Ycas, Martynas A. 1992. "The New Beneficiary Data System: The First Phase." *Social Security Bulletin*, Vol. 55, No. 2, pp. 20–35.

Zabel, Jeffrey E. 1998. "An Analysis of Attrition in the Panel Study of Income Dynamics in the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior," *Journal of Human Resources,* Vol. 33, No. 2, pp. 480–506.

Appendix A

Variable Definitions

Sex: dummy variable
= 1 if the respondent is male
= 0 otherwise

Race: dummy variable
= 1 if the respondent indicates that he/she is Asian, black, Indian, or other
= 0 if the respondent is white

Married: dummy variable
= 1 if the respondent is married in 1982
= 0 otherwise

Income: The respondent's annual family income in 1982 in 1994 dollars.

Number of minors in household in 1982: The number of individuals aged 18 or younger living in the household in 1982.

Fatality rate (at last job worked): The fatality rate of the respondent's longest occupation. If the respondent's longest occupation was not held at some point between the ages of 45 and 50, then the fatality rate of the respondent's 1982 occupation was used. If the respondent was not employed in 1982, then the fatality rate of the last occupation was used. The data on occupational fatality rates were drawn from the U.S. Department of Labor, Bureau of Labor Statistics (1994).

Time since last employment: dummy variables
=1 if the respondent's last job ended in given period of time before NBS interview
=0 otherwise
Less than 1 year
1–2 years (omitted)
3–5 years
6–10 years
11 or more years

Region of residence: dummy variables
=1 if the respondent lives in the following region
=0 otherwise
South
Northeast
West
North Central (omitted)

Health problems: dummy variables
= 1 if the respondent has each variable problem
= 0 otherwise

- Cancer (malignant tumor): cancer or a malignant growth not already mentioned
- Any heart problem: heart problems, such as hardening of the arteries, high blood pressure, or chest pain
- Blindness or eye trouble: blindness or serious trouble seeing with one or both eyes, even when wearing glasses
- Eye or retina condition: cataracts, glaucoma, or any other condition affecting the eye or retina
- Trouble hearing with aid: deafness or serious trouble hearing with one or both ears, even when wearing a hearing aid
- Missing arm, leg, hand: a missing arm, leg, foot, or hand
- Bone or muscle condition: arthritis, rheumatism, or any other condition affecting the bones or muscles
- Stiffness or deformity: permanent stiffness or any other deformity of the foot, leg, fingers, arm, or back
- Nervous system condition: multiple sclerosis, cerebral palsy, epilepsy, or any other condition affecting the nervous system
- Other paralysis: paralysis of any kind not already mentioned
- Lung or respiratory condition: asthma, emphysema, or any other condition affecting the lungs or respiratory system, including work-related respiratory condtions such as silicosis or pneumoconiosis
- Digestive system condition: gallbladder, stomach, kidney, or liver trouble; diabetes; or any other condition affecting the digestive system

SMSA residence: dummy variable
= 1 if lives in SMSA (urban area) in 1982
= 0 otherwise

Reference

U.S. Department of Labor, Bureau of Labor Statistics. 1994. *Census of Fatal Occupational Injuries*. Washington, DC: Government Printing Office.

Appendix B

Means of independent variables used in probits

Variable	Disabled workers				Retired workers			
	Entire sample		Survivors in 1991		Entire sample		Survivors in 1991	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Demographic variables:								
Sex (male = 1, female = 0)	0.70	0.46	0.67	0.47	0.59	0.49	0.55	0.50
Race (nonwhite = 1, white = 0)	.19	.39	.19	.39	.11	.31	.10	.30
Years of school	10.16	3.34	10.17	3.38	11.27	3.35	11.38	3.32
Age	53.29	10.41	51.80	11.10	66.07	2.35	65.96	2.27
Married (married = 1, otherwise = 0)	.68	.47	.67	.47	.74	.44	.74	.44
Number of minors in household	.50	1.03	.55	1.08	.09	.43	.09	.44
Economic variables:								
Income	21.10	16.32	20.94	16.60	32.25	34.30	32.85	35.49
Private health insurance (yes = 1, no = 0)	.55	.50	.54	.50	.79	.41	.80	.40
Health conditions:[1]								
Number of health problems	4.02	1.97	3.91	1.98	1.92	1.68	1.81	1.62
Cancer (malignant tumor)	.08	.28	.05	.22	.04	.19	.03	.16
Any heart problem	.62	.49	.58	.49	.37	.48	.36	.48
Blindness or eye trouble	.28	.45	.27	.45	.10	.30	.09	.29
Eye or retina condition	.15	.36	.14	.35	.11	.32	.11	.32
Trouble hearing with aid	.19	.39	.20	.40	.14	.34	.13	.34
Missing arm, leg, hand	.03	.16	.02	.14	.01	.07	.00	.06
Bone or muscle condition	.64	.48	.65	.48	.49	.50	.49	.50
Stiffness or deformity	.47	.50	.50	.50	.16	.37	.15	.36
Nervous system condition	.12	.32	.12	.33	.01	.12	.01	.11
Other paralysis	.08	.28	.08	.28	.01	.10	.01	.09
Lung or respiratory condition	.29	.45	.24	.43	.13	.33	.11	.31
Digestive system condition	.37	.48	.35	.48	.18	.38	.16	.37
Work-related variables:[2]								
Fatality rate at longest occupation	9.31	12.19	9.08	11.83	7.38	11.60	7.03	11.16
Less than 1 year since last worked	.06	.24	.08	.27	.34	.47	.34	.47
1–2 years since last worked	.63	.48	.63	.48	.48	.50	.48	.50
3–5 years since last worked	.27	.44	.26	.44	.08	.28	.08	.27
6–10 years since last worked	.04	.19	.03	.18	.05	.22	.05	.22
Greater than 11 years since last worked	.00	.07	.00	.07	.05	.21	.05	.21
Geographic variables:								
SMSA (resides in an SMSA)	.69	.46	.68	.47	.73	.44	.73	.44
Resides in South	.38	.49	.38	.49	.33	.47	.32	.47
Resides in Northeast	.20	.40	.19	.40	.22	.41	.22	.41
Resides in North Central	.26	.44	.26	.44	.29	.45	.29	.45
Resides in West	.16	.36	.16	.36	.16	.37	.17	.37
Number of observations	4,797		3,280		8,697		6,817	

[1] If have health condition, variable value = 1. Respondents may have more than one condition.

[2] For interaction variable, sex = 1 if male, 0 otherwise.