THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

A RANDOM-EFFECTS APPROACH TO ATTRITION BIAS IN THE SIPP HEALTH INSURANCE DATA

No. 144 9104

J. A. Klerman
The Rand Corporation

U. S. Department of Commerce BUREAU OF THE CENSUS

A Random-Effects Approach to Attrition Bias in the SIPP Health Insurance Data

Jacob Alex Klerman
The Rand Corporation

ABSTRACT

Using black male health insurance data from the 1984 SIPP Panel as an example, this paper describes a test for the presence of attrition bias and a consistent estimator for the level of health insurance coverage in the presence of attrition bias. The test and estimator jointly model the attrition and health insurance coverage processes using random effects panel probit models in which the random effects are allowed to be correlated. The empirical results suggest that for black males, ignoring attrition bias leads to a positive time trend for health insurance coverage, when the true, corrected for attrition bias, time trend is negative.

KEYWORDS

attrition, random effects

INTRODUCTION

The SIPP health insurance coverage data are being widely cited in the policy debate concerning how to extend health insurance coverage to the approximately 15 percent of Americans who are currently uninsured. SIPP results, computed from panel data, often differ from results from the CPS, computed from a dwelling-unit survey. One possible explanation for the differences is non-random panel attrition. Less than seventy percent of the eligible individuals complete the full set of SIPP interviews. Furthermore, two- thirds of those with any missing interviews are missing two or more. The combination of the stringent data requirements of conventional longitudinal data analysis (analyzing only cases with a complete set of interviews) and these high non-interview rates focuses attention on sample attrition and its implications for the validity of conclusions drawn from analysis of the SIPP data.[1]

SIPP attrition is not random. Attriters are disproportionately young, male, black, poor, program participants and frequent movers. (McArthur 1988) Differential attrition with respect to *observable characteristics* (age, sex, race) is correctable by traditional reweighting techniques. The differential attrition with respect to income and program participation, however, suggests that attrition may also be differential with respect to *unobservable characteristics*. If the unobserved characteristics that influence attrition also effect the behavior of interest, then traditional reweighting schemes will not yield consistent estimates of the levels of health insurance coverage.

Using the SIPP health insurance data for the black males in the 1984 panel as an example, this paper describes and implements a random effects probit scheme which yields a test for the presence of attrition bias and consistent estimates of the trend in health insurance coverage levels even when such attrition bias is present. The paper proceeds in four sections. The next section discusses the policy context of the controversy over the levels of health insurance coverage. The second section describes the SIPP data used in this study and presents

[1]The author wishes to thank Lee Lillard whose comments during an discussion of weighting in the SIPP stimulated the ideas developed here. The programming support of Chris Peterson and Marion Oshiro are also gratefully acknowledged. Finally Sharon Koga provided outstanding secretarial support. Responsibility for all remaining errors remain with the author.

some simple heuristics for the detection of attrition bias. The third section formalizes the insights of those heuristics. It outlines a random effects probit model which incorporates the heuristic and presents the results of estimating the proposed model The fourth section discusses the implications of the empirical results and directions for future development of the formal model.

HEALTH INSURANCE COVERAGE

The United States is the only major western country without a system of universal national health insurance. Instead, receipt of health care and corresponding payment mechanisms are a patchwork quilt of private and government programs. Most Americans receive health insurance coverage (hereafter, despite the potential confusion with the sampling usage of the term, simply coverage) as a fringe benefit or an employment relation; either their own employment, or that of a spouse or parent. Most Americans over the age of 65 receive coverage through Medicare. Some poor Americans are covered by Medicaid. Many Americans fall into the cracks between these systems and are left without any health insurance coverage.

The standard information on levels and sources of coverage is the Current Population Survey (CPS). Since 1980, the March Demographic Supplement to the CPS has included a battery of questions on health insurance coverage. Table 1 summarizes a set of results from the 1988 CPS.[2] They suggest that over 30 million Americans, more than ten percent of the population, is not covered by any health insurance.

Table 2 presents similar tabulations for blacks, who are the focus of the empirical work that follows. The numbers are from Long (1987) and are based on the March 1985 CPS. They suggest that compared to whites, blacks are less likely to have employer based insurance and more likely to have public insurance. In addition, they are considerably more likely to be uninsured: 23% vs. 16% in the 18-64 age group.

Beyond the sheer size of the uninsured population, there has been a marked shift in the trend over time. As a result of the implementation of Medicaid, the percentage of the non-elderly population without health insurance fell sharply from 30.4% to 13.9% over the period 1963 to 1986 (Swarz 1984). The later figure appears to have been a trough. Since that time non-coverage rates have drifted up to the 16.6% figure implied in Table 1 (Swarz 1986). To advocates this arresting of the previous decline and possible increase are causes for considerable concern.

General concern for the welfare of the working poor and the apparent failure of the situation to improve over time has resulted in a plethora of proposals to ameliorate the situation. These proposals range from calls for radical reform, to proposals for incremental extensions of the current system to fill the gaps between the various pieces. Radical reforms usually involve some form of national health insurance. Incremental proposals often involve requiring all employers to provide health insurance for their employees (Monheit and Short 1989).

Such incremental plans were an issue in the 1988 presidential debates. Then Governor Dukakis, holding Massachusetts up as an example, called for national implementation of an incremental plan based on employer mandates. Then Vice President Bush opposed such proposals as unnecessary intervention in the employer-employee relationship. Various solutions continue to be considered in Congress and in the state legislatures, though the budget situation at both levels seems to be preventing much action. Given the active political status of these health insurance reform proposals, it is not surprising that advocates on both sides study each new estimate of the levels of coverage with great care (see the discussion in footnote 1 above).

---

[2]The numbers quoted here are from Moyer (1989). He is an economist in the office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Swarz and Purcell (1989) dispute Moyer's methods and conclusions. Their differences with Moyer concern the treatment of internally contradictory responses about the coverage of children in the household. They conclude "We believe the number of people without health insurance from the 1988 CPS, however, is about 33.5 million.. . We do not believe that there has been a dramatic decline in the real number of people without health insurance based on the 1988 CPS" (Swarz and Purcell, 1989, p. 196). Suffice it to say that the exact numbers are politically sensitive and contentious.

Type of Coverage	Total	Under 15	15-17	18-24	25-34	35-64	65 and Older
Any private or public insurance							
Persons covered	210.0	45.9	9.2	20.0	35.8	70.9	28.3
Percent covered	100.0%	21.9%	4.4%	9.5%	17.0%	33.8%	13.5%
Percent of population	87.1	87.1	85.5	76.7	83.2	88.5	99.1
Private health insurance							
Persons covered	181.4	38.3	8.0	17.5	32.3	65.2	20.1
Percent covered	100.0%	21.1%	4.4%	9.6%	17.8%	35.9%	11.1%
Percent of population	75.2	72.7	74.3	67.1	75.3	81.3	70.5
Employer-sponsored health insurance							
Persons covered	147.6	33.0	6.5	13.3	29.1	57.0	8.6
Percent covered	100.0%	22.4%	4.4%	9.0%	19.7%	38.6%	5.8%
Percent of population	61.2	62.7	60.3	51.1	67.7	71.1	30.2
Medicaid							
Persons covered	20.9	8.6	1.1	2.1	2.9	3.8	2.5
Percent covered	100.0%	40.9%	5.4%	10.0%	13.8%	18.1%	11.8%
Percent of population	8.7	16.3	10.5	8.0	6.7	4.7	8.6
Uninsured							
Persons not covered	31.1	6.8	1.6	6.1	7.2	9.2	0.3
Percent not covered	100.0%	21.8%	5.0%	19.5%	23.1%	29.7%	0.9%
Percent of population	12.9	12.9	14.5	23.3	16.8	11.5	0.9

Source: Preliminary tabulations from the March 1988 Current Population Survey, 17 April 1989.
Note: Medicare and other coverages are not shown separately. Persons can be in more than one insurance category.

TABLE 1: Health Insurance Status by Age and Employment Status
Preliminary Results from the March 1988 CPS source G.M. Moyer (1989)

	Employment Related	Public	Uninsured
Black	47	31	22
17 and under	42	33	25
18-64	55	21	23
65 and over	8	90	2
Nonblack	62	23	15
17 and under	67	14	19
18-64	70	14	16
65 and over	10	89	1

TABLE 2: Health Insurance Status
by Age and Employment Status for Black American
Tabulations Results from March 1985 CPS
source S. H. Long (1988)

Year: Quarter	Percentage Not Covered by Health Insurance
1986:1	14.4
1986:2	14.3
1986:3	14.3
1986:4	13.8
1987:1	13.8
1987:2	13.5
1987:3	13.5
1987:4	13.2
1988:1	12.8
1988:2	12.7
1988:3	12.7
1988:4	13.0

TABLE 3: Percent Not Covered by Health Insurance
Computed from the SIPP
Source C. Nelson and K. Short (1990)

SIPP HEALTH INSURANCE DATA

After a long pilot period dating back to the Income Survey Development Program (ISDP) effort which began in 1975, the Bureau of the Census launched a new longitudinal survey effort in October of 1983. That dataset, the Survey of Income and Program Participation (SIPP), was explicitly designed to provide detailed longitudinal information on the levels and changes in the well being of the populations, especially those sub-populations currently receiving or who might receive federal program assistance. From its inception, the SIPP has included detailed questions on health insurance coverage as potential support for national health insurance or other legislative initiatives to provide health insurance to the currently uncovered.

The basic SIPP design is a rolling panel. Each panel is interviewed eight times at four month intervals. The resulting responses describe 32 consecutive months. A new panel is introduced each year. Thus, in a given calender month there will be two or three different SIPP panels active. Table 3 presents the official SIPP estimates of the levels of health insurance coverage for the entire population; not just the non-elderly (though most individuals over the age of 65 are covered by Medicare). If anything, they show a slight increase in the levels of coverage. This slight increase has prompted some advocates to claim that the SIPP numbers in later months are artificially inflated by attrition. It is claimed that attriters are less likely to be insured, causing rates computed on the basis of those who *are interviewed* to be too high.

Attrition is a constant concern in longitudinal surveys. A recent review of longitudinal data collection efforts listed among the disadvantages of longitudinal data (Subcommittee on Longitudinal Surveys 1986):

> *Beginning refusal rates may be comparable to those of cross-sectional surveys. but the attrition suffered over time may create serious biases in the analysis.*
>
> *Longitudinal surveys are often improperly analyzed, not taking into account longitudinal characteristics or attrition.*

This paper attempts to assess the importance of the first concern by demonstrating one way directly to address the second.

In this paper, we consider the worst case scenario — black males from the 1984 panel. They are the sub-group with the highest attrition. Specifically, our sample consists of the 1196 black males with positive month-one-weight who were between the ages of 18 and 64 at the first interview and were not among the individuals who were cut in the sample reduction. We jointly model their response to the survey (the complement of attrition) and their health insurance coverage status, in the month closest to the interview. Our definition of coverage aggregates private and government coverage. We concentrate on status at the month closest to the interview because of concerns about seam bias. Changes in status tend to occur at between months which are reported on at two different interviews.

Table 4 summarizes the (unweighted interview patterns). Only 65% of this sample completed all eight interviews. Figure 1 summarizes that information graphically. It plots the percentage of people surveyed in each month and the percentage of people who have been present for all interviews up to that one. It depicts a continuous fall in the percentage interviewed from 100% at interview 1, to 84% at interview 2, to 75% at the last interview.

Most of those missing an interview (29 of the 38 percent) are missing more than one interview. The lower line in Figure 1 depicts the percentage of people who have been present for all interviews through that point. By the eighth interview it is ten percentage points below the line for those surveyed at a given interview. The difference is people who miss a least one (almost always exactly one) interview and are found for a later interview. Thus, attrition is a potentially serious problem for this population. If those who are not interviewed are different (in an appropriately defined way) from those who are interviewed then the levels of health insurance coverage computed from those who were interviewed in a given month, or from those who completed all of the interviews will diverge from true coverage rates of the population of interest.

Pattern	Percentage
11111111	62.4
11111110	2.9
11111100	2.9
11111000	3.1
11110000	3.5
11100000	4.0
11000000	6.0
10000000	4.8
Other missing one interview	6.0
Other	4.4

TABLE 4: SIPP Interview Completion Patterns
Black Males in 1984 SIPP Panel

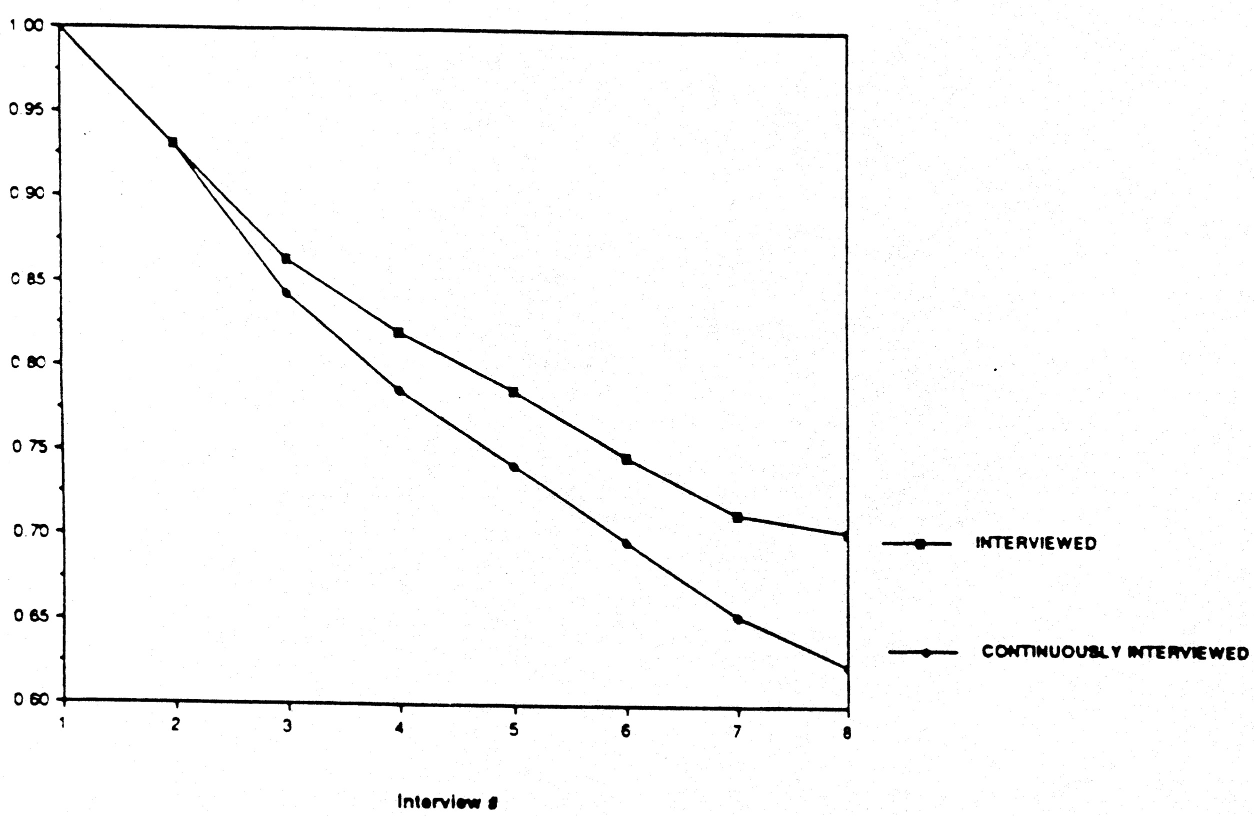

FIGURE 1: Proportion of Individuals Interviewed at Each Wave

This Interview and All Previous Interviews Black Males in 1984 SIPP Panel

Formally, we define the population of interest as those people who were interviewed at the first interview. This paper investigates whether estimates of health insurance coverage computed from those successfully interviewed consistently estimate the true health insurance coverage for the population interviewed in the first month. Failure to obtain an initial interview is a standard coverage problem (in the standard survey sampling use of the term) which we do not deal with here.

By the definition of the population of interest, we observe everyone's health insurance status at the first interview. Thereafter, some people are lost to follow-up. If attrition is random, then in every month including the first, the coverage rates among those who completed all eight interviews should be identical to the coverage rates of those who did not complete all eight interviews.

They are not. The overall coverage rate at the first SIPP interview is 69.0 percent. The coverage rates for those completing all interviews is 73.4 percent. For those with at least one missing interview, the coverage rate is only 62.0 percent. A difference of 11.4 percentage points which is clearly significant and conventional significance levels.

These figures permit some simple calculations of the magnitude of the attrition bias problem. If everyone who misses any interview is dropped from the analysis sample, interview 1 coverage rates would be computed as 73.4 percent rather than the correct figure of 69.0 percent, a difference of 4.4 percentage points. Similarly, if individuals who miss an interview are never reinterviewed, then in a time stationary environment insurance coverage would appear to decline over the 28 months between the first and last interviews from the entire interview 1 sample value of 69.0 percent to the all interview figure of 73.4 This is the same 4.4 percentage point figure, which would now be interpreted as a secular decline in coverage/increase in the levels of uncovered individuals. This is a large decline compared to the secular trends computed from the cross-sectional CPS. Little and Su (1989) call the assumption that anyone who misses one interview misses all subsequent interviews monotonically. As they note, it is not exactly true in most data. Calculations from Table 5 imply that it is correct for about 75 percent of those missing an interview. Furthermore, some analysis strategies require complete data up to the interview, so analysts might impose such a requirement.

This assessment of the situation is overly pessimistic. The requirement of the previous paragraph is equivalent to Little and Rubin's *missing completely at random* concept (as discussed in Little and Su, 1989). A traditional survey analysis approach to attrition is reweighting (Lepkowski, 1989). Attriters differ from non-attriters in terms of demographic characteristics which are incorporated in the survey control totals. Borrowing terminology from Heckman and Robb's (1989) work on estimating program effects, we term this *selection on observables*. In the terminology of Little and Rubin, this would be *missing at random*, but not missing completely at random. SIPP control totals are derived from the CPS and cross-classify by race, sex, and age. Our sample was selected on the first two. Age, however, remains. One possible explanation for the differential health insurance coverage of those present for all interviews and those missing at least one is that the attriters are drawn disproportionately from age groups with lower health insurance coverage. If attrition is random within age groups then reweighting will eliminate the attrition bias.

Specifically, we define the weights as follows.[3] In month 1 each sample individual is assigned a sample weight w_i. Summing these weights over all individuals in age group j, yields, by construction, the control totals for age group j.

$$\sum_{i \in j} w_i = C_j$$

We can compute equivalent weights for the two sub-groups (those who completed all the interviews and those missing some interviews). Define the dummy variable $a_i = 1$, if an individual completed all of the interviews, and $a_i = 0$ otherwise. Then, we define new weights, for the two groups as:

$$\tilde{w}_i = \frac{1}{\sum_{i \in j} a_i w_i} w_i \quad \text{and} \quad \tilde{w}_i = \frac{1}{\sum_{i \in j}(1 - a_i) w_i} w_i$$

[3] This is a generalizations of what Lepkowski (1986, pp. 352-353) calls "One popular method of weighting for nonresponse."

Black Males in 1984 Panel

	Month 1 Coverage all interviews	Month 1 Coverage some missing interviews
Month 1 Weights	73.4	62.0
Reweighted within Each Group	73.2	64.3

The first row uses the original month 1 weights. The second row reweights each column so that it alone reproduces the month 1 control totals.

TABLE 5: Coverage in Period 1 as a Function of Later Interview Status With and Without Reweighting

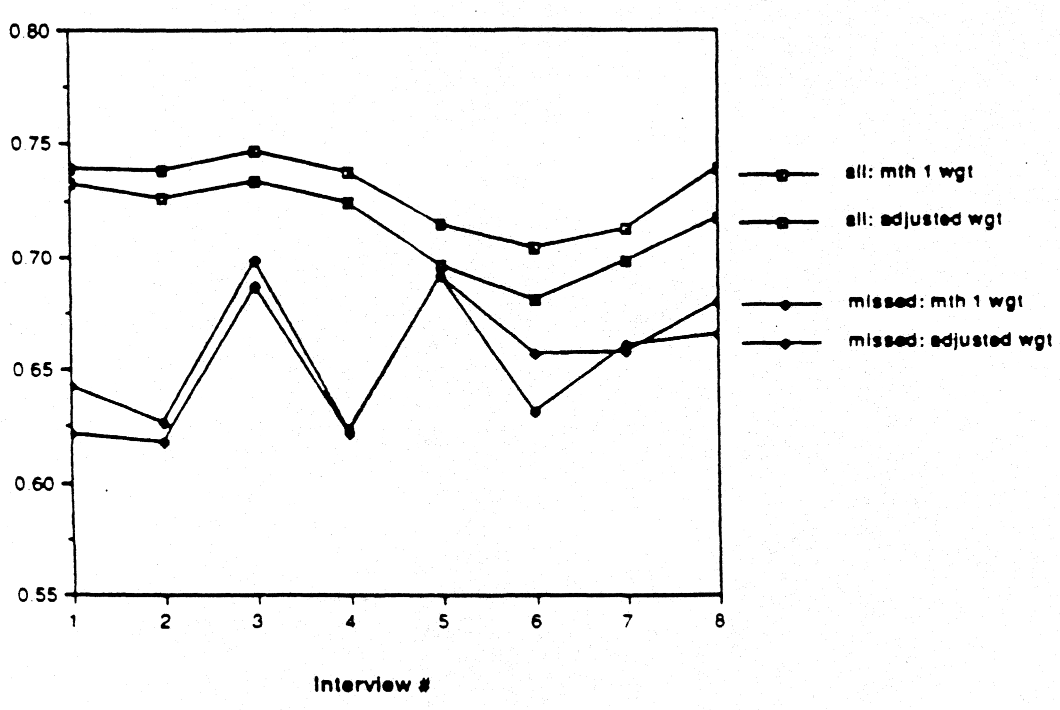

FIGURE 2: Insurance Status at Each Wave by Interviews Completed

All interviews completed. This wave. Missing at least one wave Black Males in 1984 SIPP Panel

With these weights each group separately will sum to the control totals. We can extend this concept to interviews after the first by defining a second dummy variable $d_i(t) = 1$ if an individual completed all of the interviews, and $d_i(t) = 0$ otherwise. Then, we define new weights, for the two groups as:

$$\tilde{w}_{it1} = \frac{1}{\sum\limits_{i \in j} a_i d_i(t) w_i} d_i(t) w_i \quad \text{and} \quad \tilde{w}_{it0} = \frac{1}{\sum\limits_{i \in j} (1 - a_i) d_i(t) w_i} d_i(t) w_i$$

Figure 2 plots the results of such reweighting. The top line and the bottom line are the coverage rates - among those present for each interview (1- 8) - for those present for all interviews and those missing at least one interview respectively. The inner two lines plot the coverage rates reweighting as described above. The figure confirms that attrition is differential with respect to the age groups and that these age groups are correlated with health insurance coverage. Thus, the two reweighted (adjusted) lines are inside the two unadjusted lines; the reweighting explains some of the differential; but not much. At interview one, the unadjusted figures were 73.4 and 62.0. A difference of 11.4 percentage points. The adjusted figures are 73.2 and 64.3. A difference of 8.9 percentage points. Reweighting explains 21.9 percent of the differential health insurance responses between attriters and non-attriters at interview 1. The lines converge for later interviews because individuals who will miss a large number of interviews are no longer interviewed.

The remaining difference is due to differential attrition within the age groups. Implicitly weighting assumes that they are drawn randomly from within the age groups. Evidently attrition is correlated with health insurance coverage status even after controlling for age. Again borrowing terminology from Heckman and Robb (1989), we term this *selection on unobservables*. In the language of Little and Rubin (reviewed in Little and Su, 1989), this attrition is nonignorable.

In the next section, we translate this simple graphical analysis into a formal specification for estimation. Using a conventional joint normal specification, we derive a correlated random effects probit model. The model allows us to test for the presence of attrition bias and to estimate the true time trend for health insurance coverage when attrition bias is present.

A RANDOM EFFECTS MODEL FOR ATTRITION BIAS

We begin this section with some simple parametric data analysis which will form the building blocks for the more complicated joint models of health insurance coverage and attrition which follow. It is simplest to formalize this problem in an index function framework (Heckman 1986). Consider the index functions for interview status, I_i; and for health insurance coverage status, I_h.

$$I_i = X\beta_i + u_i$$
$$I_h = X\beta_h + u_h$$

where the individual is interviewed in a period interview occurs in any period when $I_i > 0$ and the individual is covered by health insurance when $I_h > 0$. Define dummy variables for interview status (where $d_i = 1$ implies that an individual was interviewed, $d_i = 0$ that an individual was not interviewed) and health insurance coverage status (where $d_h = 1$ implies that an individual had coverage, $d_h = 0$ implies no coverage). Assuming that the u's are distributed normally, then we have a standard probit equation:

$$P[d_i = 1] = P[I_i > 0] = \Phi[-X\beta_i]$$

for interview status and:

$$P[d_h = 1] = P[I_h > 0] = \Phi[-X\beta_h]$$

for health insurance coverage, where $\Phi[\,]$ is the standard cumulative normal distribution.

Table 6 presents the results of estimating these probit models on the pooled sample of all individuals at each interval. The sample for interview probit is all interviews after the first . By construction everyone was present for the first interview. The sample for the insurance probit is all completed interviews. We have no information

	Parameter	Std. Error	T-Statistic
Interview			
Constant	1.7462	0.1675	10.4272
Time Trend	-0.1415	0.0077	-18.3109
18-19	-0.3184	0.2237	-1.4233
20-21	-0.3784	0.1842	-2.0550
22-24	-0.3508	0.1793	-1.9568
25-29	-0.3630	0.1763	-2.0593
30-34	-0.2288	0.1765	-1.2964
35-39	-0.2890	0.1814	-1.5926
40-44	0.0467	0.1928	0.2420
45-49	-0.1172	0.1965	-0.5965
50-54	0.0321	0.2015	0.1592
55-59	0.0583	0.1969	0.2958
Health Insurance			
Constant	0.7998	0.1692	4.7257
Time Trend	0.0047	0.0071	0.6596
18-19	-0.5147	0.2085	-2.4684
20-21	-0.6790	0.1833	-3.7038
22-24	-0.6207	0.1820	-3.4110
25-29	-0.3265	0.1819	-1.7946
30-34	-0.2600	0.1808	-1.4383
35-39	0.0434	0.1924	0.2258
40-44	-0.1307	0.1951	-0.6697
45-49	-0.1030	0.2004	-0.5137
50-54	0.1065	0.2055	0.5185
55-59	0.0104	0.1995	0.0523

fval = 0.86317664D+04 nind = 1196 nobs = 9568

TABLE 6: No Permanent Component
Black Males in 1984 Panel

on health insurance status for individuals who were not interviewed. The vector X of covariates for both probit equations consists of a constant, a linear time trend (the interview number 1-8), and dummy variables for the age grouping in the control totals (18-19, 20-21, 22-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, and 55-59). The oldest age category 60-64 is the excluded category.

There is a strong negative time trend in interviews ($t > 18$), later interviews are more likely to be missed (as in Figure 1). The cell sizes are small in the age groups but there is a clear positive age trend. Young black men are less likely to be interviewed. For health insurance there is a small but insignificant positive time trend ($t < 1.0$). These simple probit models imply that health insurance coverage became more common over the panel period. Again, there is a clear positive age trend. Young black me are less likely to be covered by health insurance. None of these results is surprising. Note that the similar patterns of the age effects imply that there will be some selection on observables, young black men are less likely to be interviewed and less likely to be covered. Failure to account for this selection on observables would bias our estimates towards a positive time trend; exactly what we saw in Figure 2.

In fact, the assumption of independence across time periods for a given individual is poor. There is clear evidence of correlation across periods in both interview status and health insurance status. Some form of MANCOVA (multivariate analysis of covariance) is called for. In this binary response context, the simplest specification is normal random effects. We respecify the index functions using a variance components scheme:

$$I_i = X\beta_i + \sigma_i\mu_i + \epsilon_i$$
$$I_h = X\beta_h + \sigma_h\mu_h + \epsilon_h$$

where η, is a time invariant individual specific component component; μ is a factor loading, and ϵ is a time varying (independent across periods) component. We normalize μ and ϵ to have unit variance. By construction these three elements (the X's, the η's, and the ϵ's) are mutually orthogonal.

Table 7 presents the parameter estimates from this model. The variance components are estimated by numerical integration using five point Gaussian Quadrature (Butler and Moffitt, 1986). Estimation requires approximately an hour on a SPARCstation II. Turning first to the variance components each of which have been normalized to have unit variance. In that case the square of the m coefficients is the relative variance of the permanent component. Thus. the point estimates of 1.9 and 1.4 imply that variance of the permanent component is over three times as large as that of the transitory component in the interview equation and nearly twice as large in the health insurance coverage equation. The sharp drop in the likelihood nearly 3,000 points with the addition of only two degrees of freedom is further confirmation of the importance of the individual effects.

The strong negative time trend on interview status remains, as do the age effects. The negative age effects in the health insurance coverage equation remain. The health insurance time trend, which was positive in the uncorrelated model, is now negative, though not significant. The new point estimate is 1.6 standard errors from the point estimate and standard error implied by the old model.

The sign reversal can be understood as follows. The random effects model attributes some of the observed higher rates of coverage in the later period to the fact that the individuals who happen to be interviewed had higher values for their random effects, as revealed by the fact that at their earlier interviews they were more likely than average to be covered by health insurance.

These random effect estimates are probably still not right. They assume that the probability of interview was independent of health insurance status. In that case, we want to weight the individuals with more completed interviews more than those individuals with more interviews. In fact, the optimal weight weights each individual with a full set of interviews by more than his fraction of completed interviews. This occurs because as we accumulate more interviews for each individual we get a narrower posterior on the value of his random effect.

Figure 2 suggests that the assumption of independence between the unobservables *i* interview status and health insurance coverage status is false. If the two unobservables are correlated then the inferences based on that assumption will be incorrect. Formally, the event (the interview is successfully completed or the person

	Parameter	Std. Error	T-Statistic
Interview			
Constant	3.4135	0.2999	11.3825
Time Trend	-0.3328	0.0215	-15.4878
18-19	-0.1363	0.4196	-0.3248
20-21	0.0088	0.3374	0.0262
22-24	0.0869	0.3171	0.2741
25-29	-0.0233	0.2923	-0.0798
30-34	0.1978	0.2865	0.6901
35-39	0.1728	0.2932	0.5893
40-44	0.3428	0.3197	1.0722
45-49	0.2404	0.3129	0.7682
50-54	0.5250	0.2997	1.7514
55-59	0.7772	0.4719	1.6471
Health Insurance			
Constant	1.9235	0.1539	12.4949
Time Trend	-0.0067	0.0131	-0.5122
18-19	-1.5308	0.3079	-4.9723
20-21	-1.7711	0.2748	-6.4449
22-24	-1.6014	0.2589	-6.1861
25-29	-1.0511	0.1956	-5.3749
30-34	-0.9061	0.2186	-4.1455
35-39	-0.4675	0.1867	-2.5042
40-44	-0.6817	0.2804	-2.4312
45-49	-0.4441	0.2099	-2.1157
50-54	-0.2259	0.2418	-0.9339
55-59	-0.3164	0.2000	-1.5815
μ_i	1.8936	0.0616	30.7290
μ_h	1.3533	0.0476	28.4444

fval = 0.56883577D+04 nind = 1196 nobs = 9568

TABLE 7: Independent Permanent Components
Black Males in 1984 Panel

is covered) is observed to occur if $I > 0$. Given this formalism, attrition bias will be present whenever the characteristics of those sample are different from the characteristics of those not sampled. That is, when we compute sample statistics over those who are surveyed at a given interview, or at all of the interview, we are assuming:

$$EP[I_h > 0] = E\{P[I_i > 0]P[I_h > 0]\}/EP[I_i > 0]$$

Attrition bias will not be a problem is the first term is constant. This would require that β in the interview equation be exactly equal to zero, that the η's be uncorrelated and that the ϵ's be uncorrelated. If the β's in the interview equation are not zero, but the other conditions hold, then we have selection, only on observables. In that case, within each homogeneous group of observables, we can estimate the probability of coverage. Reweighting back to the control totals yields consistent estimates of the population coverage probabilities.

Otherwise, to get consistent estimates of the parameters, we need to jointly model the interview and the health insurance coverage processes allowing for correlation of the unobservables. Correlation between the ϵ's is the case discussed by Hausman and Wise (1979). If the time varying components of the two discrete choices are correlated, then some additional information is needed. The standard econometric analysis of the selection bias model implies that we require some variable which affects interview which does not directly affect the behavior of interest (health insurance coverage). No obvious candidate suggests itself and for now we assume away that possibility. We return to it in the conclusion.

In this paper we focus on the third possibility, that the η's are correlated. In our case, that means that the components of the two behaviors which are orthogonal to the observed covariates (age dummies and time trend), but correlated over time are correlated with each other. Put differently, among observationally equivalent people (those in the same age category), those who are more likely to be interviewed are also more likely to be insured.

This is exactly what Figure 2 and Table 5 were demonstrating. In this model, there are two reasons why later interview status could be correlated with interview 1 health insurance coverage status. Either the β's are non-zero, or the η's are correlated. The reweighting addresses the possibility of non-zero β's. That there is still a discrepancy suggests that the h's are correlated.

This formalism can be translated directly into an estimation strategy. We reformulate our index functions as:

$$I_i = X\beta_i + \sigma_i\mu_i + \epsilon_i$$
$$I_h = X\beta_h + \sigma_h\mu_h + \rho\sigma_i\mu_i + \epsilon_h$$

Then, the hypothesis of no correlation of the unobservables is equivalent to the hypothesis of $\rho = 0$. Table 5 and our story that individuals with higher attrition probabilities also have lower health insurance coverage probabilities implies $\rho > 0$ (note that we are modelling interview status which is the complement of attrition).

The final likelihood, is thus a probit for health insurance coverage status in the first period. In each following period, there is a probit for interview status, and a second probit for health insurance status for those who were interviewed. Define the dummy variables:

So the full likelihood is:

$$L_i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{\Phi[X_{h1}\beta_i - \sigma_h\mu_{h1}]\}^{1-d_{h1}}\{1 - \Phi[-X_{h1}\beta_h - \sigma_h\mu_{h1}]\}^{d_{h1}}$$

$$\prod_{t=2}^{8}\{\Phi[-X_{it}\beta_i - \sigma_i\mu_{it}]\}^{1-d_{it}}\{\{1 - \Phi[-X_{it}\beta_i - \sigma_i\mu_{it}]\}\{\Phi[-X_{ht}\beta_i - \sigma_h\mu_{ht}]\}^{1-d_{ht}}$$

$$\{1 - \Phi[-X_{ht}\beta_h - \sigma_h\mu_{ht}]\}^{d_{ht}}\}^{d_{it}}\phi(\mu_i)\phi(\mu_h)d\mu_i d\mu_h$$

These probits are estimated conditional on values of the random effects. The full estimation, jointly estimates the variance and covariance of the random effects along with the vector of coefficients. Numerically, we implement this bi-variate random effects strategy using Gaussian Quadrature (Butler and Moffit, 1986). We use five point quadrature in each of the orthogonal directions (for the two μ terms), yielding 25 different points in the two-dimensional integration space. Thus, we estimate:

$$\sum_{k=1}^{K}\{w_k\Phi[X_{h1}\beta_i - \sigma_h\mu_{h1k}]\}^{1-d_{h1}}\{1-\Phi[-X_{h1}\beta_h - \sigma_h\mu_{h1k}]\}^{d_{h1}}$$

$$\prod_{t=2}^{8}\{\Phi[-X_{it}\beta_i - \sigma_i\mu_{itk}]\}^{1-d_{it}}\{\{1-\Phi[-X_{it}\beta_i - \sigma_i\mu_{itk}]\}\{\Phi[-X_{ht}\beta_i - \sigma_h\mu_{htk}]\}^{1-d_{ht}}$$

$$\{1-\Phi[-X_{ht}\beta_h - \sigma_h\mu_{htk}]\}^{d_{ht}}\}^{d_{it}}$$

where $K = 5$, and the μ's and w's come from a table of Gaussian Quadrature points.

The increase in the log-likelihood between Tables 7 and 8 (from -5688.36 to -5669.68) suggests that there is evidence of correlation between the two permanent components and thus of the presence of attrition bias. Clearly a likelihood ratio test would reject the hypothesis of no correlation (the absolute value of twice the difference in the likelihood is distributed as chi-square with one degree of freedom). The t-statistic of 5.6 on rho implies a similar inference. Thus, there is clear evidence in this sample that people with higher propensities to attrit (conditional on observables), also have a lower propensity to be covered.

Table 8 presents the results of estimation for the model allowing correlation across the two random effects. Most of the parameter estimates are similar to the uncorrelated random effects values from Table 7. The interview time trend continues to be negative, the age trend is positive in both equations, and the permanent components continue to be much more important than the transitory components. The parameter of interest is the attrition bias corrected estimate of the time trend for health insurance. It has fallen again from 0.0047 in the no permanent component model (Table 6), to -0.0067 in the independent permanent components model (Table 7), to -0.0178 in the correlated permanent components specification (Table 8). This value is over three standard deviations away from the point estimate and standard error implied by the simple no permanent effects model which ignored attrition bias and fixed effects. Thus inference based on the simple no permanent effects model, ignoring attrition bias, would yield incorrect conclusions.The absolute value of the implied time trend is however quite small, under two-hundredths of a standard deviation every three months, and none of the estimates are significantly different from zero.

CONCLUSIONS

This paper has proposed a random effects model for estimating the level of health insurance coverage in the presence of differential attrition. The model was estimated on a sample of black males from the 1984 SIPP panel. The estimates suggest some evidence that there is attrition bias, but the effect on the estimates of the time trends in coverage are minimal and the standard errors are large.

Three areas for future research are suggested by these results. First, these results were deliberately obtained on the most attrition prone sub-sample of the SIPP. Extension to the full SIPP sample, with a corresponding increase in computational burden, is an obvious next step. Second, there are several later SIPP panels. Comparing coverage rates of an ongoing panel against those of a newly beginning panel provide an alternative test of the importance of attrition bias. The P-70 report on Health Insurance (Nelson and Short, 1990) provides some simple tests along this line for the full sample. It finds no evidence of attrition bias. The random effects methods proposed here yield mor powerful tests.

Finally, the permanent-transitory scheme used here is clearly inappropriate for both the health insurance and the interview status data. Both of these series are highly serially correlated. It is possible to generalize the models presented here to allow the within period probabilities to follow a Markov process. There is a problem of initial conditions. Although everyone starts out interviewed in period 1, the period 1 health insurance status is random. Steady state approximations can be used to correct for the random initial conditions and their dependence on the permanent components.

	Parameter	Std. Error	T-Statistic
Interview			
Constant	3.3953	0.3023	11.2306
Time Trend	-0.3317	0.0217	-15.2695
18-19	-0.1672	0.4195	-0.3986
20-21	-0.0199	0.3551	-0.0562
22-24	0.0661	0.3299	0.2004
25-29	0.0025	0.2993	0.0083
30-34	0.1775	0.2897	0.6126
35-39	0.1578	0.2971	0.5311
40-44	0.3748	0.3314	1.1310
45-49	0.2743	0.3190	0.8599
50-54	0.5694	0.2897	1.9657
55-59	0.7404	0.5554	1.3331
Health Insurance			
Constant	1.9360	0.2011	9.6283
Time Trend	-0.0178	0.0136	-1.3046
18-19	-1.3839	0.3341	-4.1419
20-21	-1.6637	0.2611	-6.3712
22-24	-1.5149	0.2566	-5.9041
25-29	-1.0454	0.2528	-4.1348
30-34	-0.9084	0.2721	-3.3381
35-39	-0.4169	0.2818	-1.4792
40-44	-0.6794	0.3203	-2.1215
45-49	-0.4436	0.2580	-1.7192
50-54	-0.2096	0.2614	-0.8019
55-59	-0.2789	0.2541	-1.0977
μ_i	1.8888	0.0617	30.6148
μ_h	1.4064	0.0565	24.8706
ρ	0.2309	0.0412	5.5990

fval = 0.56696872D+04 nind = 1196 nobs = 9568

TABLE 8: Correlated Permanent Components
Black Males in 1984 Panel

BIBLIOGRAPHY

BUTLER, J.S. and MOFFITT, Robert (1982), "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica* May, 761-764.

HAUSMAN, Jerry A. and WISE, David A. (1979), "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica*, 47(2):455-473.

HECKMAN, James J. (1976) "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*," 5, Fall, 475-492.

HECKMAN, James J. and ROBB, Richard (1989), "The Value of Longitudinal Data for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in Daniel Kasprzyk, et.al. Panel Surveys, John Wiley and Sons, New York.

LEPKOWSKI, James J. (1989), "Treatment of Wave Nonresponse in Panel Surveys" in Daniel Kasprzyk, et.al. Panel Surveys, John Wiley and Sons, New York.

LILLARD, Lee (1989), "Sample Dynamics: Some Behavioral Issues" in Daniel Kasprzyk, et.al. Panel Surveys, John Wiley and Sons, New York.

LITTLE, Roderick J.A. and SU, Hong-Lin (1989), "Item Nonresponse in Panel Surveys" in Daniel Kasprzyk, et.al. Panel Surveys, John Wiley and Sons, New York.

LONG, Stephen H. (1987), "Public versus Employment-related Health Insurance: Experience and Implications for Black and Non-black Americans," *The Milbank Quarterly*, 65, Suppl. 1, 200-212.

MCARTHUR, Edith (1988), "Measurement of Attrition through the Completed SIPP 1984 Panel: Preliminary Results," letter.

MCNEIL, Jack (1988), "CPS and SIPP Estimates of Health Insurance Coverage Status," Prepared for the United States Department of Commerce, Bureau of the Census, Washington, D.C.

MONHEIT, Alan C. and SHORT, Pamela Farley (1989), "Mandating Health Coverage for Working Americans," *Health Affairs*, (Winter) 22-38.

MOYER, M. Eugene (1989), "A Revised Look at the Number of Uninsured Americans," *Health Affairs*, Summer: 102-110.

NELSON, Charles and SHORT, Kathleen (1990), "Health Insurance Coverage: 1986-1988," U.S. Bureau of the Census, Current Population Reports, Series, P-70, No. 27, U.S. Government Printing Office, Washington, D.C.

SUBCOMMITTEE ON FEDERAL LONGITUDINAL SURVEYS, Federal Committee on Statistical Methodology (1986), "Federal Longitudinal Surveys," Statistical Policy Working Paper 13, Statistical Policy Office, Office of Management and Budget.

SWARTZ, Katherine (1984), "Who has been without Health Insurance? Changes Between 1963 and 1979," Working Paper #3308-01, The Urban Institute, Washington, D.C.

SWARTZ, Katherine and PURCELL, Patrick J. (1989), "Counting Uninsured Americans," *Health Affairs*, Winter: 192-196.