City University of New York (CUNY)

CUNY Academic Works

Publications and Research

Hunter College

1970

Systematic Biases in Panel Surveys Due to Differential Nonresponse.

William (Bill) H. Williams
CUNY Hunter College

Colin L. Mallows PhD

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/hc_pubs/72

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

Systematic Biases in Panel Surveys Due to Differential Nonresponse

W. H. WILLIAMS and C. L. MALLOWS*

Panel surveys involve repeated observations made on the same sample of population units. In some of these studies systematic biases have appeared; specifically, if R_i is an estimate made from panels appearing in the sample for the i^{th} time, then it has been observed that these estimates sometimes vary systematically with i . It is tempting to interpret these phenomena as being due to conditioning of the panel of respondents. In this article it is shown that these systematic changes can also be a result of the characteristics of the response probabilities. This is true whether the estimates are computed only for those individuals who are identical from one interview time to the next, or on all available persons, or on those persons appearing only once.

Numerical examples are presented to show that in some situations, which appear to be innocuous, the potential bias effects of the differential response probabilities can be very large.

1. INTRODUCTION

Panel surveys involve the periodic observation of the *same* population units. In some such studies, systematic biases have appeared which do not seem to have been fully explained. For example, in a Bell System study of customers in midwestern United States [9], those families interviewed for the first time had an average of 3.2 children per family; those interviewed for the second time had 2.5; and those for the third time had 2.4. The average within-group standard error of the estimates was 0.1. Consequently, it appears that the first group is significantly different from the second and third. One of the main purposes of this article is to show that, under certain simple hypotheses, each of several natural estimates made at two different interview times may exhibit systematic changes simply as a result of nonresponse characteristics.

Panel surveys have many variations. Sometimes the same sample is repeated at each of the subsequent observation times. Such samples are referred to as "complete follow-up" surveys. On the other hand, rotation designs involve the month-to-month (say) retention of some sampling units and the replacement of others. Both of these designs can be used to produce estimates for specific points in time as well as estimates of change. The Current Population Survey conducted monthly by the U. S. Bureau of the Census [10] is a good example of a rotation design. In it the unemployment of those persons interviewed for the first time appears to be considerably higher than on subsequent interviews (see [8 and 5]). The details of rotation sampling and panel studies in general will not be described here because there is a large literature on the subject (see, for example, [1, 2, 3, 4, 6, 7]).

Generally, if R_i is the estimate made from those rotation groups which are

* W. H. Williams is a supervisor in the Statistics and Behavioral Science Research Center, and C. L. Mallows is head, Statistical Models and Methods Research Department, both at Bell Telephone Laboratories.

appearing in the sample for the i th time, then it has been sometimes observed that there are significant and systematic changes in these estimates when considered as a function of i . A study of this problem by one of the authors leads to a possible explanation of these phenomena; see [9]. In that article it was shown that if (a) the probability of nonresponse for a selected unit is monotonically related to the characteristic under study, and if (b) this same probability changes from one observation time to the next, then systematic changes in the expectation of the estimator *must* occur.

These two hypotheses are very reasonable ones. It is well known that non-response is sometimes related to the characteristic being measured. For example, families with no children are much more likely to be missed than families with children. There is a large literature on this type of behavior (see, for example, [4]). Also there can be little doubt that the probabilities change from one interview time to the next. This is often clearly revealed by a changing nonresponse rate, which can change *only* if the nonresponse probabilities change and, after all, it is the continuing goal of every survey manager to achieve an ever higher response rate.

The question remains, however, as to whether any of the observed systematic changes could be caused by other factors. For example, is it true that people tend to answer questions about their economic status differently in a second interview from in a first? To examine this question it is tempting to construct estimates which are based only on those individuals who appear in the survey *both* at T_1 and T_2 . The argument is that systematic changes in the estimate for this matched set must be the result of factors other than purely statistical ones. This argument is false, although the conclusion may be partially true in practice. The reason is that systematic changes in the probability of non-response may be ensuring that the matched set of individuals is not representative.

The same difficulty can arise in *any* study in which the same individuals are observed at repeated intervals and in which nonresponse occurs. It can occur, for example, in the complete follow-up surveys so frequently used to examine socioeconomic and medical characteristics. It is true that such studies give a detailed account of the history of the selected individuals, and that changes in the observation from T_1 to T_2 are certainly due to real changes in these individuals or at least to the response given by them. As we shall show, however, it is not necessarily appropriate to estimate changes in the population from the available observations on identical persons.

Also in this article we undertake a comparison of estimates based on identical persons, on all persons, and on those persons interviewed only at one observation period. Since these comparisons appear to be awkward algebraically, some numerical results are included.

Finally a crude "adjusted" estimation scheme is described. It is not recommended unreservedly, however, because it can exhibit poor behavior. The authors hope to report on additional work on the estimation problem in a future article.

Since it was examination of the problem of systematic bias in unemployment statistics that led to the work described here, we have chosen to describe the analysis in terms of employment and unemployment. However, it is important

to notice that none of the subsequent data in this article are real and that the relationship of these models to the real problem has yet to be examined.

2. THE STATISTICAL MODEL

We have chosen to use a simple model for illustration. The increased algebra required for more general cases would detract from the presentation sufficiently, so it seems better to avoid it. Consequently a two-category model, employed and unemployed, is considered.

Suppose that a survey design results in the selection of a certain geographical area and that the employment status of all persons living in that area is to be determined both at time T_1 and again at a later time, T_2 . For simplicity, we neglect the effect of population mobility. Many sample designs will specify a subsampling of persons in the selected area; the remarks we shall make apply to these cases also but for simplicity they are not discussed explicitly. Similarly, the higher structure of the sample design is also ignored with no loss in generality. Given these assumptions, all persons in the area to be sampled can be classified in the manner described in Table 1:

Table 1. NUMBERS OF PERSONS EMPLOYED AND UNEMPLOYED

		<i>Employment status at time T_2</i>	
		<i>Unemployed</i>	<i>Employed</i>
<i>Employment status at time T_1</i>	<i>Unemployed</i>	N_{uu}	N_{ue}
	<i>Employed</i>	N_{eu}	N_{ee}

Using the notation of Table 1, the true unemployment rate at T_1 is given by

$$R_1 = (N_{uu} + N_{ue})/N \quad (2.1)$$

and at T_2 by

$$R_2 = (N_{uu} + N_{eu})/N \quad (2.2)$$

where

$$N = N_{uu} + N_{ue} + N_{eu} + N_{ee}. \quad (2.3)$$

Algebraically, it can be seen that $R_1 = R_2$ if and only if $N_{ue} = N_{eu}$, i.e., when the number of persons who have found employment in the period from T_1 to T_2 is equal to the number of persons who have lost it. Intuitively, this is an obvious condition for the unemployment rate to remain constant.

When the actual sampling of the selected area begins, the classification of the tabulation will not be adequate, since not all persons designated for the survey will be interviewed. Consequently, each individual can be classified as employed, unemployed, or not interviewed at each of T_1 and T_2 , as in Table 2. After the second round of interviews, the nine frequencies in this table will be known. We are assuming here that those persons who are not interviewed at either T_1 or T_2 can still be counted, so that the frequency F_{oo} is known. In some kinds of surveys F_{oo} may remain unknown, or a rough estimate of it may become

Table 2. OBSERVED NUMBERS OF PERSONS IN VARIOUS CATEGORIES

		Status at time T_2		
		Unemployed	Employed	Not interviewed
Status at time T_1	Unemployed	F_{uu}	F_{ue}	F_{uo}
	Employed	F_{eu}	F_{ee}	F_{eo}
	Not interviewed	F_{ou}	F_{oe}	F_{oo}

Table 3. EXPECTED SAMPLE NUMBERS

		Status at time T_2		
		Unemployed	Employed	Not interviewed
Status at time T_1	Unemployed	$N_{uu}P_uP_{uu}$	$N_{ue}P_uP_{ue}$	$N_{uu}P_u(1-P_{uu}) + N_{ue}P_u(1-P_{ue})$
	Employed	$N_{eu}P_eP_{eu}$	$N_{ee}P_eP_{ee}$	$N_{eu}P_e(1-P_{eu}) + N_{ee}P_e(1-P_{ee})$
	Not interviewed	$N_{uu}(1-P_u)Q_{uu}$	$N_{ue}(1-P_u)Q_{ue}$	$N_{uu}(1-P_u)(1-Q_{uu}) + N_{ue}(1-P_u)(1-Q_{ue})$
	interviewed	$+N_{eu}(1-P_e)Q_{eu}$	$+N_{ee}(1-P_e)Q_{ee}$	$+N_{eu}(1-P_e)(1-Q_{eu}) + N_{ee}(1-P_e)(1-Q_{ee})$

available. This has no effect on the subsequent discussion in this article, which is directed towards estimation of the unemployment rates using the other eight frequencies, and which largely ignores sampling variations in the observed frequencies.

To construct an elementary model, let

P_u = the probability that an unemployed person actually appears in the sample at T_1 ;

P_e = the probability that an employed person actually appears in the sample at T_1 ;

P_{uu} = the probability that an individual is interviewed at T_2 given that he was interviewed at T_1 and was unemployed at both T_1 and T_2 ;

P_{ue} = the probability that an individual is interviewed at T_2 given that he was interviewed at T_1 and was unemployed at T_1 and employed at T_2 ;

P_{eu} = the probability that an individual is interviewed at T_2 given that he was interviewed at T_1 and was employed at T_1 and unemployed at T_2 ;

P_{ee} = the probability that an individual is interviewed at T_2 given that he was interviewed at T_1 and was employed at both T_1 and T_2 .

Finally, let Q_{uu} , Q_{ue} , Q_{eu} and Q_{ee} represent probabilities similar to P_{uu} , P_{ue} , P_{eu} and P_{ee} except that the Q 's are conditional on the individual *not* being interviewed at T_1 . For example, Q_{uu} = the probability that an individual is interviewed at T_2 given that he was *not* interviewed at T_1 and was unemployed at both T_1 and T_2 .

Ideally each of these probabilities would equal unity because all persons in the selected area are theoretically to be included in the sample. However, non-response problems will virtually always ensure that these probabilities are *not* unity. Consequently it is interesting to construct a table of expected sample

numbers based on the given three-by-three classification. These expectations are displayed in Table 3.

3. THE STUDY OF IDENTICAL PERSONS

The array of observed sample numbers in Table 2 can be used to construct an estimator based *only* on individuals who are observed at *both* T_1 and T_2 . Using the table as a guide, it can be seen that the number of persons who are unemployed at T_1 and who are interviewed again at T_2 (i.e., in either the employed or unemployed category at T_2) is given by $F_{uu} + F_{ue}$. Similarly, the number of persons employed at T_1 and interviewed at T_2 is given by $F_{eu} + F_{ee}$. Consequently, the unemployment rate at T_1 , based *only* on those individuals who appear both at T_1 and T_2 is given by:

$$\hat{R}_1 = (F_{uu} + F_{ue})/F, \quad (3.1)$$

where

$$F = F_{uu} + F_{ue} + F_{eu} + F_{ee}. \quad (3.2)$$

Similarly, the unemployment rate at time T_2 for this same group of identical individuals is given by,

$$\hat{R}_2 = (F_{uu} + F_{eu})/F, \quad (3.3)$$

so that $\hat{R}_1 = \hat{R}_2$ if and only if

$$F_{ue} = F_{eu}. \quad (3.4)$$

We assume that we are dealing with a large survey, so that under the model of Section 2 and neglecting sampling variability, (3.4) becomes

$$N_{ue}P_uP_{ue} = N_{eu}P_eP_{eu}. \quad (3.4a)$$

In the case in which there is no overall change in employment, (i.e., $N_{ue} = N_{eu}$), this equation can be written as,

$$P_u/P_e = P_{eu}/P_{ue}. \quad (3.5)$$

The result of (3.5) means that there will be a change in the *observed* unemployment rate unless the ratio of the probability of interviewing an unemployed person to the probability of interviewing an employed person, at time T_1 , is the same as the ratio of the corresponding probabilities at time T_2 for persons who were observed at T_1 and whose employment status has changed between T_1 to T_2 . This change in the observed unemployment rate will occur even though there is *no change* in the true unemployment rate.

To explain this in a simpler case, consider the model in which the probabilities at T_2 are independent of the status at T_1 and of whether the individual is observed at T_1 , as follows: $P_{uu} = P_{eu} = P_{2u}$, $P_{ue} = P_{ee} = P_{2e}$. Also we now write $P_u = P_{1u}$ and $P_e = P_{1e}$. Then (3.4a) can be written as,

$$N_{ue}P_{1u}P_{2e} = N_{eu}P_{1e}P_{2u}, \quad (3.6)$$

and (3.5) as,

$$P_{1u}/P_{1e} = P_{2u}/P_{2e}. \quad (3.7)$$

The interpretation of (3.7) is that the ratio of the probability of interviewing an unemployed person to the probability of interviewing an employed person must be the same at T_1 and T_2 ; otherwise, there will be a change in the expected unemployment rate from T_1 to T_2 . This is true even though there has been no change in the true unemployment rate and the estimates are based on identical individuals. Is (3.7) likely to hold in practice? It has been reported by Williams [9] that some practitioners feel that $P_{1u} > P_{1e}$. If this is so and if a survey manager subsequently reduces the nonresponse rate so effectively that P_{2u} and P_{2e} are approximately unity, then the difference $\hat{R}_e - \hat{R}_2$ must be positive, and a systematic change in the estimator based on the identical persons will be observed. A similar result was also found by Williams [9] for estimators based on the complete sample at each of T_1 and T_2 .

There appears to be little hard information on the magnitude of the P 's and Q 's. Some information can be obtained from survey managers and people closely involved with survey methodology. It was pointed out in the previous paragraph that some practitioners feel that $P_{1u} > P_{1e}$. Many of these same practitioners also feel that at some later go-arounds this inequality probably reverses to $P_{2u} < P_{2e}$. The argument for this is based on possible existence of a hard core of unobservable unemployed persons. Other arguments can be created to justify other relationships among the P 's and Q 's. Unfortunately, there appears to be little information on which to distinguish among the theoretical proposals.

4. COMPARISON OF THE ESTIMATES BASED ON IDENTICAL, UNMATCHED, AND ALL PERSONS

In Section 3 an analysis was made of the estimates based only on those persons appearing in the sample both at T_1 and T_2 . Altogether three pairs of estimates of R_1 and R_2 might be in hand; one based on identical persons, one on "single" persons (who appear only once, either at T_1 or at T_2) and one on all persons. Consequently, it is interesting to ask whether all three of these estimates must necessarily exhibit a systematic change? Is it possible for one estimate of the difference, $R_1 - R_2$, to behave differently from the others? And what are the magnitudes of the possible biases in the various estimates of $R_1 - R_2$?

The estimates based on identical individuals have already been described in Section 3. They were formed by simply picking out the appropriate sample numbers from Table 2. The estimate based on "single" individuals and the estimate based on all available persons can easily be formed in the same way. For example, the estimates $\hat{R}_{1t}$, $\hat{R}_{2t}$ based on the total number of persons, and their values under our model in the simple case of independence, are given by

$$\frac{\hat{R}_{1t}}{1 - \hat{R}_{1t}} = \frac{F_{uu} + F_{ue} + F_{uo}}{F_{eu} + F_{ee} + F_{eo}} \approx \frac{N_{uu} + N_{ue} \cdot P_{1u}}{N_{eu} + N_{ee} \cdot P_{1e}}, \quad (4.1)$$

and

$$\frac{\hat{R}_{2t}}{1 - \hat{R}_{2t}} = \frac{F_{uu} + F_{eu} + F_{ou}}{F_{ue} + F_{ee} + F_{oe}} \approx \frac{N_{uu} + N_{eu} \cdot P_{2u}}{N_{eu} + N_{ee} \cdot P_{2e}}. \quad (4.2)$$

The estimates $\hat{R}_{1s}$, $\hat{R}_{2s}$ based on the individuals who appear only on a single occasion are given (in the simple case of independence) by

$$\frac{\hat{R}_{1s}}{1 - R_{1s}} = \frac{F_{uo}}{F_{eo}} \approx \frac{N_{uu}(1 - P_{2u}) + N_{ue}(1 - P_{2e})}{N_{eu}(1 - P_{2u}) + N_{ee}(1 - P_{2e})} \cdot \frac{P_{1u}}{P_{1e}}, \quad (4.3)$$

and

$$\frac{\hat{R}_{2s}}{1 - \hat{R}_{2s}} = \frac{F_{ou}}{F_{oe}} \approx \frac{N_{uu}(1 - P_{1u}) + N_{eu}(1 - P_{1e})}{N_{ue}(1 - P_{1u}) + N_{ee}(1 - P_{1e})} \cdot \frac{P_{2u}}{P_{2e}}. \quad (4.4)$$

Unfortunately, the differences $\hat{R}_1 - \hat{R}_2$, $\hat{R}_{1t} - \hat{R}_{2t}$, and $\hat{R}_{1s} - \hat{R}_{2s}$ do not appear to have algebraic forms which facilitate easy comparisons. Consequently, a numerical study was undertaken. A large number of cases were computed; three are presented in the appendix. Each of the examples given there is arrayed in the same way.

Examination of the numerical results enables one to make certain comments:

1. Substantial biases appear with apparently innocuous probability differences and with very low nonresponse rates. In Example A, the response rate at T_1 is over 89 percent and at T_2 it is well over 95 percent. These are response rates which are characteristic of some of the surveys run by the U.S. Bureau of the Census but are not matched consistently by any survey group. Nevertheless, the estimates based on all sampled persons suggest a change of about 14 percent in the true unemployment rate. Keep in mind that $N_{ue} = N_{eu}$, so that there is *no real* change in unemployment, and also that sampling variability is being ignored. Even the estimate based on an identical set of individuals has a bias of about five percent. The estimates of R_1 , R_2 and $R_1 - R_2$ based on the "singles" are so bad that it is disturbing to consider the possibility of their use.

Example B is included to further illustrate the major effect that these probability differences can have on the estimates. The only difference between Example A and Example B is that the second-stage response probabilities for unemployed persons, i.e., P_{uu} , Q_{uu} , P_{eu} , Q_{eu} have been dropped substantially. It is tempting to react to this case by taking the attitude that these probabilities are unrealistically low. But are they? And how would one know it, because the response rate at T_2 is higher than at T_1 ; and it has dropped only one and one-half percent from that in Example A and in general is still a very high response rate. The important point, however, is that with almost no warning the biases in R_1 , R_2 and $R_1 - R_2$ have gone from bad, in Example A, to disastrous in Example B. The "total" change estimate now has a bias of over 55 percent of the true unemployment rate. The estimates of R_1 and R_2 based on the identicals are clearly very bad, but in addition the "identical" estimate of $R_1 - R_2$ is now about three times as big as it was in Example A.

2. Examples A and B have the same amount of shifting from one category to the other, i.e., $N_{ue} = N_{eu} = 100$ in both cases. The amount of shifting does affect the biases. For example, suppose the probabilities of Case A are used with the following populations:

	N_{uu}	N_{ue}	N_{eu}	N_{ee}
<i>i</i>	400	00	00	9600
<i>ii</i>	300	100	100	9500
<i>iii</i>	200	200	200	9400
<i>iv</i>	100	300	300	9300
<i>v</i>	0	400	400	9200

These five examples are constructed so that the true unemployment rate remains constant at 4.0 percent. It will be found that the biases in both the "total" and "identical" estimates get worse as the shifting around increases. On the other hand, the singles estimates get better. The latter are still not satisfactory but at least do not exhibit the extremely bad behavior of Examples A and B.

3. Calculation of Case *v* in Item 2 reveals the distressing fact that the three estimates of the change $R_1 - R_2$ from identicals, total, and singles do not even necessarily have the same sign.

4. The Q probabilities do not seem to play a major role in the biases. These probabilities cannot affect the "identical" estimates at all and have only a small effect on the bias in the "singles" estimate. The estimates based on total persons are affected even more slightly because these are a combination of the single and the matched.

5. An examination of the tables of expected numbers reveals characteristics which are likely to be quite perplexing. For example, the number of interviewed, employed persons (at T_2), who were not interviewed at T_1 may be two or three times the number of persons who were employed at T_1 and not interviewed at T_2 .

Specifically, in Case A, the number of persons interviewed at T_2 and found to be employed but who were *not* interviewed at T_1 is 1010. This means that 1010 employed persons who were not interviewed at T_1 were found in the sample area at T_2 . On the other hand, only 352 employed persons were interviewed at T_1 and lost to the sample at T_2 . Thus it appears that one is "finding" far more employed persons at T_2 than were lost at T_1 . This result might be interpreted as showing mobility of the population, so that unemployed persons move and show up somewhere else as employed persons. Example A shows that this mobility feature may well be only a reflection of the response probability structure. This is not to say that the real population does not have such a characteristic; this example shows only that it will have to be demonstrated with other evidence.

6. The examples presented in the statistical appendix have what must be considered high expected response rates. If the probabilities are dropped so that the response rates are in the vicinity of 60 to 70 percent, the biases become far worse than those shown in Cases A and B. Unfortunately, in practice, more surveys seem to operate in this range than with response rates of 90 percent.

7. Example C is constructed so that there is an even split of the population between the two categories. Again $N_{ue} = N_{eu}$ so that there is no change in the ratio in the two categories from T_1 to T_2 . Since unemployment rates of this magnitude occur very seldom, it may be more helpful to discuss this example

in terms of persons who say that they are going to vote for Candidate U or for Candidate E . The probabilities are numerically the same as those used in Example A. Now, however, they are interpreted as indicating that the voters for one candidate are more easily interviewed than those of the other. Both at T_1 and T_2 there is a response rate of approximately 91 percent which again can be considered to be high. However, the three estimates of $R_1 - R_2$, "identicals," "total" and "singles" give the appearance of a major swing to Candidate E . Just how big a swing is determined by which of the three estimates is being considered. Unfortunately, in reality there is no swing to either of the candidates.

8. In practice the eight numbers (apart from the marginal totals) given in the first table of each of the examples are available for estimation purposes. Unfortunately our model has 14 parameters, six P 's, four Q 's and four N 's. In this situation, there are several simplifying assumptions that can be made to reduce the 14 parameters to eight. One has been used here. It consists of assuming that the four Q 's are equal to the corresponding P 's. In addition, the assumptions $P_{uu} = P_{eu} = P_{2u}$ and $P_{ee} = P_{ue} = P_{2e}$ reduce the number of parameters to eight, specifically, two first stage P 's, two second stage P 's and four N 's. Now the expressions for the expected cell numbers can be equated to the realized numbers and the resultant equations solved for the unknown eight parameters. This crude "adjusted" estimation scheme was used in all the examples shown in the appendix. While it behaves better than the other estimators in all of the present cases, and indeed did so in many examples which are not included in this article, we do not wish to recommend its general use pending further investigation which we hope to report on in a future article.

Numerical studies of some examples of "complete follow-up" surveys showed that, even in these cases, the estimates can show substantial biases.

5. SUMMARY AND GENERALIZATIONS

The important point made in this article is that systematic changes will occur in estimates from one stage of a survey to the next if the probabilities of non-response are not the same for all of the population categories. Williams [9] showed that this was possible for the overall sample, and in this article the same systematic biases have been shown to be possible even if one computes estimates based on an identical set of individuals.

This article also has mentioned complete follow-up surveys and a simple "adjusted" estimation scheme. Some numerical examples have been included which show the extreme biases that can easily be encountered in the estimates of change from one observation period to the next. This is true whether the estimate of change is based on identical persons, on all persons or on those persons appearing only for a single interview.

Distortions caused by the incomplete responses can mislead a researcher. For example, it is pointed out that a "mobility" characteristic could well be a result of these nonresponse problems.

The models presented here are very simple. They can be generalized easily to more categories and to additional reinterviews. With some increase in complexity the case of a continuous variate y_i can also be covered. We have not

attempted extensive generalization here. The list of cases that might be worth detailing is as long as the number of different schemes used in practice. It seems important simply to point out that these awkward biases can and probably do occur in practice.

APPENDIX

Example A

First Stage Response Probabilities:

$$P_u = 0.93 \quad P_e = 0.89$$

Second Stage Response Probabilities:

$$P_{uu} = 0.87 \quad P_{ue} = 0.96 \quad P_{eu} = 0.84 \quad P_{ee} = 0.96$$

$$Q_{uu} = 0.87 \quad Q_{ue} = 0.96 \quad Q_{eu} = 0.84 \quad Q_{ee} = 0.96$$

True Population Figures:

$$N_{uu} = 300 \quad N_{ue} = 100 \quad N_{eu} = 100 \quad N_{ee} = 9500$$

Table A1. EXPECTED SAMPLE NUMBERS

	<i>U2</i>	<i>E2</i>	<i>N2</i>	<i>Total</i>
<i>U1</i>	243	89	40	372
<i>E1</i>	75	8117	352	8544
<i>N1</i>	28	1010		
<i>Total</i>	345	9216		

NOTE: $E(n_1) = 8916$
 $E(n_2) = 9561$

Table A2. ESTIMATES OF UNEMPLOYMENT RATES (PERCENT)

<i>R</i>	<i>True</i>	<i>Identicals</i>	<i>Total</i>	<i>Singles</i>	<i>Adjusted</i>
<i>R1</i>	4.00	3.90	4.17	10.19	3.84
<i>R2</i>	4.00	3.72	3.61	2.65	3.83
<i>R1 - R2</i>	0.00	0.17	0.56	7.54	0.01

Example B

First State Response Probabilities:

$$P_u = 0.93 \quad P_e = 0.89$$

Second Stage Response Probabilities:

$$P_{uu} = 0.50 \quad P_{ue} = 0.96 \quad P_{eu} = 0.50 \quad P_{ee} = 0.96$$

$$Q_{uu} = 0.40 \quad Q_{ue} = 0.96 \quad Q_{eu} = 0.40 \quad Q_{ee} = 0.96$$

True Population Figures:

$$N_{uu} = 300 \quad N_{ue} = 100 \quad N_{eu} = 100 \quad N_{ee} = 9500$$

Table B1. EXPECTED SAMPLE NUMBERS

	<i>U2</i>	<i>E2</i>	<i>N2</i>	<i>Total</i>
<i>U1</i>	139	89	143	372
<i>E1</i>	44	8117	383	8544
<i>N1</i>	13	1010		
<i>Total</i>	197	9216		

NOTE: $E(n_1) = 8916$
 $E(n_2) = 9413$

Table B2. ESTIMATES OF UNEMPLOYMENT RATES (PERCENT)

<i>R</i>	<i>True</i>	<i>Identicals</i>	<i>Total</i>	<i>Singles</i>	<i>Adjusted</i>
<i>R1</i>	4.00	2.73	4.17	27.23	3.76
<i>R2</i>	4.00	2.19	2.09	1.25	3.86
<i>R1 - R2</i>	0.00	0.53	2.08	25.98	-0.10

Example C

First Stage Response Probabilities:

$$P_u = 0.93 \quad P_e = 0.89$$

Second Stage Response Probabilities:

$$P_{uu} = 0.87 \quad P_{ue} = 0.96 \quad P_{eu} = 0.84 \quad P_{ee} = 0.96$$

$$Q_{uu} = 0.87 \quad Q_{ue} = 0.96 \quad Q_{eu} = 0.84 \quad Q_{ee} = 0.96$$

True Population Figures:

$$N_{uu} = 4000 \quad N_{ue} = 1000 \quad N_{eu} = 1000 \quad N_{ee} = 4000$$

Table C1. EXPECTED SAMPLE NUMBERS

	<i>U2</i>	<i>E2</i>	<i>N2</i>	<i>Total</i>
<i>U1</i>	3236	893	521	4650
<i>E1</i>	748	3418	285	4450
<i>N1</i>	336	490		
<i>Total</i>	4320	4800		

NOTE: $E(n_1) = 9100$
 $E(n_2) = 9120$

Table C2. ESTIMATES OF UNEMPLOYMENT RATES (PERCENT)

<i>R</i>	<i>True</i>	<i>Identicals</i>	<i>Total</i>	<i>Singles</i>	<i>Adjusted</i>
<i>R1</i>	50.00	49.78	51.10	64.65	50.51
<i>R2</i>	50.00	48.03	47.37	40.70	49.75
<i>R1 - R2</i>	00.00	1.75	3.73	23.95	0.76

REFERENCES

- [1] Cochran, W. G., *Sampling Techniques*, 2nd ed., New York: John Wiley and Sons, Inc., 1963.
- [2] Eckler, A. R., "Rotation Sampling," *The Annals of Mathematical Statistics*, 26 (1955), 664-85.
- [3] Hansen, M. H., Hurwitz, W. N. and Madow, W. G., *Sample Survey Methods and Theory*, New York: John Wiley and Sons, Inc., Vols. 1 and 2, 1953.
- [4] Kish, Leslie, *Survey Sampling*, New York: John Wiley and Sons, Inc., 1965.
- [5] "Measuring Employment and Unemployment," Report of the President's Committee to Appraise Employment and Unemployment Statistics, Washington, D.C.: The White House, 1962.
- [6] Patterson, H. D., "Sampling on Successive Occasions with Partial Replacement of Units," *Journal Royal Statistical Society, B*, 12 (1950), 241-255.
- [7] Rao, J. N. K., and Graham, Jack E., "Rotation Designs for Sampling on Repeated Occasions," *Journal American Statistical Association* 59 (1964), 495-509.
- [8] Waksberg, Joseph, and Pearl, Robert B., "The Effects of Repeated Interviews in the Current Population Survey," Paper presented at 47th National Conference of the American Marketing Association, Dallas, Texas, 1964.
- [9] Williams, W. H., "The Systematic Bias Effects of Incomplete Responses in Rotation Samples," *Public Opinion Quarterly*, 32, No. 4 (1970), 593-602.
- [10] U. S. Bureau of the Census, "The Current Population Survey—A Report on Methodology," Technical Paper No. 7, Washington, D. C.: U. S. Government Printing Office, 1963.