

THE SYSTEMATIC BIAS EFFECTS OF INCOMPLETE RESPONSES

W. H. Williams, Bell Telephone Laboratories and U.S. Bureau of the Census

I. INTRODUCTION

Rotation sampling schemes are used for continuing studies in which there is interest in the estimation of change from month-to-month (say) as well as in separate estimates for the individual months. These rotation designs involve the month-to-month retention of some sampling units and the replacement of others. The details of rotation sampling will not be described in this paper because there is a large literature on the subject, see for example, Hansen, Hurwitz and Madow[1], Cochran[2], Patterson[3], Eckler[4], Rao and Graham[5], and Kish[8]. It is to the point, however, to describe some applications.

A study of Bell System customers in western United States used a monthly sample consisting of three separate rotation groups. Each month one group appeared in sample for the first time, another for the second time, and the third had been in the two previous months. After three months in the sample each rotation group was dropped and did not reappear. The duration of study was eighteen months.

In the Current Population Survey (CPS), [6], conducted monthly by the U. S. Bureau of the Census one-eighth of the sample is new each month. Each new one-eighth group is retained in the sample for four consecutive months. It is then dropped for the next eight months, after which it is brought back into the sample again for four consecutive months. In this way each rotation group appears in the sample for a total of eight months.

In rotation group studies systematic biases have been observed in practice and the following examples appear to be typical.

In the Bell System study the average number of children per family for rotation groups appearing in the sample for the first time was 3.2. For rotation groups appearing in the sample for the second and third times the averages were 2.5 and 2.4 respectively. The average within rotation group variance of the monthly estimates was 0.1. Consequently, it appears that the first month may be significantly different from the second and third. How does one explain this apparent falling off in the number of children per household? Is it a systematic bias introduced by the interviewer or respondent? Or can a characteristic of the survey design or its implementation be responsible?

A similar characteristic appears in the CPS survey. Table 1 shows unemployment versus number of times in the survey for the CPS study. The data are taken from Waksberg and Pearl[15]. Unemployment appears to be higher for units which appear in the sample for the first and fifth times. [Recall that there is an eight-month lapse between the fourth and fifth interviews.] Why do these two peaks appear?*

Does the interviewer influence the respondent in such a way that he gives different responses from one month to the next? Such an hypothesis may be acceptable for the unemployment estimates but seems less likely for the number of children in the Bell System study. Similar behaviour exists for other characteristics in the CPS study, for example for estimated vacancy rates and families with salaries over \$15,000, see Waksberg and Pearl[15].

Appearance in Sample								
	1	2	3	4	5	6	7	8
In-dex	107.3	100.3	100.3	98.9	100.7	99.6	96.6	95.0

Table 1. CPS TOTAL UNEMPLOYMENT 1955-61
[INDEX NUMBERS, ALL GROUPS COMBINED EQUALS 100]

Before leaving this description of the problem, it is relevant to introduce "one-time" surveys which involve call-backs. These will be compared with rotation samples in the next section but it is to the point to present some data from one now. The data were taken from the paper by Finkner[7] and are presented in Table 2 below. Actually, the Finkner study was a multiple mail survey, but it is interesting because a systematic behaviour similar to the rotation group bias appears. Experienced practitioners will of course recognize that this behaviour is common in call-back and mail surveys. It will be pointed out in this paper that this behaviour and the rotation group bias can have a similar cause.

	Number of Growers	% of pop ⁿ	Avg. # of Fruit Trees per Grower
Response to first mailing	300	10	456
Response to 2nd mailing	543	17	382
Response to 3rd mailing	434	14	340
Nonrespondents after 3rd mailing	1839	59	290
Total population	3116	100	329

Table 2. FINKNER DATA - MULTIPLE MAIL SURVEY

II. INCOMPLETE SAMPLES

Population surveys are frequently conducted in such a way that all of the persons in a randomly selected area are to be included in the sample; other schemes will specify a subsampling of these persons, say by selecting every k^{th} household on a block. The remarks to be made in this paper apply to both cases, but to simplify the

discussion and the formulas, it is assumed that all persons in the selected area are to be drawn into the sample. For the same reason, the higher structure of the sampling design is ignored. No loss in generality will result.

The number of persons in the selected area is denoted N , which may be known or unknown in practice, but seems more often to be unknown. The sampling scheme specifies that N persons are to be interviewed at some point in time, but in practice they rarely all are. To be specific, the objective of the survey is to interview N individuals in an area in such a way that the probability of inclusion, p_i , equals one, $i = 1, 2, \dots, N$. In practice, however, these probabilities may well be less than one with the result that a sample of $n < N$ persons is obtained.

The expected number of persons is $\sum_{i=1}^N p_i$ which equals N if all $p_i = 1$ and is less than N otherwise.

It was stated earlier that the survey which uses call-backs to obtain estimates at a single point in time has characteristics similar to rotation sampling. These can be seen by looking at the first visit as the first appearance in the rotation sample. The second visit (first call-back) is the same as the second appearance in the rotation sample if those persons interviewed at the first visit are considered to be included at the second visit with probability one. [They are not actually visited twice but the data obtained are simply carried over.] A difference is that call-back surveys use the assumption that the characteristics under observation do not change with time, while rotation samples are designed to estimate this change.

In both call-back and rotation surveys, estimation difficulties arise because the probabilities with which a response is obtained are unknown. Estimation is usually carried out by assuming that these response probabilities are equal. What are the effects of this practice? For call-back sampling, the problem has long been recognized and papers have appeared on the subject. It seems unnecessary to trace these in this paper except to point out that a good description of the work has been given by Kish[8], pp. 532-62. The papers by Politz and Simmons[10],[11], and Hartley[12] are relevant to the work in this paper, in that an attempt is made to estimate the individual response probability.

In rotation sampling the effects of these unknown probabilities do not seem to have been discussed. An additional difficulty is that these probabilities are undoubtedly changing from one appearance in the survey to the next, and probably are doing it in a systematic way. This problem is discussed in the next section in such a way that the results are applicable to any design which involves periodic re-interviews.

III. THE EFFECTS OF THE UNKNOWN PROBABILITIES

3.1 At the First Appearance

Suppose that the N units in the sampled area have characteristics y_i , $i = 1, 2, \dots, N$. The objective is that responses be obtained from each of them with probability $p_i = 1$, $i = 1, 2, \dots, N$. However, as pointed out earlier the interviewing method is not likely to be that successful and $p_i = 1$ will not be achieved for all i units. Then the expected sample size (number of responses) is $n_1^* = E(n_1) = \sum_{i=1}^N p_i$, where n_1 is the number of interviews actually obtained. Next, an estimate

of the mean is formed as $\hat{Y}_1 = \left(\sum_{i=1}^{n_1} y_i \right) / n_1$, which is a ratio estimator with expectation,

$E(\hat{Y}_1) = \left(\sum_{i=1}^N p_i y_i \right) / \left(\sum_{i=1}^N p_i \right)$. This expectation is approximate but the technical bias of the ratio estimator is not important here.

The incomplete response has effectively introduced an additional level of sampling into the overall design. The effect on total variance is probably not large because this additional component of variance comes in at the lowest level in the sampling design. The bias effects may be quite another matter however, since the probabilities of inclusion at the last stage are unknown and may very well have a systematic behaviour.

3.2 Rotation Sampling and Call-Backs

The second time the selected persons are to be interviewed there can be little doubt that the probabilities of actual inclusion will have changed from the first interview. There are a number of reasons for this. One is that it would be expected that the information gained at the time of the first interview period, (T_1), would increase the probability of a response at the second, (T_2). The interview team probably knows the area and the availability characteristics of some of the individuals better at T_2 than at T_1 . Consequently, a survey manager would naturally expect that the number of responses obtained would tend to go up at T_2 . It seems unlikely, however, that every unit will have a larger probability at T_2 , some could conceivably decline. The number of refusals for example typically increases the longer a group has been in the sample. Specifically, the units will have probabilities p_i' associated with them at T_2 and many of these will be different from the p_i at T_1 . In rotation sampling it will also be expected that some of the characteristics y_i , $i = 1, 2, \dots, N$ will have changed. One of the purposes of rotation sampling is to obtain efficient estimates of this change. However, in this paper we wish to study the possible effects of the changes in probabilities and so to insure that there are no confounded factors, it is assumed that the y_i do not change from T_1 to T_2 .

Given this hypothesis, rotation sampling and callback surveys are very similar.

Consequently, with the above assumptions,

$n_2^* = E(n_2) = \sum_{i=1}^N p_i''$, is the expected sample size,

and the

estimator, $\hat{Y}_2 = \left(\sum_{i=1}^{n_2} y_i \right) / n_2$, has the approximate

expectation $E(\hat{Y}_2) = \left(\sum_{i=1}^N p_i'' y_i \right) / \left(\sum_{i=1}^N p_i'' \right)$.

3.3 The Special Case of Proportions

A case of special interest is that in which there are two classifications such as employed and unemployed. It should be emphasized that these two categories are referred to as employed and unemployed because this work was originally suggested by consideration of the characteristics of unemployment statistics. The extent to which these models actually apply to unemployment statistics has not yet been determined. Then if (i) p_e' denotes the probability of an employed person actually being interviewed at T_1 , and (ii) p_u' denotes the analogous probability for an unemployed person, and (iii) y_i equals one if unemployed and equals zero if employed, then

$\hat{R}_{ul} = n_{ul} / (n_{el} + n_{ul})$ and $E(\hat{R}_{ul}) = N p_u' / (N p_e' + N p_u')$. Similar expressions can be written

down for the unemployment rates at T_2 . The generalization to more categories presents no difficulties.

3.4 The Bias Effects of the Unknown and Changing Probabilities

Under the assumption of no changes in the characteristic y_i it would be hoped that the expectations of \hat{Y}_1 and \hat{Y}_2 would be the same and equal to \bar{Y} the population mean. Is this true? And if not, what statements can be made?

The technical question being asked is how does

$\left(\sum p_i'' y_i \right) / \left(\sum p_i'' \right)$ compare with $\left(\sum p_i' y_i \right) / \left(\sum p_i' \right)$? To this end the following points can be easily made.

(1) If $p_i' = k p_i''$, the expectations at T_1 and T_2 are the same.

(2) If the p_i' 's are randomly associated with the y_i 's, the expectation at T_1 is equal to \bar{Y} . Similarly, if the p_i'' 's are randomly associated with the y_i 's the expectation at T_2 is equal to \bar{Y} . Consequently there is no bias at T_1 or T_2 and no systematic change from T_1 to T_2 .

(3) What happens under the more realistic assumption that the probabilities at T_1 are related to the characteristic y_i and that more information and experience on the part of the interviewers at T_2 brings these probabilities closer to equality and to one? To answer this,

suppose that $p_i = k y_i^\alpha$, and that all the y_i 's are positive. Then it can be easily shown that the estimator $\sum p_i y_i / \sum p_i$ increases monotonically with α .

(i) As a first example suppose that $p_i' \sim y_i$ and $p_i'' = 1.0$. This means that at T_1 the units with larger y values have a higher probability of entering the sample and that at T_2 all units enter the sample. The latter choice of all $p_i = 1$ is the survey manager's idealized goal and would be a result of an efficient interview program at T_2 . Since $p_i' \sim y_i$ at T_1 corresponds to $\alpha = 1$ and $p_i'' = 1$ at T_2 to $\alpha = 0$, it follows from above that $E\hat{Y}_2 \leq E\hat{Y}_1$, the equality occurring if all y_i 's are equal. It is important to notice that this systematic change comes about solely as a result of changes in the probabilities and will occur even though there has been no change in the characteristic being measured.

(ii) As a second example, suppose that $p_i' \sim 1/y_i$ and $p_i'' = 1$, so that the larger units have a smaller chance of appearing in the sample at T_1 . Then $E\hat{Y}_2 \geq E\hat{Y}_1$ and a systematic change appears in the opposite direction. This again is solely a result of changing probabilities because the y characteristics have been assumed to be constant in the time period from T_1 to T_2 .

What can be said about the specific case of unemployment? First, it can be easily shown that

$E(R_u) \leq \text{True Rate}$, $\Leftrightarrow p_u' \leq p_e$, which is an obvious intuitive result. Secondly, if the probabilities for employment and unemployment each change in different proportions from T_1 to T_2 , as follows, $p_e'' = k_1 p_e'$, $p_u'' = k_2 p_u' = c k_1 p_u'$, then it can be easily shown that,

$E(R_{u2}) \leq E(R_{u1}) \Leftrightarrow c \leq 1$. For example, if $c < 1$, $k_2 < k_1$, and $E(R_{u2}) < E(R_{u1})$. This means

that if the biggest change in probability from T_1 to T_2 is associated with employed persons then you must get a decrease in the expected value of the estimator solely as a result of this change. In this situation, it would seem likely that at T_1 , $p_u > p_e$, which in fact concurs with the field experience of Deming[13], Harris[14], and Kish[8].

In this case it is interesting to look at some numerical results. Suppose that $N = 10,000$, $N_u = 400$, so that $R_u = 0.04$, then simple calculations yield the figures in Table 3. If case i represents the situation at T_1 , and an effort is made at T_2 to improve the response so that case ii describes the resultant situation then we see that there has been a five percent change in the expectation of the estimate with no change in actual unemployment and in spite of a high response rate. Case iii simply shows that without knowledge of the p 's there is no way of knowing whether R_u is being over or underestimated. Case iv shows that a three percent bias is possible with probability differences which intuitively one would probably judge to be very small.

Cases iv and v are interesting to consider together. If at T_1 , (case iv), p_u is slightly higher than p_e (as indicated), and if as a result of any "unobservable" characteristic of

unemployed persons p_u drops, then a comparison of the cases shows a ten percent drop in $E(\hat{R}_u)$ with virtually no change in the response rate. In practice, however, the response rate does in fact improve from T_1 to T_2 . If this response increase was a result of an increase in p_e , and if p_u was prevented from improving by a hard core of unobservable unemployed persons then cases vi and vii show what may happen. Specifically, there has been a five percent change in the expectation of the estimator. It is possible to construct examples like this indefinitely. To what extent any of these factors apply to a specific survey, each practioner will have to decide for himself.

Case	p_e	p_u	$E(\hat{R}_u)$	$E(n)$
i	0.90	0.95	0.0421	9,020
ii	0.95	0.95	0.0400	9,500
iii	0.95	0.90	0.0380	9,480
iv	0.95	0.98	0.0412	9,512
v	0.96	0.90	0.0376	9,576
vi	0.92	0.95	0.0413	9,192
vii	0.98	0.95	0.0388	9,788

Table 3. POSSIBLE UNEMPLOYMENT BIASES

3.5 Coverage

The case in which some p_i equal zero is usually referred to as a coverage problem. It means that some persons who should appear in the sample have no chance of actually entering. It follows from the earlier discussion that efforts to improve the coverage will contribute to the rotation bias effects by increasing some of the p_i . Unfortunately, if the group which is not being covered tends to have a certain characteristic the bias effects can be dramatic. For example, suppose that there is a hard core of "unobservables" who tend mostly to be unemployed. To be specific, consider the example of section 3.4 in which $N = 10,000$, $N_u = 400$ and $R_u = 0.040$. In addition assume that there has been a coverage loss of one half percent or fifty persons and that twenty percent of these are unemployed. Then with equal probabilities $p_u = p_e = 0.95$ it is easy to calculate that $E(n) = 9452.5$ and $E(\hat{R}_u) = 0.0392$ so that a two percent bias has been introduced. Next suppose that the "uncovered" group has even more unemployment than supposed, specifically that out of the fifty persons missed, twenty are unemployed. Then $E(n) = 9452.5$ as before, but $E(\hat{R}_u) = 0.0382$, a four and one half percent bias. To push the example still further, suppose that the coverage problem jumps to one percent with $p_u = p_e = 0.95$ and forth of the "unobservables" are unemployed. Then $E(n) = 9405$, $E(\hat{R}_u) = 0.0364$, and the bias has jumped to nearly ten percent. Finally, if there is a one percent coverage error, forty of whom are unemployed, coupled with $p_e = 0.95$ and $p_u = 0.90$,

then $E(n) = 9387$ and $E(\hat{R}_u) = 0.0345$ which is a bias of about fourteen percent. Notice that the response rate is not necessarily indicative of the bias behaviour. In order of their presentation above, the values of $E(n)$ were 9452.5, 9452.5, 9405 9387 which for most practical considerations would be considered to be unchanged.

It will be recalled from section 3.4 that in the experience of a number of practioners $p_u > p_e$ at T_1 , and it was shown that, if true, this would cause an upward bias. For example if $p_u = 0.95$ and $p_e = 0.90$ $E(\hat{R}_u) = 0.0421$. Consequently, if, (i) $p_u > p_e$ at T_1 , (ii) $p_e > p_u$ at T_2 , (iii) a coverage problem appears which is associated with unemployed persons, then combining the calculations made above shows that $E(\hat{R}_u)$ may drop from 0.0421 to 0.0345. This is a change of twenty percent without any real change in unemployment. It is relevant that the data of Waksberg and Pearl[15] suggest that coverage tends to have a rotation group bias type behaviour. This has also been the Canadian experience[16].

IV. SUMMARY AND DISCUSSION

(1) In this paper it has been shown that systematic changes in the response probabilities can cause the type of systematic bias that has been observed in rotation sampling. Under certain assumptions, the expected value of the estimator must change from the first time to the second time that a rotation group appears in the sample.

(2) Are the basic hypotheses reasonable?

(i) The first necessary hypothesis is that the probability of a response actually being obtained is related monotonically to the characteristic exhibiting the bias. It seems clear from experience that this can actually occur. Indeed in the case of number of children per family, it would be surprising if it were otherwise. Surely the families with children are more likely to be found at home. The suggestion of such an association is not new. There is a large literature on this problem, see Kish's discussion, [8].

(ii) The second hypothesis required is that the probability of response changes from T_1 to T_2 . In many studies there can be little doubt that this is true, because there is a systematic, significant increase in the response rate. Such a significant change in the response rate must be a result of changing probabilities. In particular, an increase in the response rate must mean that an overall increase in the response probabilities has occurred. This is not surprising because the managers of every survey are working towards the goal of maximization of the response rate. On the other hand, it is important to notice that there can be systematic biases without any noticeable change in the response rate.

(3) The hypothesis that the y_i do not change from T_1 to T_2 has also been used. This hypothesis was made in order to study the effect of changes in the p_i 's in an otherwise static

situation. In a forthcoming paper by Mallows and Williams[17], this assumption has not been made and some very interesting results have been obtained. Three of these are as follows.

(i) The magnitude of a potential bias can be made very large, even in very innocent-looking situations. For example, biases of one hundred percent can be obtained even with response rates over ninety percent.

(ii) To attempt to study socio-economic patterns by the study of matched individuals may be highly misleading. For example, is it true that people tend to answer questions about their economic status differently in a first interview than a second? To examine this, it is tempting to construct estimates which are based only on those individuals who appear in the survey at both 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 is shown to be a false statement by Mallows and Williams[17].

Obviously, however, if there are systematic reporting changes, these changes will evidence themselves in the estimates. Such phenomena may or may not exist and this paper does not concern itself with their presence or absence.

(iii) An estimation scheme has been developed and is described in the paper.

The author wishes to convey his appreciation to David Brillinger and Colin Mallows for a number of helpful discussions.

V. LITERATURE CITED

- [1] Hansen, M. H., Hurwitz, W. N. and Madow, W. G., Sample Survey Methods and Theory, John Wiley and Sons, New York, Vols. I and II, 1953.
- [2] Cochran, W. G., Sampling Techniques, Second Edition, John Wiley and Sons, New York, 1963.
- [3] Patterson, H. D., Sampling on successive occasions with partial replacement of units, Jour. Roy. Stat. Soc., B, 12, 241-255, 1950.
- [4] Eckler, A. R., Rotation sampling. Ann. Math. Stat., 26, 664-658, [1955].
- [5] Rao, J. N. K., and Graham, Jack E., Rotation designs for sampling on repeated occasions. Jour. Amer. Stat. Assoc. 59, 492-509, 1964.
- [6] U. S. Bureau of the Census, The Current Population Survey--A Report on Methodology. Technical Paper No. 7, U.S. Government Printing Office, Washington, D.C., 1963.
- [7] Finkner, A. L., Methods of sampling for estimating commercial peach production in North Carolina. North Carolina Agr. Exp. Stat. Tech. Bull. 91, 1950.
- [8] Kish, Leslie, Survey Sampling, John Wiley and Sons, New York, 1965.
- [9] "Measuring Employment and Unemployment." Report of President's Committee to Appraise Employment and Unemployment Statistics. The White House, 1962.
- [10] Politz, A. N. and Simmons, W. R., An attempt to get the "not at homes" into the sample without call-backs. Jour. Amer. Stat. Assoc. 44, 9-31, 1949.
- [11] Politz, A. N. and Simmons, W. R., An attempt to get the "not at homes" into the sample without call-backs. Jour. Amer. Assoc. 45, 136-137, 1950.
- [12] Hartley, H. O., Discussion of a paper by F. Yates, Jour. Roy. Stat. Soc., 109, 37, 1946.
- [13] Deming, W. Edwards, Verbal discussion, Consultant in Statistical Surveys, Washington, D.C., 1967.
- [14] Harris, Louis, Verbal discussion, Louis Harris Associates, New York, New York, 1967.
- [15] Waksberg, Joseph, and Pearl, Robert B., The effects of repeated interviews in the current population survey. Paper presented at the 47th National Conference of the American Marketing Association, Dallas, Texas, 1964.
- [16] Fellegi, Ivan, Verbal Discussion, 1967.
- [17] Mallows, C. L.; and Williams, W. H., Systematic biases in panel surveys. To be submitted to the Journal of the American Statistical Association.