1970

The Systematic Bias Effects of Incomplete Responses in Rotation Samples

William (Bill) H. Williams
CUNY Hunter College

Recommended Citation
W.H. Williams, "The Systematic Bias Effects of Incomplete Responses in Rotation Samples" Public Opinion Quarterly 32 (1970): 593 - 602.

THE SYSTEMATIC BIAS EFFECTS OF INCOMPLETE RESPONSES IN ROTATION SAMPLES*

BY W. H. WILLIAMS

Rotation samples are frequently used in continuing surveys in order to obtain estimates of changes in a characteristic over time as well as separate estimates of the characteristic at specific points in time. Rotation designs involve the retention of some sampling units and the replacement of others.

It has been observed in some studies that there are systematic changes in the estimate of a characteristic, depending on the frequency of appearance of a rotation group in the sample. It is shown in this paper that these systematic changes *must* occur provided (1) the probability of a selected unit actually appearing in the sample is monotonically related to the characteristic under measurement, and (2) the probability of a selected unit actually appearing in the sample *changes* monotonically from one observation point to the next. Some numerical examples showing the form and magnitude of the potential biases are included.

W. H. Williams is Professor of Statistics at the University of Michigan.

ROTATION SAMPLING is used for continuing studies in which there is interest in estimating change from month to month (say) as well as in obtaining separate estimates for individual months. 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.[1]

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 the sample for the first time, another for the second, and the third had been in the two previous months. After

* The author wishes to convey his appreciation to David Brillinger and Colin Mallows for a number of helpful discussions. The work reported here was supported by the U. S. Bureau of the Census and Bell Telephone Laboratories while the author was employed by both organizations.

1 See, for example, M. H. Hansen, W. N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory*, New York, Wiley, vols. I and II, 1953; W. G. Cochran, *Sampling Techniques*, 2d ed., New York, Wiley, 1963; H. D. Patterson, "Sampling on Successive Occasions with Partial Replacement of Units," *Journal of the Royal Statistical Society*, B, Vol. 12, 1950, pp. 241-255; A. R. Eckler, "Rotation Sampling," *Annals of Mathematical Statistics*, Vol. 26, 1955, pp. 664-458; J. N. K. Rao and Jack E. Graham, "Rotation Designs for Sampling on Repeated Occasions," *Journal of the American Statistical Association*, Vol. 59, 1964, pp. 492-509, and Leslie Kish, *Survey Sampling*, New York, Wiley, 1965.

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),[2] conducted monthly by the U. S. Bureau of the Census, one-eighth of the sample is new each month. Each new 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.

Systematic biases have been observed in rotation group studies 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.[3] 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?[4] Does the interviewer influence the respondent in such a way that he gives different responses from one month to the next? Such a hypothesis may be acceptable for the unemployment estimates but seems less likely for the number of children in the Bell System study. Similar behavior exists for other characteristics in the CPS study, for example for estimated vacancy rates and families with salaries over $15,000.[5]

Before leaving this description of the problem, it is relevant to intro-

[2] 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.

[3] Joseph Waksberg and Robert B. Pearl, "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.

[4] The problem of rotation group biases in unemployment statistics was a point of concern of the President's Committee on Employment and Unemployment, "Measuring Employment and Unemployment," Washington, D. C., The White House, 1962, p. 300.

[5] See Waksberg and Pearl, *op. cit.*

TABLE 1
CPS TOTAL UNEMPLOYMENT 1955–61

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

(Index numbers, all groups combined equals 100)

duce "one-time" surveys with call-backs. These are compared with rotation samples in the next section, but it is to the point to present some data from one now. The data, taken from a multiple mail survey by Finkner,[6] are presented in Table 2 below, and it is clear that a systematic behavior similar to the rotation group bias appears. Experienced practioners will of course recognize that this pattern is common in call-back and mail surveys. It will be shown later that this pattern and the rotation group bias can have a similar cause.

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

TABLE 2
FINKNER DATA: MULTIPLE MAIL SURVEY II INCOMPLETE SAMPLES

	Number of Growers	*Per cent of Population*	*Average No. of Fruit Trees per Grower*
Response to first mailing	300	10	456
Response to 2d mailing	543	17	382
Response to 3d mailing	434	14	340
Nonrespondents after 3d mailing	1,839	59	290
Total population	3,116	100	329

[6] A. L. Finkner, "Methods of Sampling and Estimating Commercial Peach Production in North Carolina," North Carolina Agricultural Experiment Station Technical Bulletin 91, 1950.

such a way that the probability of inclusion, p_i', equals one, $i = 1, 2, \cdots, N$. 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 here except to point out that a good description of the work has been given by Kish (*op. cit.*, pp. 532–562). The papers by Politz and Simmons[7] and by Hartley[8] are relevant to this work in that an attempt is made to estimate the individual response probabilities.

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 involving periodic reinterviews.

THE EFFECTS OF THE UNKNOWN PROBABILITIES

At the first appearance. Suppose that the N units in the sampled area have characteristics y_i, $i = 1, 2, \cdots, N$. The objective is that responses be obtained from each of them with probability $p_i' = 1$, $i =$

[7] A. N. Politz and W. R. Simmons, "An Attempt to Get the 'Not at Homes' into the Sample without Callbacks," *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 9-31; A. N. Politz and W. R. Simmons, "An Attempt to Get the 'Not at Homes' into the Sample without Callbacks," *Journal of the American Statistical Association*, Vol. 45, 1950, pp. 136-137.

[8] H. O. Hartley, "Discussion of a Paper by F. Yates," *Journal of the Royal Statistical Society*, Vol. 109, 1946, p. 37.

$1, 2, \cdots, N$. 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 $\hat{\bar{Y}}_1 = (\sum_{i=1}^{n_1} y_i)/n_1$, which is a ratio estimator with expectation, $E(\hat{\bar{Y}}_1) \doteq (\sum_{i=1}^{N} p_i' y_i)/(\sum_{i=1}^{N} p_i')$. 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 over-all 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* behavior.

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 is also expected that some of the characteristics y_i, $i = 1, 2, \cdots, N$ will have changed. One of the purposes of rotation sampling is to obtain efficient estimates of this change. However, since 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 call-back 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 at T_2, and the estimator, $\hat{\bar{Y}}_2 = (\sum_{i=1}^{n_2} y_i)/n_2$, has the approximate expectation $E(\hat{\bar{Y}}_2) \doteq (\sum_{i=1}^{N} p_i'' y_i)/(\sum_{i=1}^{N} p_i'')$.

The special case of proportions. A case of special interest is that in which there are two classifications, such as employed and unemployed.[9]

[9] It should be emphasized that these two categories are referred to as "employed"

If p'_e denotes the probability of an employed person actually being interviewed at T_1, and p'_u denotes the analogous probability for an unemployed person, and $y_i = 1$ if unemployed and 0 if employed, then $\hat{R}_{u1} = n_{u1}/(n_{e1} + n_{u1})$ and $E(\hat{R}_{u1}) = N_u p'_u/(N_e p'_e + N_u p'_u)$. Similar expressions can be written for the unemployment rates at T_2. The generalization to more categories presents no difficulties.

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 $\hat{\bar{Y}}_1$ and $\hat{\bar{Y}}_2$ 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 $(\sum p''_i y_i)/(\sum p''_i)$ compare with $(\sum p'_i y_i)/(\sum p'_i)$? To this end the following points can be easily made.

1. If $p''_i = kp'_i$, the expectations at T_1 and T_2 are the same, but are not necessarily equal to $\bar{Y}$.

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 1? To answer this suppose that $p_i = ky_i^\alpha$, 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.

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 the larger y values have a higher probability of entering the sample and that at T_2 all units enter the sample. This 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 $E\hat{\bar{Y}}_2 \leq E\hat{\bar{Y}}_1$, 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.*

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 $E\hat{\bar{Y}}_2 \geq E\hat{\bar{Y}}_1$ and a systematic change appears in the opposite direction. This again is solely a result of changing probabilities because

and "unemployed" simply 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.

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 $E(\hat{R}_u) \lesseqgtr$ True Rate, iff $p_u \lesseqgtr p_e$, which is an obvious intuitive result. Second, if the probabilities for employment and unemployment each change in different proportions from T_1 to T_2, as follows, $p''_e = k_1 p'_e$, $p''_u = k_2 p'_u = ck_1 p'_u$, then it can be easily shown that $E(\hat{R}_{u2}) \lesseqgtr E(\hat{R}_{u1})$ iff $c \lesseqgtr 1$. For example, if $c < 1$, $k_2 < k_1$, and $E(\hat{R}_{u2}) < E(\hat{R}_{u1})$. This means that if the biggest change in probability from T_1 to T_2 is associated with employed persons, then a decrease in the expected value of the estimator must occur 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, and Harris,[10] and Kish.[11]

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 5 per cent 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 3 per cent 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" characteristics of unemployed persons p_u drops, then a comparison of the cases shows a 10 per cent 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 resulted from 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 5 per cent 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 practitioner will have to decide for himself.

Coverage. The case in which some $p_i = 0$ 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

10 W. Edwards Deming and Louis Harris, discussion.
11 *Op. cit.*

TABLE 3

POSSIBLE UNEMPLOYMENT BIASES

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

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 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 per cent or fifty persons and that 20 per cent of these are unemployed. Then with equal probabilities $p_u = p_e = 0.95$ it is easy to calculate that $E(n) = 9{,}452.5$ and $E(\hat{R}_u) = 0.0392$, so that a 2 per cent 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 $E(n) = 9{,}452.5$ as before, but $E(\hat{R}_u) = 0.0382$, a 4.5 per cent bias. To push the example still further, suppose that the coverage problem jumps to one percent with $p_u = p_e = 0.95$ and forty of the "unobservables" are unemployed. Then $E(n) = 9{,}405$, $E(R_u) = 0.0364$, and the bias has jumped to nearly 10 per cent. Finally, if there is a one per cent coverage error, forty of whom are unemployed, coupled with $p_e = 0.95$ and $p_u = 0.90$, then $E(n) = 9{,}387$ and $E(R_u) = 0.0345$, which is a bias of about 14 per cent. Notice that the response rate is not necessarily indicative of the bias behavior. In order of their presentation above, the values of $E(n)$ were 9,452.5, 9,452.5, 9,405 and 9,387, which for most practical considerations would be considered to be the same.

It will be recalled that in the experience of a number of practitioners $p_u > p_e$ 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 $p_u > p_e$ at T_1, $p_e > p_u$ at T_2, and 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 20 per cent without any real change in unemployment.* It is relevant that the data of Waksberg and Pearl (*op. cit.*)

suggest that coverage tends to have a rotation group bias type behavior. This has also been the Canadian experience.[12]

SUMMARY AND DISCUSSION

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.

Are the basic assumptions reasonable?

The first necessary assumption 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, for example, Kish's discussion, *op. cit.*).

The second assumption 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 over-all increase in the response probabilities has occurred. This is not surprising, because the managers of every survey are working toward this goal. On the other hand, it is important to notice that there can be systematic biases without any noticeable change in the response rate.

The assumption that the y_i do not change from T_1 to T_2 has also been made. This is a convenient assumption because it permits an unconfounded examination of the effects of the unknown probabilities. In an actual survey the characteristics of the changes in the y_i would be superimposed on any effects due to the systematic behavior of the response probabilities. In a forthcoming paper by C. L. Mallows and the author,[13] this assumption has been dropped and some interesting results obtained. One of these results is that extremely large biases can occur in *very* innocent-looking situations. Another is that the study of matched sets of individuals, as for example in complete follow-up surveys, can be highly misleading.

[12] Ivan Fellegi, discussion.

[13] C. L. Mallows and W. H. Williams, "Systematic Sampling Biases in Panel Surveys," submitted to the *Journal of the American Statistical Association.*

The hypothesis that rotation group biases are caused by socio-psychological conditioning has been put forward by various people.[14] Obviously, if there are systematic reporting changes, these will evidence themselves in the estimators. Such phenomena may or may not exist and this paper does not concern itself with their presence or absence.

The problem of estimation has not been discussed in this paper. A procedure has been suggested in the forthcoming Mallows–Williams paper for use in a restricted set of circumstances.

[14] See Waksberg and Pearl, *op. cit.*