

## Re-interview, Revisited

Dhirendra N. Ghosh

Synectics for Management Decisions, Inc., 1901 North Moore St. Suite 900 Arlington VA 22209

### Abstract

The reinterview method is the usual technique for estimating the response variance and the response bias. It is assumed that the respondent does not remember or is not influenced by the original response. The gross discrepancy rate (GDR) is the ratio of the sum of the two types of discrepancies to the total number of respondents in a yes/no type of question. The response variance is half of the GDR. Response bias is also estimated from the ratio of the net difference between the two discrepancies and the total. However this assumption that the respondent neither remembers nor influenced by the original response is not realistic. A measure of the memory effect is derived in this paper. The GDR, the sample response variance and the response bias are rectified based on this measure of the memory effect.

**Key Words:** Gross discrepancy rate, index of inconsistency

Evaluation of survey data has become increasingly important. It involves investigation into the quality of data generated from the field. Evaluation stresses the validity aspects of data, rather than the precision or reliability aspects, which is measured in sampling variance.

A large sample survey, which collects data for many individuals, has the inherent possibility that the data harbors a large amount of error. These errors arise due to non-coverage, non-response, measurement errors and data analysis. It is essential that each of the above sources should be thoroughly investigated and some measure assigned to each of the above mentioned sources. Then one can approximately estimate the total of the errors and can gauge the quality of the data. This is essential since policy makers, researchers and other users of the data obtained from a survey should have a reasonable idea regarding the quality and dependability of the estimates generated from a survey.

The technique of re-interviewing a subsample of respondents is a powerful tool in this task of evaluating survey data. At present, the U.S. Census Bureau in its demographic surveys utilizes the re-interview technique primarily for detecting interview falsification and estimating the response variance. However, re-interview techniques have tremendous potential for estimating and reducing nonsampling error. In order to use it cost effectively, it is essential that a carefully planned research program is undertaken to rest the different possible applications. We outline below such possible methodologies to confront various sources of sampling error.

The re-interview technique is the usual technique for estimating the response variance and response bias. It is assumed that the respondent does not remember, or is not influenced by, the data reported in the first interview. The set up is illustrated below for a dichotomous variable.

		1st interview	
		1	0
2nd interview	1	$a$	$b$
	0	$c$	$d$

$$N=a+b+c+d$$

Where  $a, b, c, d$  are the cell frequencies.

$\frac{b+c}{N}$  is called the gross discrepancy rate (GDR). It can be shown that  $\frac{b+c}{2N}$  is the response variance. Also if we assume that if the respondent reports the same answer at both times, it represents the true value, the difference  $(b-c)/N$  represents the response bias. The index of inconsistency (IOI) is given by  $\frac{GDR}{2\sigma^2}$ . Otherwise a reconciled value is obtained from the respondent which leads to the direct calculation of response bias.

Reconciliation is a means of improving responses. Since it is very unlikely that the reconciled responses are actually error free, they can be used to identify the expected direction of bias and the relative amount of bias, but cannot provide precise estimates of the size of the bias. Furthermore the reconciliation process does not alter consistent errors made in both the original and re-interview. Of course the main reason for doing the reconciliation is to provide at least a rough guide to the size of response bias. An unbiased estimate of the response bias under the above model is given by net difference rate (NDR).

$$NDR = \frac{c - b}{n}$$

In the calculation of NDR equal numbers of errors in opposite directions offset each other and the remaining non-offsetting part of the total error is counted. In the GDR calculation there is no opportunity for one error to offset resulting in every error being counted. For numerical data, the NDR is the average difference between the original and the re-interview. For categorical unordered data the net difference rate is more of a general indicator of offsetting error than a direct measure

We described above the present usage of the re-interview technique as used in the U.S. Census Bureau and other survey agencies. It indicates a moderate level of use of this technique. However this technique can be more extensively utilized to improve the quality of the data collected in a demographic survey. We indicate below the areas where it can make a difference.

1. Ethnic groups can be identified which are more prone to response error. Specifically bias. Other characteristics of the population such as gender, age group, can be factors in response error or bias.
2. The strata (rural or urban), (geographical regions) where the response error bias is more pronounced can be identified. This can be taken into account in the allocation of sample size to different strata (compensating for large non-sampling error in areas the sample size to reduce sampling variance).
3. Questions can be identified which are most vulnerable for response error. The optimum format for a question may be determined from well planned re-interview programs.
4. The re-interview can be conducted with a different eligible member of the household (with the original respondent not present at the time of the re-interview). This response variance is a component of overall response variance. If the re-interview is conducted with the original respondent, this component is missed. And this component of response variance can only be measured for re-interviewers.
5. The effect of the respondents interviewer interaction can also be measured from the re-interviewing program.

One problem of re-interview which plagues the statisticians is that if the respondent remembers the response he/she provided at the first interview, he/she will tend to provide the same answer. We indicate below a somewhat novel method for measuring this component.

At the time of the re-interview the respondent may be asked if he/she remembers the original response and it will be noted. This can be asked before the re-interview or after the re-interview. Thus the following table is constructed from the two responses.

	Remembers original response	Does not remember original response	
Same response	$a$	$b$	$a+b$
Different response	$c$	$d$	$c+d$
	$a+c$	$b+d$	

The difference between  $\frac{a}{a+c}$  and  $\frac{b}{b+d}$  can be taken as a measure of the memory effect and can be eliminated from the measures discussed above.

We had assumed that if the responses in the first interview and the second interview agree then the answer is true. However, some of these include the cases where the respondent remembers the first response and repeats it in the second interview. Thus a fraction of the “true” values must be eliminated before a rectified measure of GDR, response variance and response bias can be constructed.