Survey and Statistical Computing 1996
Edited by R Banks *et al.*

Weighting for Non-Response

Peter Lynn

Abstract

Non-response is a problem which pervades almost all survey research. This paper reviews ways in which adjustments for the biasing effects of non-response can be made at the analysis stage. The emphasis is practical rather than theoretical. Practical constraints on the choice of weighting method (or whether to weight at all) are discussed, as well as statistical implications and how to assess them, and methods of developing a weighting scheme. Ways of obtaining information about non-respondents, or about the population, are discussed. Some suggestions are made for improving the prevalence and quality of non-response weighting. The discussion is illustrated with real examples from surveys carried out by SCPR.

Keywords

Survey non-response; weighting; bias; precision; sampling frames; post-stratification

1. Introduction

Almost all survey research suffers from non-response. Non-response is a failure to collect required data from a sample member. It can take one of two forms: unit non-response is a failure to collect any information at all from a sampled unit, whereas item non-response is a failure to collect a particular item of information from a sampled unit which has supplied other information. This paper concerns unit non-response, and will limit discussion to the case of quantitative surveys with pre-selected random samples. The term non-response will be used as shorthand for unit non-response.

Statistically, non-response causes two types of problems. First, by reducing the sample size it causes an increase in the standard errors of estimates. This is a minor issue, as the approximate level of non-response can usually be predicted, so the selected sample size can be set at an appropriate level to yield (approximately) a required achieved sample size. The second and more important problem caused by non-response is the introduction of bias. If the non-respondents are not a random subset of the sample - in other words, if their characteristics differ systematically - then the achieved sample will produce biased population estimates. It is for this reason that survey researchers are concerned about non-response.

There are two approaches to tackling the effects of non-response. One is to minimise the effects of non-response at the data collection stage. This may involve introducing measures which aim to maximise the response rate, for example (Morton-Williams, 1993; Lievesley, 1986). The other approach is to make statistical adjustments at the analysis stage. It is desirable to combine both approaches, but this paper in concerned only with the latter.

The paper begins by presenting examples of the nature of non-response bias and sources of information about non-response bias (sections 2 and 3) and description of common weighting strategies (section 4). There follows discussion of the practical considerations that typically constrain the development of a weighting strategy (section 5) and the statistical implications of weighting (section 6). The paper concludes by summarising key aspects of current practice and identifying a few challenges for the future (section 7).

2. The Effects of Non-Response

Table 1 shows the nature of non-response bias in the Scottish School Leavers Survey (SSLS), in terms of qualifications gained at school (see also Lynn and Farrant, 1994; Lynn and Purdon, 1994; Lynn 1996a; Taylor 1996). It is clear that the better-qualified are more likely to respond to the survey, and consequently that the achieved sample is biased against the less well qualified. For example, 52% of the achieved sample have one or more Higher grade, compared with 46% of the issued sample. However, this information alone says little about the bias of any survey estimates that may be required. Interest is almost certainly not restricted to simply estimating the proportion of pupils who leave school with a particular level of qualifications. Indeed, that information is already available on the sampling frame - that is how table 1 was constructed. Most uses to which the data are put are likely to be a little more sophisticated.

Table 1: Response to the 1994 Scottish School Leavers Survey, by Qualifications

Highest qualification	Response rate	Profile of issued sample %	Profile of achieved sample %
5+ Higher grades	91.1%	18.0	21.4
3-4 Higher grades	85.1%	13.0	14.5
1-2 Higher grades	81.7%	15.0	16.1
5+ Standard grades 1-3	76.4%	8.1	8.1
3-4 Standard grades 1-3	74.1%	9.1	8.8
1-2 Standard grades 1-3	69.1%	14.5	13.1
Standard grades 4-7 only	62.6%	14.4	11.8
No qualifications	59.6%	7.8	6.1
Base		*4,542*	*3,469*

Other studies too have found lower response amongst the less well educated or qualified (e.g. Lynn et al, 1994; Farrant and O'Muircheartaigh, 1991). This seems to be a fairly consistent pattern, particularly on self-completion surveys. Lower response is also often associated with urban areas, the lower social classes, men, and the youngest and oldest age groups (McDaniel et al, 1987; Barnes, 1992). However, while many of the correlates of response rate are common across surveys, the extent of the correlation varies, and many other factors are survey-specific. Thus, there is no substitute for careful examination of non-response on each survey.

3. Sources of Information about Non-Respondents

To study the nature of non-response bias it is necessary to compare respondents with non-respondents. Typically, there is a wealth of relevant information available about the respondents (from the survey), but little is known about non-respondents.

Sampling Frame

One source of information about non-respondents is the sampling frame. The SSLS data in table 1 is fairly typical of the information that might be available in a survey where the sample is drawn from administrative records of some kind. The information comes from the sampling frame, and is limited to a few basic descriptors. In the case of the SSLS, the sampling frame contains information on qualifications gained at school, sex, stage of leaving school, school type, and region. In fact, this particular sampling frame is perhaps more informative than many, as qualifications and stage of leaving in particular are quite highly correlated with a number of key survey estimates. Even so, the information is limited.

Geographical Information

For surveys of the general population (in Britain), relevant information on the sampling frame is extremely limited. If the frame is the Postcode Address File (Lynn and Lievesley, 1991; Lynn, 1992) the only information is the postcode itself. There is no data about the people resident at each sampled address. However, the postcode can be used to link to a wide range of useful geographical information, such as Census of Population Small Area Statistics, population densities (e.g. of postcode sectors), credit scores, and consumer databases. This area-level information can then be used to analyse response rates, as has been done in table 2.

Table 2: Response to the Health Survey for England 1994, by Type of Area

	Response rate	Profile of issued sample %	Profile of achieved sample %
Large urban/ city centre	73%	21.5	20.1
Other urban/ suburban	78%	58.9	59.0
Rural	83%	19.6	20.9
Base		*11,703*	*9,117*

Source: SCPR. The survey methodology and response are described in Colhoun and Prescott-Clarke (1996)

Interviewer Reports

It is usually possible to obtain some useful information about all sample members (and therefore about non-respondents) either from the sampling frame itself, or by linking to other data sources via geographical or other classificatory indicators on the frame. But even if no such information is available, there are other ways of obtaining information about non-respondents. One is to collect some information as part of the main survey data collection exercise.

For example, on surveys which involve interviewing people in their own homes, interviewers can be asked to record information about the address and the immediate surrounding area. This might include characteristics of the dwelling and surrounding properties. This information can then be related to response rates. Table 3 presents response rates to the 1992 British Crime Survey in terms of housing type. The information on housing type was recorded by interviewers in the field. It can be seen that response rate was highest at detached houses, and lowest at purpose-built flats (and addresses which the interviewer could not code).

Table 3: Response to the 1992 British Crime Survey, by Housing Type

		Response rate	Profile of issued sample	Profile of achieved sample
			%	%
House:	detached	82.6%	19.5	21.0
	semi-detached	79.6%	30.2	31.3
	end terrace	79.2%	7.3	7.6
	mid-terrace	77.7%	20.4	20.6
Maisonette		74.9%	1.7	1.7
Flat:	converted	72.3%	2.9	2.7
	purpose-built	70.3%	11.7	10.7
Rooms/bedsit		75.6%	0.3	0.3
Unable to code		51.2%	6.0	4.0
Base			*13,117*	*10,059*

Source: SCPR. For details of the survey methodology, see Hales (1993). For further analysis of the survey response, see Lynn (1996b)

Surveys of Non-Respondents

Other methods of obtaining information about non-respondents (e.g. Kersten and Bethlehem, 1984; Lynn, 1996b) include follow-up surveys of non-respondents and very short interviews as part of the main data collection exercise (along the lines of, "I'm sorry to hear that you do not wish to take part in our survey, but would you just answer three quick questions for me"). But any strategy which involves an attempt to collect information directly from survey non-respondents is, almost by definition, likely to suffer severe non-response problems itself.

Modelling

An alternative to the use of direct information about non-respondents is to model their characteristics (Bartholomew, 1961; Drew and Fuller, 1980; Colombo, 1992). The model might combine theory with data about the respondents - for example, relating respondents' characteristics to speed/ ease of response, and positing a relationship between the non-respondents and the slowest/ most difficult respondents. Table 4 shows the relationship between social class (collapsed to three categories) and the number of visits that an interviewer needed to make in order to achieve an interview on the British General Election Study. Those interviewed at first call were disproportionately manual workers, whereas those for whom three or more calls were necessary were most likely to be non-manual workers. Social class of the non-respondents is unknown, but it would seem reasonable to propose that they might be similar to those respondents who were most difficult to contact and interview, or perhaps even slightly more extreme in their social class profile.

Table 4: Ease of Response to the 1987 British General Election Study, by Social Class

	Interviewed at 1st call	Interviewed at 2nd call	Interviewed at 3rd call	Interviewed after 4+ calls (inc. reissues)	Non-Respondents
	%	%	%	%	%
Non-manual	41.3	44.3	49.7	50.4	Unknown
Manual	46.4	44.5	39.5	36.9	Unknown
Self-employed	6.2	6.3	6.1	7.9	Unknown
Unclassifiable	6.0	4.9	4.7	4.8	Unknown
Base	*726*	*1,024*	*722*	*1,354*	*1,637*

Source: SCPR. Heath et al (1991), appendices 2 and 3, give details of the survey methodology and response.

Panel and Follow-up Surveys

If the sample is selected from the responding sample to a previous survey, the data collected by that previous survey can be used to compare respondents and non-respondents. This is likely to mean that a much larger amount of relevant information is available than for other surveys. Table 5 analyses response to a survey carried out in 1993 on discrimination against gay men and lesbians (Snape et al, 1995). The sample was selected from the respondents to the National Survey of Sexual Attitudes and Lifestyles (NATSAL). The discrimination survey achieved a relatively poor response rate, perhaps partly due to the highly sensitive topic and partly due to the elapse of a couple of years since the NATSAL, leading to out-of-date sample details. The availability of full NATSAL data for sample members enabled the identification of variables which correlated highly with response rate. However, it must be remembered that this only refers to non-response to the follow-up survey conditional upon response to the NATSAL. In considering overall response to a follow-up survey, non-response to the initial survey must also be taken into account.

Table 5: Response to 1993 Discrimination Survey, by Age and Activity Status (Gay sample)

	Response rate	Profile of issued sample	Profile of achieved sample
		%	%
Aged under 25	34.9%	13.7	8.0
Aged 25-34, in work*	58.4%	26.5	25.7
Aged 25-34, unemployed	22.7%	3.5	1.3
Aged 35 or over	69.4%	56.3	65.0
Base		*627*	*377*

Source: SCPR. Note: "in work" includes a small number in training or education.

4. Methods of Weighting For Non-Response

Most strategies for weighting for non-response (Elliot, 1992) involve dividing the respondents into a set of comprehensive and mutually exclusive groups, referred to as weighting classes. A weight is then applied to each class, or to each member of the class.

Sample-Based Weighting Using Response Rates

Typically, the weight applied is equal to the reciprocal of the class response rate (in other words, the ratio of selected sample to achieved sample size). The success of this approach depends on the classes discriminating in terms of response rate, and in terms of the survey measures. The classes may be defined more or less arbitrarily, by looking at tabulations of response by the available variables, and picking one, or a couple in combination, that seem to produce a range of response rates and which are also presumed to be associated with the survey variables. Alternatively, a more systematic approach may be used, involving a multivariate analysis technique such as regression or AID analysis to identify the best predictors from amongst the set of variables available. However, although one is looking for classes which produce a range of response rates, one should be wary of classes with particularly low response rates, as these will lead to large weights and hence increased variance. As a rule of thumb, it might be advisable to avoid creating classes with a response rate of less than, say, one-fifth of the overall survey response rate.

Other Forms of Sample-Based Weighting

Alternatively, the weights may be something other than the reciprocals of class response rates. For example, one of the simplest models that could be applied to the data of table 4 would assume that the social class distribution of non-respondents was equal to that of those interviewed after four or more calls. Then, there would be two weighting classes. Weights would be (1,637+1,354)/1,354 for those interviewed after four or more calls and 1 for others. The weight for the first class is the reciprocal of the response rate conditional upon there having been no response at the first three calls.

Another example of weights which are not simple inverse response rates is when weights are defined by the estimated coefficients of a multiple regression model (where survey response is the dependent variable). With this strategy, the weights are reciprocals of estimated (by the model) response rates for classes, where the classes are defined as all possible combinations (represented in the sample) of categories of the predictor variables. (Note that an alternative use of a regression model is simply to define the classes, to which simple inverse response rate weights can then be applied.)

Population-Based Weighting

If there is no information about non-respondents, an alternative to sample-based weighting is to create weighting classes based on variables which are known both for respondents (from the survey data) and for the population as a whole (from an external source, such as a Census). Weights are then applied in proportion to the ratio of population to achieved sample. This strategy is frequently used, sometimes in conjunction with sample-based weighting, and perhaps other weighting too. As well as correcting for non-response, weighting in this way simultaneously incorporates an element of post-stratification (Holt and Smith, 1979). However, caution is needed as observed differences between the survey and population distributions may, at least partly, be due to the effects of differences in the data gathering procedures, such as question wording, data source and mode of data collection.

5. Practical Constraints

An important constraint is the existence of information which can be used to define weighting classes. This information must be available for both respondents and non-respondents in order to implement sample-based weighting. In the examples presented in tables 1 to 3 above, classes could be defined in terms of qualifications, population density, and housing type respectively. If more than one variable is available, classes can be defined by combinations of variables, as in table 5.

To be successful, the weighting classes must discriminate in terms of response rate. If they do not, weights will not vary between classes, so the weighting will not have any effect. Often, variables available from the sampling frame do not discriminate well. The survey researcher may have to anticipate that this is likely to be the case, and consequently decide to ask interviewers to obtain other information. This decision cannot be taken after the event, but it is not a cost-free option.

Weighting classes should also discriminate in terms of important survey variables. If the mean amongst respondents for a certain variable is the same in each weighting class, then the weighting cannot have any effect on the estimate of the mean, no matter how large the range of weights (though it will have a detrimental effect on the standard error of the estimate). Usually, the only way of assessing empirically the relationship between potential weighting variables and survey variables is to base an analysis on the respondents. The assumption then being made is that the relationship between the weighting and survey variables is the same for non-respondents as for respondents within weighting classes. Expert knowledge must be called upon to assess whether this is likely to be true. Non-response weighting will only remove that part of the non-response bias that can be explained by the weighting classes. Any bias within classes will remain.

On some occasions, it may not even be possible to assess the relationship between the weighting and survey variables amongst respondents. In this case, recourse to expert knowledge and expectations may be necessary in order to create the classes. An example is the situation where a data set is being prepared for distribution to a number of users, each of whom might have different analysis objectives, many of which may not have been precisely formulated at the time that the weighting is being developed. The users may not be sophisticated enough or willing to develop their own weights and there may in any case be a strong argument for ensuring that all users use the same set of weights.

One very important practical consideration is the time available to develop a weighting strategy. Most surveys are run to fairly strict timetables, and weighting is not an element of the survey process to which clients typically attach great importance (at least, not at the survey planning stage). Consequently, little if any time and resources tends to be set aside for weighting. And yet clients usually want clean, weighted, data very soon after the completion of field work. It is therefore important to be able to develop weighting strategies quickly and efficiently, as the alternative might be that no weighting gets done at all. This consideration suggests favouring the multivariate analysis methods referred to in section 4 above, as these can be implemented relatively quickly.

6. Statistical Implications

The aim of non-response weighting is to reduce bias in survey estimates. The statistical price to be paid for this is an increase in the variance of estimates. There is obviously a trade-off to be made. The ideal would be to minimise the root mean square errors (rmse) of estimates.

In order to develop a weighting strategy which minimises rmse it is necessary to estimate the effect of weighting on both bias and variance. The effect on variance can usually be estimated quite accurately for any specified variable by calculating appropriate complex sampling errors with and without the weighting. But this process is resource-intensive, and not always feasible in the light of the timetable constraints referred to in section 5 above. An adequate substitute is to predict the likely effect of weighting on variance of estimation by assuming that population variances are equal within weighting classes. Under that assumption, a simple formula can be used to estimate the design effect due to weighting (Collins and Hedges, 1977).

The effect on bias is, however, somewhat more elusive. Survey practitioners often resort to simply assuming that the effect is to reduce bias by a worthwhile amount, and use some rule of thumb to limit the amount by which they are prepared to let the variance be increased. To estimate empirically the bias-reduction effect of a particular set of weights requires both an estimate of the bias initially present in the (unweighted) sample and an estimate of the bias remaining in the weighted sample.

There are a number of difficulties in estimating these quantities. First, initial bias can only be estimated in respect of any variables available for both respondents and non-respondents, or both respondents and the population as a whole. As already noted, these variables tend to be limited. Second, these variables will probably have been used to define the weighting classes, in which case the weighting will, by definition, remove any bias. For both of these reasons, estimates of bias reduction in these variables may not tell us very much about bias reduction for survey variables. Third, point estimates of bias will be subject to large standard errors. In other words, estimates of rmse will themselves tend to be subject to large rmse, making it difficult to be very conclusive about the effects of weighting. Nevertheless, creditable attempts to estimate rmses have been made (e.g. Ecob, 1990; Lynn *et al*, 1994) and this should be encouraged.

Another issue to be aware of is the risk that weighting designed to reduce bias could actually increase or introduce bias. The risk arises only with population-based weighting. Systematic differences in the nature of the population data and sample data could cause apparent differences in the distributions, which are due to factors other than non-response bias. Removing those differences could introduce bias. For example, suppose 85% of survey respondents classify themselves as “white”, with the other 15% distributed across a number of other categories. This is compared with, say, Census data for the geographical area that was covered by the survey, which shows 90% “white”. It might be suggested that whites are under-represented in the survey, and so should be given a weight of 90/85, while the others are weighted by 10/15. But suppose the difference was entirely due to the existence of 5% of people who would have been classified as white by the Census, but would have chosen one of the other categories on the more extensive list provided on the survey. Then, the proposed weighting will result in a weighted proportion of 93.33% who are “Census white”, compared to the correct 90% in the unweighted sample.

7. Conclusion

Perhaps as a consequence of the difficulty of estimating rmses, coupled with a general cautious attitude towards weighting, statisticians are not particularly successful at convincing survey researchers of the importance of weighting for non-response. In turn, researchers often do not or can not convince survey clients that the effect is likely to be beneficial, let alone that it warrants a commitment of resources. There seems to exist a common attitude that weighting for non-response is a specialist technique of interest only to statisticians or sophisticated secondary analysts of survey data, and not something that really matters at the stage of producing a basic survey report.

This attitude is misguided. Non-response bias is present, and observable, on most surveys. It is highly desirable, and usually not difficult, to reduce this bias considerably through the use of appropriate weighting. The techniques needed to employ successful weighting for non-response are not new, are not difficult to understand, and are becoming increasingly easy to implement with the continual production of new and better software. For example, multiple regression and AID analyses can be performed easily within major survey data analysis packages. Survey researchers should perform non-response analyses as a standard part of the survey process, and be prepared to implement a weighting strategy if necessary.

Survey clients could benefit from being a little less obsessed about the survey response rate, and examining other quality measures. Ultimately, some of the most important quality measures associated with a survey are the rmses of key estimates. These are affected not just by non-response bias, but by the net effect of non-response bias and non-response weighting (“residual non-response bias”). It can be very inefficient for a survey client to spend tens or even hundreds of thousands of pounds on a survey, some of which will be targeted on maximising response rates, and then fail to make a relatively small investment which could dramatically reduce the residual bias.

In summary, the following suggestions are made:

- Weighting for non-response should become more widespread for all types of surveys.
- Survey researchers should be familiar with the issues involved, and able to persuade their clients of the importance of weighting for non-response.
- Analysis of non-response bias should be a standard part of the survey process, including the use of multivariate techniques, and should be routinely reported in survey technical reports, along with descriptions of any non-response weighting.
- Where sampling frame information is lacking or not relevant, survey researchers should consider building into the design the collection of data by interviewers or linking to geographical data.
- Researchers should attempt to assess rmses more often, and publish the results of these attempts.
- Survey clients should be persuaded of the importance as quality indicators of measures of overall error, and made aware that non-response weighting can be effective in reducing overall error.

About the Author

Peter Lynn is deputy director of the Survey Methods Centre at Social and Community Planning Research (SCPR), an independent social research institute. He provides advice and support to SCPR’s survey research teams on statistical aspects of survey design and administration, as well as running methodological projects. He has worked on a wide range of surveys and has specialised in sample design issues, as well as the methodology of postal and self-completion surveys. He teaches and runs courses on aspects of survey methodology, and is currently honorary secretary of the Social Statistics Section of the Royal Statistical Society. He can be contacted at SCPR, 35 Northampton Square, London EC1V 0AX; tel. 0171 250 1866; fax 0171 250 1524; e-mail p.lynn@scpr.ac.uk.

References

Barnes R (1992) Studies of non-respondents in OPCS household surveys using census data, *Surv Meth Bull.* 30.

Bartholomew D (1961) A method of allowing for not-at-home bias in sample surveys. *Appl. Stat.s* 10, 52-59.

Colhoun H and Prescott-Clarke P (eds) (1996) *The Health Survey for England 1994. Volume II: Survey Methodology and Documentation.* HMSO, London.

Collins M and Hedges B (1977) *Variable sampling fractions and the effect of weighting*. SCPR methodological working paper no.4.

Colombo R (1992) Using call-backs to adjust for non-response bias, in *Survey and Statistical Computing* (eds Westlake A et al) Elsevier, North Holland.

Drew J and Fuller W (1980) Modelling non-response in surveys with call-backs. *Proceedings of the American Statistical Association section on survey research methods*.

Ecob R (1990) *A comparison of possible weighting schemes for the baseline wave of the West of Scotland Twenty-07 study*. MRC Medical Sociology Unit working paper no.22.

Elliot D (1992) *Weighting for Non-Response: A survey researcher's guide*. OPCS, London.

Farrant G and O'Muircheartaigh C (1991) Components of nonresponse bias in the British Election Surveys, in *Understanding Political Change* (Heath et al). Pergamon, Oxford.

Hales J (1993) *1992 British Crime Survey (England and Wales): Technical Report*. SCPR, London.

Heath A, Jowell R, Curtice J, Evans G, Field J and Witherspoon S (1991) *Understanding Political Change*. Pergamon, Oxford.

Holt D and Smith T M F (1979) Post stratification *Journal of the Royal Statistical Society* A, 142, 33-46.

Kersten H and Bethlehem J (1984) Exploring and reducing the nonresponse bias by asking the basic question. *Statistical Journal of the United Nations*, vol.2, 369-380.

Lievesley D (1986) *Unit non response in interview surveys*. SCPR Survey Methods Centre unpublished paper.

Lynn P (1992) Applications of computers in sample design and selection, in *Survey and Statistical Computing* (ed.s Westlake A et al). Elsevier, North Holland.

Lynn P (1996a) *Scottish School Leavers Survey Series: The Survey of 1993 Leavers - Technical Report*. SCPR, London.

Lynn P (1996b) Who responds to the British Crime Survey? Paper presented to the Fourth ISA Conference on Social Science Methodology, Colchester, 2-5 July.

Lynn P and Farrant G (1994) *Scottish School Leavers Survey Series: The Survey of 1992 Leavers - Technical Report*. SCPR, London.

Lynn P and Lievesley D (1991) *Drawing General Population Samples in Great Britain*. SCPR, London.

Lynn P and Purdon S (1994) An analysis of factors related to non-response in a postal survey of young people. Paper presented to the RSS International Conference, Newcastle upon Tyne, 13-15 September.

Lynn P, Purdon S, Hedges B and McAleese I (1994) *England and Wales Youth Cohort Study: An assessment of alternative weighting strategies*. Employment Department Research Series YCS no.30.

McDaniel S, Madden C and Verille P (1987) Do topic differences affect survey non-response? *Journal of the Market Research Society* vol.29, 55-66.

Morton-Williams J (1993) *Interviewer Approaches*. Dartmouth, Aldershot.

Snape D, Thomson K and Chetwynd M (1995) *Discrimination against Gay Men and Lesbians*. SCPR, London.

Taylor S (1996) *Scottish School Leavers Survey Series: The 1995 Survey of 19-year-olds - Technical Report.* SCPR, London.