

---

# **The General Household Survey as a sampling frame: the example of the Survey of Poverty and Social Exclusion<sup>1</sup>**

**Joanne Maher**

## **1. Introduction**

The practice of using an existing survey to identify a sample of people or households with certain characteristics is commonly used as a relatively inexpensive method of carrying out special population surveys when an alternative sampling frame is not readily available. The use of these follow-up surveys raises some interesting methodological issues which are explored in this article in connection with a survey of Poverty and Social Exclusion (PSE) which was carried out as a follow-up to the 1998/99 General Household Survey<sup>2</sup> (GHS).

This article will describe the background to the PSE, the methodological issues raised and considerations made by conducting the survey as a follow-up to the GHS, principles which can also apply to other follow-up surveys. Because aspects of this work have not been published elsewhere, this article gives a fuller report of the project for researchers who are interested in the PSE<sup>3</sup> itself as well as in the general methodological issues raised.

A follow-up survey is designed to re-interview former survey respondents who are chosen on characteristics identified from responses given at an earlier interview. As such it enjoys the benefits (and inherits the difficulties) of the ‘parent’ survey as well as presenting a number of methodological issues peculiar to this type of survey. The issues that need to be addressed include sample design, questionnaire design, respondent burden and non-response bias (including opportunities for weighting). These issues are foremost in many survey designs but take on another dimension in the context of the follow-up survey. The aims of the follow-up survey are constrained by the parent survey’s design and content but are also enhanced by the rich source of information on which the follow-up survey can be based.

The follow-up survey for the PSE was drawn from respondents to the 1998/9 GHS, who were interviewed in detail about their circumstances and views on a range of issues associated with poverty and social exclusion. By using the GHS it was possible to:

- Draw a sample with known characteristics;
- Use GHS data for cross-topic analysis in conjunction with PSE data;
- Reduce the length of the PSE questionnaire by not repeating questions previously asked in the GHS interview;
- Identify characteristics of PSE non-responders in the GHS dataset.

## **2. The Sample**

Findings from the PSE preparatory research called for a reliable source of income data, on which to measure poverty on the basis of income. The PSE needed to interview enough people in low income households to allow analysis of this group. If a readily available and useable list of such households had been available, then this could have been used as a sampling frame or a large scale sift could have been carried out to identify such households. However, no such reliable frame exists and a sift would have been expensive, particularly as income data is difficult to collect reliably in this way. A follow-up survey was, therefore, deemed the most efficient way of identifying the sample and the GHS was the survey that collected all the information necessary for a follow-up. As the authors of *Poverty and Social Exclusion in Britain*<sup>4</sup> note, “the PSE survey makes major use of income data from the GHS but measures poverty in terms of both deprivation and income level: whether people lack

items that the majority of the population perceive to be necessities, and whether they have incomes too low to afford them.”

The PSE benefited from using the GHS as a sampling-frame by being able to select a sample with known characteristics. The sample design had to meet three main objectives:

1. Sufficient cases were required for the analysis of key variables by sub-groups.
2. Sufficient cases were required for separate analysis of households and individuals in Scotland.
3. Sufficient cases of low income households and respondents were required to examine their characteristics.

The fact that the GHS dataset contains income details, household composition and regional information allowed the sample to be selected on these key variables.

Identifying individuals involved a three-stage process<sup>5</sup>:

1. a number of areas were selected from all those used for the 1998/9 GHS;
2. a number of households were selected from each of the areas;
3. one individual was chosen from each sampled household.

### 3. Sampling from edited data

#### 3.1. Calculating equivalised income

The sample gave a greater probability of selection to people in lower income groups. An equivalised income measure was developed by Jonathan Bradshaw of York University, in conjunction with ONS. The McClements equivalence scale, which is used as the standard by ONS, was felt not to be appropriate for the PSE, as it does not assign sufficient weight to children, particularly young children. The scale used for the PSE was designed to take account of this. Each member of the household was assigned a value, shown in Table 1:

**Table 1 Equivalised income scale**

Type of household member	Equivalence value
Head of household	0.70
Partner	0.30
Each additional adult (anyone over 16)	0.45
Add for first child	0.35
Add for each additional child	0.30
If head of household is a lone parent, add	0.10

The values for each household member were added together to give the total equivalence value for that household. This number was then divided into the gross income for that household. For example, the equivalence value for a lone-parent household with two children is  $0.7 + 0.35 + 0.3 + 0.1 = 1.45$ . If the household's gross income is £10,000, its equivalised income is £6,897 ( $=£10,000/1.45$ ).

The income, household composition and relationship information collected on the GHS allowed the equivalised income to be created for the purpose of drawing the sample. However, it was not only in the sample design that information collected during the GHS

interview was put to use in the design of the PSE. GHS data was also used to enhance PSE fieldwork procedures, questionnaire design and weighting for non-response.

### ***3.2. Selecting areas***

Because the GHS is a nationally representative sample of private households in Britain, any probability sub-sample from the GHS will also be nationally representative. For the smaller sample size required for this survey, it would have been most economical to sub-sample the GHS areas and select all GHS respondents in those areas. However because of the requirement to oversample lower income households, a larger sample of areas was required. In fact, 70% of GHS areas in England and Wales were selected (360 areas from a total of 518 ). All of the 54 Scottish areas were sampled to provide sufficient cases for separate analysis of the Scottish data.

To allow for variation in income within the areas the list of PSUs was sorted on area and equivalised income quintile groups before any selections were made.

### ***3.3. Selecting Households***

A sample of 2846 households was taken from the selected areas. PSE required sufficient cases in lower income households to allow analysis of those found to be in poverty by an income based measure. However, for the purposes of testing theories of poverty and social exclusion, a sample of all income groups was required. For this reason households were selected in different proportions according to household income. The equivalised income of the household was grouped into quintiles which were then sampled in the following proportions:

**Table 2 Proportions sampled in each quintile group**

<b>Quintile group</b>	<b>Proportion sampled</b>
Bottom quintile (lowest income)	40%
Fourth quintile	30%
Third quintile	10%
Second quintile	10%
Top quintile (highest income)	10%

### ***3.4. Selecting individuals***

One adult aged 16 or over was selected at random from each sampled household using a Kish grid. This was done in preference to interviewing all eligible adults to reduce interview length and costs. It was also believed that when asking opinion questions (on which the PSE was heavily based), household members may influence each other's answers; the risk of this was hoped to be reduced by interviewing one person.

### ***3.5. Partial and proxy interviews***

Only those who had given a full interview in the 1998/9 GHS were eligible for selection. As the intention was to analyse the PSE in combination with GHS data, proxy and partial interviews were excluded from the sample because of their limited usefulness for analysis. By removing proxies and partials from the sample item non-response in the GHS data was reduced for PSE cases when analysing combined PSE and GHS data.

#### **4. Contacting the respondent**

An advantage of using a survey such as the GHS as a means of drawing a follow-up sample is that a more personalised approach could be adopted when re-contacting the GHS respondent than could normally be used when contacting an unknown household at a sampled address for the first time. Advance letters were addressed to the selected respondent, except when a name was not given during the GHS interview, when the letter was addressed to 'the resident'.

Where contact telephone numbers were available, interviewers made initial contact with the respondent by telephone. Once an appointment was made with the respondent the interviews were conducted face-to-face. In the event of a broken appointment, interviewers were instructed to make a maximum of two visits to an address before recording a non-contact, unless they were already in the area and could make an extra call without driving out of their way.

#### **5. Questionnaire Design**

Choosing a survey design based on a follow-up of the GHS meant that detailed information was already available on those topics covered by the GHS. Questions asked during the GHS interview need not be repeated again in the follow-up survey. The 1998/9 GHS questionnaire held 68 questions on income which, if included again in the PSE would have taken some time from the PSE interview that could be used for other questions. By following up the GHS respondents the PSE was able to concentrate on operationalising concepts of poverty and social exclusion within the survey instrument and was able to reduce the length of the PSE questionnaire.

As the follow-up interviews<sup>6</sup> took place between 6 and 18 months after the original GHS interview, a small number of follow-up questions were included in the PSE questionnaire to record changes to the household composition, tenure, employment and income. Where appropriate this was carried out by the technique of dependent interviewing. "Dependent interviewing can be used to remind the respondent of previously reported information or to probe for inconsistent responses between data provided in the current interview and data previously provided" (Mathiowetz & McGonagle, 2000).

Questions were included in the PSE questionnaire which asked whether there had been any changes to the respondent's individual and household income since the GHS interview. The amount by which the income had changed and reasons for the change in income were also recorded. It was possible to feed the date of the GHS interview into the question to aid the respondent's recall.

The PSE questionnaire included questions relating to housing conditions. The GHS housing tenure question was fed-forward to the PSE questionnaire to check for changes in tenure over time. Questions were repeated in the PSE for the purposes of verification, but did not have to be asked again unless the situation had changed<sup>7</sup>.

The PSE dependent interviewing was designed to remind the respondent of the answer they gave during the GHS interview. This approach was adopted to aid the respondent's recall and reduce the length of the interview. Such proactive interviewing is adopted to reduce respondent burden; however it is not always considered suitable for follow-up interviewing as it allows the respondent to agree too readily with his/her previous response but it was believed that the tenure question would not be disadvantaged by this approach.

Little research into the effects of dependent interviewing and what constitutes best practice exists (Mathiowetz and McGonagle, 2000). The effects of rewording questions were therefore carefully considered and care was taken to change the question wording as little as

possible. This approach seemed to have worked well and interviewers did not report any difficulties with feed forward questions.

Dependent interviewing is not always an appropriate approach. It is not appropriate where a respondent is asked an opinion-based question and they may change their response over time. For example, the GHS health questions were repeated in the PSE questionnaire as experience from other surveys, such as the Health Education Monitoring Survey (HEMS), showed that respondents gave different answers on their health status over time. Of particular interest in HEMS were those who changed their reports of having a limiting long-standing illness. Work by ONS's Question Testing Unit has shown that the question asking whether the respondent has a longstanding illness that limits their activities presents a number of difficulties to the respondent, which may explain why there is variation in the answers given over time. These health questions were repeated in the PSE interview as future analysis of health and social exclusion was anticipated.

It would have been possible to route respondents to the PSE questions on health and well-being and use of health and social services by their answers to the GHS health questions. However, it was feared that routing in this way would increase the risk of asking respondents inappropriate questions or under-reporting health problems in the PSE questionnaire, where the respondent's health might have changed or their opinion on their health changed since the GHS interview. Comparison of the GHS and PSE responses to the limiting and long-standing health questions showed that some respondents' opinion about their health had changed between the two surveys. The responses can be seen in Table 4 below.

Although the distribution of longstanding illness for PSE and GHS respondents were broadly similar (see Table 3 below) there was some variation in the answers given by individuals between surveys (see Table 4). For example, of those who reported a limiting long-standing illness in the GHS 14% reported that they did not have a longstanding illness in the PSE (see Table 4 below). These changes could reflect real changes in health but could also be attributed to context effects or to the fact that answers to questions of this nature are often affected by how a person feels on the day they are interviewed and this illustrates a common finding, that such questions should not be treated as factual.

**Table 3 The distribution of longstanding illnesses for PSE and GHS respondents**

	GHS (%)	PSE (%)
Limiting Longstanding Illness	30.2	32.0
Non-Limiting Longstanding Illness	15.7	14.0
No Longstanding Illness	54.0	54.0
Total	100.0	100.0
base (numbers)	1534	1534

**Table 4 GHS responses to Limiting Longstanding illness by PSE responses**

GHS		%	%	%	%
		Limiting Longstanding Illness	Non-Limiting Longstanding Illness	No Longstanding illness	Total
PSE	Limiting Longstanding Illness	<b>22.9</b>	4.0	5.1	<i>32.0</i>
	Non-Limiting Longstanding Illness	3.1	<b>6.5</b>	4.5	<i>14.1</i>
	No Longstanding Illness	4.2	5.3	<b>44.3</b>	<i>53.8</i>
	Don't know	0	0	0.1	<i>0.1</i>
	Total %	<i>30.2</i>	<i>15.7</i>	<i>54.0</i>	<i>100</i>
	Base	464	541	829	1534

## 6. Interview length and Respondent Burden

Respondent burden was an issue for the content of the PSE. In particular, the interview length was of greatest concern for respondent burden and the ultimate effects on the response rate. "Long interviews, or unpredictable lengths, can interact with interviewer behaviour, and can limit the number of interviews which can be completed in any one day. Where fieldwork periods are restricted, or days in an area limited to control costs, response may be affected." (Martin and Matheson; 1999) A great deal of work was done to keep the questions on the PSE to a minimum to reduce respondent burden.

The average length of interview was 60 minutes. With older respondents or those who had literacy problems, it took about 90 minutes. Questions requiring a lot of thought or those involving difficult concepts, such as assessments of absolute and overall poverty, added to the length of interview for some respondents. Three types of data collection were adopted in PSE: face-to-face interviews, a self-completion module and a card-sorting exercise. The mixed mode interview was adopted to allow a change of pace of the interview in an attempt to keep the respondent's interest, but for some this may have become burdensome.

## 7. Response rate

The length of the questionnaire affected the response rate. ONS interviewers are required to give an assessment of how long the interview is likely to take when making an appointment, to ensure that respondents set aside sufficient time. Some sampled individuals refused to take part on hearing that the interview was likely to last for an hour. The relatively short field period (a month), meant interviewers also did not have sufficient time to call back on many households to attempt refusal conversion.

Of the 2,431 eligible individuals, 1,534 (63%) were interviewed, the vast majority completing a full interview. One of the drawbacks of using follow-up surveys as a means to identify special populations is that it suffers from two waves of non-response; that of the original survey and that of the follow-up. As such this response rate was disappointing, and may reflect some of the factors outlined above. However, the availability of information about non-responders means that it is possible to compensate for non-response by weighting.

## 8. Weighting the data

As noted earlier, the PSE interviewed one person per household, over-sampled households in Scotland, and over-sampled households in the lowest quintile groups of equivalised income. Several weights were calculated to allow for the probability of selection and also to compensate for non-response. The final weight is made up of four elements: a weight for country, a weight for income quintiles, a weight for the probability of selection of individuals from households, and a weight for non-response. Details of the weighting for probability of selection are presented in Appendix 1.

A benefit of using the GHS as a parent survey for PSE was that it was possible to compensate for non-response by weighting. To weight for non-response at this stage, response rates were analysed to see how they varied according to categories defined by 1998 GHS data. Variables that were compatible with the PSE were selected from the GHS data to investigate the effect of non-response in PSE.

The variables examined were as follows:

- Sex
- Age
- Marital status
- Number of adults in the household
- Government Office Region
- Tenure
- Type of accommodation
- Number of vehicles
- Household type
- Limiting longstanding illness
- Equivalised income (quintiles)
- Economic status (ILO definition)
- Age of head of household
- SEG of adults
- SEG of head of household
- Highest educational qualification

Consumer durables (TV, satellite/cable receiver, video recorder, freezer, washing machine, tumble drier, dishwasher, microwave, telephone, CD player, computer).

CHAID (Chi-squared Automatic Interaction Detector) Answer Tree was used to split the sample down to find the groups which best explain the variation in (weighted) response. It is the compatibility of the subjects covered in the GHS with those in the PSE that allowed a suitable set of variables to be selected. Within these groups weights are formed on the reciprocals of the (weighted) response rates and multiplied by the weight for the probability of selection of individuals from households to create final weight. The variables which Answer Tree selected (from the list above) as having an effect on the response rate were:

- Age of Head of household

- Socio-Economic Group of Head of Household
- Government Office Region
- Equivalised income
- Consumer durables (Satellite/cable receiver, CD player, microwave, washing machine, tumble drier)
- Limiting longstanding illness
- Household type

Appendix 2 shows the eighteen weighting classes that were produced from the output and details of the variables selected by Answer Tree in column2. The percentages, response rate and weight for non-response are also shown.

## 9. Conclusion

The use of the GHS as a sampling frame for the PSE highlights some of the common issues inherent in any follow-up survey. The GHS provided a high quality data source for the PSE in a variety of ways. The GHS provided a sampling frame which could represent particular groups in the proportions required for analysis where no other sampling frame was readily or economically available. The challenge of the sample design in using equivalised income as the basis for selection could be met by the GHS income data.

The compatibility of the GHS and PSE topics made the GHS an appropriate source of data not only as a sampling frame but as a rich data source for further analysis. Use of the GHS data allowed the PSE interview to concentrate on operationalising different concepts of poverty and social exclusion rather than taking up interviewing time asking many of the questions carried on the GHS again. The consideration was not only for interview length but also because repeating questions unnecessarily would have reduced respondent tolerance of the survey. Interviewers reported that PSE respondents were concerned about the length of the interview and the impact of this on their limited leisure time.

With any follow-up survey, there will be an element of ‘double non-response’, that is the non-response inherited from the parent survey plus the additional non-response from the follow-up. However, the parent survey can provide us with detailed information about non-responders to the follow-up and this can feed into a weighting scheme.

The usefulness of a survey as a sampling frame will always depend on the group of the population that is needed for the follow-up survey but in the case of the PSE, the GHS has proved to be an efficient way of identifying the refined sample.

## References

Mathiowetz, N. and McGonagle, K.(2000) An Assessment of the current state of dependent interviewing in household surveys, *Journal of Official Statistics* p401

Martin, J and Matheson. J (1999) Response to declining response rates on government surveys, *Survey Methodology Bulletin*, Issue 45

## Notes

- 1 The author would like to thank the following people for the work they contributed to the design and implementation of the methods adopted in the PSE survey: Charles Lound, Dave Elliot; Angie Osborn, Alan Francis, Olu Alaka, Karen Irving and Michaela Pink.



- 2 The General Household Survey (GHS) has collected information on a variety of topics over the last 30 years. The survey is primarily designed to provide information for policy use by its sponsoring government departments. Its usefulness lies in the potential for cross-topic analysis, and this potential has been exploited by many GHS data users. The multi-purpose design and relatively large sample size of the GHS has also benefited those wanting to conduct research into a particular sub-sample of the population (for example those aged 65 and over). In this way the GHS has been used not only as a data source for analysis but has also been employed as a sampling frame for follow-up surveys.
- 3 The PSE was carried out by Social Survey Division of the Office for National Statistics on behalf of a consortium of the Universities of York, Bristol, Loughborough and the London School of Economics who were funded by the Joseph Rowntree Foundation. The PSE aimed to measure poverty and social exclusion in Britain today, focussing on deprivation indicators based on a societal definition of the necessities of life. The team at the four universities conducted preparatory research, defining concepts and testing ways to operationalise them, including question testing work, before commissioning ONS to advise on the survey design and to conduct the fieldwork. The consortium of academics had conducted research into poverty in conjunction with MORI in the Breadline Britain Surveys in 1983 and 1990. PSE was designed to update these surveys. In addition to the follow-up to the GHS described in this paper the ONS Omnibus Survey also asked a representative sample of the population of Great Britain their views on what constitutes the necessities of life in present-day Britain in June 1999.
- 4 From Poverty and Social Exclusion in Britain – D. Gordon et. al. 2000, p.10
- 5 The sample was drawn from edited GHS data. This was necessary due to the complexity of the data that was used to select the sample. The use of edited, rather than raw data, has implications for the timescale of such projects.
- 6 The 1998/9 GHS and PSE were conducted as a computer-assisted personal interview (CAPI) using a Blaise program. Answers recorded during the GHS interview were fed forward into the PSE CAPI interview for the interviewer to verify with the respondent and update where circumstances had changed. Feeding information from one CAPI questionnaire to another helped maintain the smooth flow of the interview.
- 7 For example, in PSE the tenure question was asked in this way:
 

“Last time we spoke you said that you occupied your accommodation in the following way.  
Buying it with the help of a mortgage. (fed-forward answer from GHS interview)  
Is this still correct?”

  1. Yes
  2. No

If the respondent answered ‘no’ the GHS tenure question was asked again in full

## Appendix 1 Weighting for the probability of selection

Since the GHS is an equal-probability sample of households and individuals, it was only necessary to build weights to account for each stage of the subsampling from the GHS sample.

### *Weighting for countries*

360 areas out of 518 were selected in England and Wales and all PSUs in Scotland. Therefore the initial weight,  $Wt_1$ , has a value of 518/360 in England and Wales and a value of 1 in Scotland.

### *Weighting by quintile group*

Households in different income quintile groups (as defined by the data collected in the 1998 GHS) were sampled at different rates. For each quintile group the sampling rate was used to create a second weight,  $Wt_2$ , which is a product of  $Wt_1$  and the sampling rate.

For the lowest quintile group:  $Wt_2 = Wt_1 * 1$ .

For the fourth quintile group:  $Wt_2 = Wt_1 * 4/3$ .

For the top three quintile groups and those who refused the GHS income section:  $Wt_2 = Wt_1 * 4$ .

### *Weighting by household size.*

The weight needed here depends on the type of analysis to be carried out. Three weights have been created, each of which is a product of  $Wt_2$  and household size and composition.

1. Analysing sampled adults. Here the chance of being included for the survey, given that the household is sampled, is '1' divided by the number of adults in the household. So the weighting factor should be the number of adults in the household:

$Wt_{3a} = Wt_2 * \text{number of adults in household}$

2. Analysing families. Here the chance of a family unit being included, given that the household is sampled, is equal to the number of adults in the family divided by the number of adults in the household.

$Wt_{3b} = Wt_2 * \text{number of adults in household} / \text{number of adults in family}$

3. Analysing adults in a sampled family. Here each adult can be included in the file either by being sampled themselves, or because their spouse is sampled. Therefore given that the household is sampled, for single adults, the probability is one divided by the number of adults in the household and for adults in couples the probability is two divided by the number of adults in the household.

$Wt_{3c}$

If the respondent has a partner,  $Wt_{3c} = Wt_2 * \text{number of adults in household} / 2$ .

If the respondent does not have a partner,  $Wt_{3c} = Wt_2 * \text{number of adults in household}$ .

## Appendix 2 Weighting for unit non-reponse

Weighting Class	Description	Response Rate (%)	Weight for Non-response	Weighted Base
<b>Age 17-28</b>				
<b>Socio- Economic Group of head of household</b>				
1	Professional; Employer-manager; skilled manual; semi-skilled manual; <missing>	32	3.13	986
2	Intermediate non-manual; junior non-manual; unskilled manual	48	2.09	384
<b>Age 29-40</b>				
<b>Government Office Region</b>				
NE; Merseyside; E Midlands; SW; Wales; Scotland;				
<b>Satellite/Cable TV</b>				
3	Have satellite/cable TV	58	1.72	559
4	Do not have satellite/cable TV	65	1.54	807
NW; Yorkshire & Humberside; W. Midlands; Eastern; London; SE				
<b>CD Player</b>				
Have CD Player				
<i>Household type</i>				
5	1 pers; 2+ unrelated adults, m. couple – no children; 2 + families; cohabiting couple – no children	27	3.75	499
M. couple - dep children; lone parent; lone parent – indep children; same sex cohab; cohab couple – dep children				
<i>Tumble drier</i>				
6	Have tumble drier	61	1.64	910
7	Do not have tumble drier	36	2.78	389
8	Do not have CD player	29	3.42	253
<b>Age 41-63</b>				
<b>Video recorder</b>				
9	Have video recorder	60	1.68	6019
10	Do not have video recorder	52	1.92	330
<b>Age 64-70</b>				
<b>Government Office Region</b>				
11	NE; E Midlands; W Midlands; SE; SW; Wales	79	1.27	825
12	NW; Merseyside; Yorkshire & Humberside; Eastern; London; Scotland	56	1.78	699

Weighting Class	Description	Response Rate (%)	Weight for Non-response	Weighted Base
	<b>Age 71-100</b>			
	<i>Equivalised income (Quintiles)</i>			
13	1 (top), 2, 3	77	1.30	649
	Quintile 4			
	<i>Microwave</i>			
	Have Microwave			
	<i>Limiting Longstanding Illness</i>			
14	Have non-limiting longstanding illness; no longstanding illness	70	1.44	306
15	Have limiting longstanding illness	53	1.88	203
16	Do not have microwave	51	1.96	275
	Quintile 5 (bottom); <missing>			
	<i>Washing machine</i>			
17	Have washing machine	47	2.15	617
18	Do not have washing machine	38	2.65	204
	All	56		14914