Journal of the Market Research Society, 28, 3, 1986, pages 269-284

Weighting survey results

Trevor Sharot
Audits of Great Britain Ltd

Introduction

The author's recent involvement on the examination board for the MRS Diploma made him aware of the lack of a single, reasonably comprehensive, introductory explanation of the process of weighting. This paper, which is based on a lecture on weighting delivered by the author at a recent MRS education course, aims to fill that vacuum, It is not in itself an exhaustive treatment, in either breadth or depth, and the references provide more detailed expositions of a number of specific techniques, However the sections on rim weighting provide information not believed to be documented elsewhere.

What is weighting? Quite simply, a weight is a multiplying factor applied to some or all of the respondents in a survey. The weight applied to anyone respondent may be less than one (though positive) or greater than one, and a whole range of weights may co-exist among the respondents in a given survey. Their effect is therefore to change the relative importance of the respondents in determining the final tables -no longer 'one man, one vote'. The essential reason for doing this is to achieve a sample 'profile' (usually in demographic or regional terms) which is closer than the actual (ie unweighted) sample to the profile of the target population. However, there are many different reasons why the actual sample may not reflect the population, and the method and effect of weighting differs accordingly. These various situations provide the substance of this paper. They will be treated under three broad groups:

-When unequal probabilities of selection are an integral part of the sample design (variable sampling fractions, or disproportionate sampling).
-When the design has equal probabilities, but the achieved sample is visibly 'unrepresentative' when compared with (say) the population age structure (post-stratification).
-When non-probability sampling methods are used, and weighting is used to achieve a 'balanced' sample.

Finally, traditional cell weighting is contrasted with the alternative method of 'rim' weighting, the properties of which have received little visibility in market research textbooks or papers.

Variable sampling fractions

Optimal allocation

This technique is sometimes employed when the population mean of each survey variable is the parameter of major interest. The sample mean from an 'epsem' design (equal probability of selection method) is an unbiased estimator, but is not necessarily efficient ie its sampling error may not be the minimum obtainable.

If the element standard deviation varies between strata, then improved efficiency results if the most heterogeneous strata are over-sampled (and vice versa). It can be shown algebraically that, for a fixed total sample size, the optimal plan is for the ratio:

$$\frac{\text{Sampling fraction for stratum}}{\text{Standard deviation of stratum}}$$

to be made equal for all strata.

One situation in which this arises is in a retail audit. The universe typically stratifies into a relatively few, large, outlets (eg superstores) and a large number of smaller outlets. The mean rate of sale is much higher in the first group and the variance is accordingly also greater . A higher proportion of this group is sampled to compensate, even reaching total enumeration for the largest retail outlets.

A similar situation occurs in industrial market research when sampling firms within a given industry; size of firm variations are often very great, and so variances again differ .

Generally speaking, the household population does not usually merit variable. sampling fractions (at least on grounds of overall efficiency), as it is reasonably homogeneous.

A further case for variable sampling fractions can be made when the cost of sampling is different in different strata. The objective might then be to minimise overall sampling error subject to a fixed total cost. If the marginal cost per response in stratum i is C_i then the optimal plan is for the quantity:

Sampling fraction x c_i

to be made equal across strata.

If both costs and variability differ between strata, then the quantity:

$$\frac{\text{Sampling fraction x } c_i}{\text{Standard deviation}}$$

is equated across strata.

As a general rule, small differences in variances (or cost) do not merit the complexity of variable sampling fractions. Moreover, even if high variances do occur in very small strata, the gain in efficiency of the overall mean arising from optimal allocation is usually quite small.

Reporting on sub-groups

The above applies when the overall population means (or proportions) are the major variables of interest. Frequently, results are also required for various sub-groups, eg regions, male vs female, social class etc. It is also true that quite small groups are sometimes of much interest (for example, small regions may be employed for test-marketing purposes).

In this situation, it makes sense to redistribute the sample towards the smaller groups to ensure that they are adequately represented in the sample. .We might, for instance, allocate a sample equally among ten regions within Britain. Since variances will usually differ little between regions, this will give equal precision to the regional means and will also maximise the precision of regional differences (as when comparing 'test' and 'control' regions).

The penalty for these improvements is that the overall (national) mean must incorporate regional weights to remove the resulting bias, and will be *less* precise than an epsem sample of the same size. The overall variance is increased by the factor:

$$\frac{n.\Sigma n_i w_i^2}{(\Sigma n_i w_i)^2}$$

This factor is sometimes called the ‘weighting effect’ or ‘weff’’ analogous to the design effect, deff; and results in an effective sample size ‘neff’ given by neff = n/weff; see Conway (1982). For illustration suppose that there are two strata (eg South and North) with sizes in the ratio 2: 1, but we allocate 1000 interviews to each. The weights will be 4/3 and 2/3, giving weighted sample sizes of 1333 and 667 as required. The above factor is then

$$\frac{2000.\ (1000(4/3)^2+1000(2/3)^2}{(1000\ .4/3+\ 1000.2/3)^2}$$

=1.111.

The standard errors of national data are therefore increased by a factor of

$(1.111)^{1/2} = 1.054.$

This may be considered a small price to pay given that the corresponding factor for the smaller of the two regions is $(2/3)^{1/2}$=0.816. (Note that gains at sub-group level are particularly valuable since the sample sizes are that much smaller than the total sample.)

There are numerous instances of the use of the technique (more so than optimal allocation). For example, the AGB Home Audit is a quarterly survey which monitors acquisition of consumer durable household appliances. One important sub-group consists of households moving into newly-built homes. In any 3-month period, they account for 0.2% of all households. Therefore, on an epsem basis, the sample would contain only 50 newly-built homes, which is insufficient for separate reporting. Since they account for 4% of cookers, 12% of telephones and 20% of central heating installations, a rather larger sample is called for (250), with a corresponding weight of 0.2 in the 'all homes' reports. In this particular case the 'all homes' sampling error is also reduced, as the sample also approximates to the optimal allocation.

Various schemes have been proposed to balance the sometimes conflicting requirements of 'total' and 'sub-group' reporting. Kish's method involves allocating the sample so that the quantity:

$$\frac{\text{Sub-group sample size}}{(D^2 + 1)^{1/2}}$$

is equal across strata, where D is the ratio of the sub-group's population to the average population of all the sub-groups. This and other schemes are discussed in Conway *op cit.*

'Unrepresentative' samples

We now look at the case where the achieved sample is visibly 'unrepresentative' even though equal probabilities were imposed at the sampling stage. A number of quite distinct situations may give rise to this.

True post-stratification

There is usually only limited opportunity to impose stratification at the selection stage. With household surveys, the smallest area on which relevant stratifying information is available is a census enumeration district. More usually, counties, constituencies, wards or polling districts form the primary and/or secondary selection units and stratification is employed at these stages. Almost no household-specific information is available at the final selection stage, yet this is potentially where stratification is of most benefit.

The traditional approach to the problem is to draw an unrestricted random sample of households (or individuals) at the final stage, and then to weight the achieved sample to the known population profile. *Providing the deviation of the sample profile from this is a purely random phenomenon*, then this technique will reduce variance, though to a lesser degree than will prior stratification. Note that both the unweighted and weighted results are unbiased in this case.

However, *in practice*, post-stratifying weights are invariably applied in situations where the sample deviations are *not* a purely random occurrence. One major cause of this is non-response.

Non-response

No survey is free of the non-response problem. In fact there are several causes of non-response, but we will group them under three main headings:

Refusals
Non-contacts
Non-effective, non-eligible or unusable elements.

The last one of these covers such situations as empty dwellings, non-domestic addresses, language and literacy problems and so on. The effect of these problems is to reduce the achieved sample size, and usually by different amounts in different strata. Refusals are more common in metropolitan areas than in rural areas. Non-contacts are more common amongst smaller households than larger households. Language and literacy problems occur most amongst the . DE social class and in foreign ethnic groups.

The design and purpose of the survey affects both the degree and effect of non-response. Postal surveys achieve much lower response rates than personal interview surveys. In both cases, chasers or recalls are commonly employed to achieve improved response levels. However, some non-response always remains. In general, its effect depends on the extent to which the response-rate is correlated with each survey variable across the population.

The application of weighting in this case should be distinguished clearly from true post-stratification. The effect of non-response is generally to introduce a bias into the survey results, and the weighting is designed to alleviate this bias. It will also increase the variance (compared with the unweighted results), but this might be considered an acceptable price to pay. (This increase in variance is additional to the increase which results from the reduction in sample size created by non-response.)

The extent to which the weighting does in fact reduce the bias is dependent on a number of factors. The variables used for weighting should be selected so as to correlate highly with the non-response rate. The usual demographic analysis variables will certainly do so to some extent, but other variables may be necessary. For example, urban/rural weights may be important, though this axis is rarely employed for analysis purposes. Sometimes, behavioural controls become necessary. One example of this is found in the BARB television viewing panel. There is a propensity for 'heavy viewers' to be more inclined to join the panel than 'light viewers' (they are at home more and are more interested in the subject matter!). This occurs within the standard demographic cells, so post-stratifying by these variables does not remove the resulting bias. The solution is to either control or weight the sample to the correct proportion of heavy/medium/light viewers (in fact, panel control rather than weighting is used).

An interesting system of weighting for non-response was proposed by Hartley (1946) and developed by Politz & Simmons (1949), by whose names it is generally known. Rather than use recalls to reduce non-response, a single call is made, let us say in the evening. Respondents are questioned as to how many of the previous five evenings they would have been at home and available for interview at about the same time. The different responses (0, 1, 2, 3, 4 and 5) define six strata whose probabilities of inclusion are assumed to be 1/6, 2/6, 3/6, 4/6, 5/6 and 1 respectively. There is a seventh stratum consisting of those not at home on any of the six nights (and therefore not sampled). The results from the sampled strata are then weighted inversely to their probabilities, ie by weights of 6/1,6/2,6/3,6/4,6/5 and 1 respectively.

There are certain drawbacks to this method. First, the scheme has low efficiency, due to the wide range of achieved sampling fractions and weights. Secondly, the weights are themselves subject to response errors, particularly as the information on which they are based is somewhat sensitive. Thirdly, the low response-rate inherent in a single call technique artificially increases the non-response, and in particular the 'seventh stratum' could be significant. Though the Politz-Simmons method is little used nowadays, a related technique is widely used in 'out-of-home' studies, as described below.

Finally, under this heading, mention should be made of work which has examined the relationship between survey results and the number of calls required to achieve contact. Typically, the approach consists of assuming that 'last round' contacts are more similar to non-respondents than is the total contacted sample. One sample treatment is to weight up the 'last round' results to represent non-contacts also, before being added to the previous contacts. (See Moser & Kalton 1972, p 185.)

Deficiencies in the sampling frame

Any omissions or duplicate entries in the sampling frame will clearly affect the probability of inclusion of the affected elements of the population. Much work has been done on this problem in respect of the Electoral Register, when used as a frame for electors. Related to this are techniques for using the Register as the basis of equal probability sampling of adults, or of individuals in any given age range. A number of alternative methods are available, some of which involve weighting. The major ones are clearly described and compared in Hoinville, Jowell *et al* (1978), pp 77-82.

Sometimes, large distortions in probability arise because there is no single sampling frame, and multiple overlapping frames are required to cover the population. In this case, elements stand a probability of selection in direct proportion to the number of frames in which they appear. If it is not practical to eliminate duplications prior to sampling, the selected sample can sometimes be checked against each of the frames to establish estimates of the correct probabilities and weights. If even this is impossible, the survey itself may be used to yield the necessary information.

Examples where this problem has been met include sampling from Yellow Pages, and in drawing a subsample of AGB's interviewers. In the latter case, interviewer lists are held individually for each panel survey. Respondents (that is, sampled interviewers) were asked to state how many surveys they worked on, and this was used to weight back the results.

Given the near-l00% response-rate achieved in this case, it was possible to estimate both the loss of precision in this sampling scheme and the bias that would have resulted in not employing weighting (by comparing unweighted and weighted results).

Non-random sampling

Quota/quasi-random sampling
Under this heading are a variety of techniques employed for 'in-home' surveys, including random walks and random location methods. They generally replace true random sampling at the final selection stage, the earlier stages of selection having been performed randomly.

Whether or not quotas are imposed, the final sample is likely to exhibit some imbalances against the demographic control (or analysis) variables being employed, and post-stratification is again employed. However, this technique cannot be theoretically justified in terms of reducing either bias or variance, at least by reference to selection-based inference, since both the actual and weighted selection probabilities remain unknown. Rather, the intuitive appeal of the approach rests on an implicit *model* which relates the means of the survey variables to the quotas and/or other control variables. Weighting does not overcome the basic theoretical weakness of non-random sampling, that is, that the selection method may introduce skewness in the sample in respect of some unanticipated and un-controllable covariate(s), with consequent measurement bias.

Out-of-home studies
Certain survey topics require that interviewing takes place away from the home. Examples include visitor surveys (eg museums or holiday

resorts), shopper surveys, transport users etc. In each case, interviewers are usually placed at purposively chosen 'sites' and interviewing may be spread over times of day and day of week, in order to achieve a 'cross-section' of responses. The 'sites' may, in the case of travel studies, be selected trains, aeroplanes 'etc.

In all such cases, a particular problem is that the probability of inclusion is not equal across elements, but will be proportional to the length of stay, frequency of shopping, frequency of travel etc. If no correction is made, the resulting sample will over-represent 'long stayers' or 'frequent travellers' and biases may result. In fact, the sampling method yields an equi-probability sample of 'visits' rather than 'visitors', or 'trips' rather than 'trippers'.

An example of this effect may be found in Dance and Hopewell-Smith (1983). Holiday-makers in the South-West of England were asked about their television viewing whilst on holiday. A positive correlation between length of stay and amount of viewing was established. Unless the selection probabilities are appropriately corrected by weighting, this leads to an over-estimate (upward bias) in the estimated mean viewing level.

The solution usually adopted in such cases is similar to the Politz-Simmons technique. Respondents are asked about their length of stay/frequency of visiting as appropriate, and this is used to down-weight the results proportionately. The 'never visit' cell is of course no longer a problem. Whilst the technique is appealing, the basic non-random nature of the sampling method is of course open to selection bias just as for quota sampling.

Weighting methods

Advances in data-processing have had a considerable influence on the way in which weighting is performed. In the days of the 'card-sorter', weighting was either applied manually, or (where appropriate) by duplicating cards in proportion to the desired weight. The advent of fast computers allowed weighting to be largely automated. Population targets are entered as parameters; the computer performs the necessary analysis of the 'achieved' sample and compares this with the targets to calculate the weights. Once vetted for 'acceptability’, these would then be applied automatically in calculating survey tables.

Cell weighting

The commonest method of applying weights is ‘cell’ weighting. The target variables, typically three or four in number (eg age, sex, size of household) are interlaced to form a matrix of target cells and the sample is broken down similarly. A weight is then calculated for each sample cell to achieve the corresponding target. The. overall target may either be the actual sample size (in which case the average weight is 1.0), some nominal sample size (useful for repeated surveys to achieve comparability) or population size (in which case grossing-up is performed simultaneously with weighting).

Complete automation of the process is not always desirable, as cells with small targets (in terms of the expected sample size) may attract an excessively high weight even if the shortfall is, in absolute terms, quite modest. Weights should therefore always be vetted before they are applied to the data. The usual remedy in. this eventuality (assuming further fieldwork is not possible) is to pool the offending cell with its 'nearest neighbour' and calculate an overall weight. The small bias introduced by this 'fix' is worthwhile given the reduction in variance which results. Typically, weights of 2 or more would be handled in this way (unless arising from disproportionate sampling).

Rim weighting

In the last few years, further increases in computing power have permitted a further development known as 'rim weighting'. This system is now available within proprietary survey analysis packages such as Quantum and Star. It is based on practical and theoretical work described in Deming & Stephan (1940). Rather than interlace the required control variables, each is treated on a marginal basis. The sample is weighted to the first such variable (eg age group targets); this set of weights is retained in comparing the balance of the sample with the targets for the second axis (eg social class groups).

Fresh weights are calculated to correct this, which are multiplied by the first weights. This process continues until the last control variable has been appropriately weighted. At this stage, the weighted sample will not be exactly balanced against any of the preceding variables, but the balance is (hopefully) better than prior to weighting. The whole process is then repeated, starting with the first variable, and continuing until either:

(a) satisfactory balance is achieved on all axes (say when all weighted sample subgroups are within 2-3% of the corresponding targets)
(b) no further convergence can be obtained
(c) the number of iterations reaches a pre-set limit.

Deming & Stephan demonstrate that this method normally converges to the 'least-squares' solution; that is, the cells defined by all combinations of the control variables have the property that the weighted cell counts c_i differ least from the raw counts n_i as measured by

$$\Sigma(c_i - n_i)^2 / n_i$$

In practical terms, the method offers the advantage that weighting to key variables may proceed even when the targets are not available on an interlaced basis, and that a larger number of variables may be

balanced simultaneously than could be achieved by cell weighting.
The properties of the method are best understood by example.
Suppose the 'true' (population) profile on two characteristics is given .
by Table 1.

TABLE 1

30	20	10	60
10	10	10	30
10	20	30	60
50	50	50	150

Table 2 shows one imaginary sample in which the row totals are wrong, but the proportions across cells in each row are correct.

TABLE 2

15	10	5	30
20	20	20	60
10	20	30	60
45	50	55	150

Although the column totals are also 'thrown', a single pass through the rows recreates the correct balance. The row weights are 60/30 = 2.0, 30/60 = 0.5 and 60/60 = 1.0 respectively.
So far so good. However, Table 3 shows a sample where all the margins are correct, but some of the individual cells are wrong.

TABLE 3

10	20	30	60
10	10	10	30
30	20	10	60
50	50	50	150

It is clear that rim weighting would make no changes to this sample, so the imbalance would remain: only cell weighting can correct them. A less extreme case is shown in Table 4.

TABLE 4

30	20	10	60
10	10	10	30
30	20	10	60
50	50	50	150

Here, only two cells in the last row are incorrect, and row margins are correct. Application of rim weighting until convergence yields the weighted profile given in Table 5.

TABLE 5

21.9	20.6	17.5	60
6.2	8.8	15.0	30
21.9	20.6	17.5	60
50	50	50	150

Comparison with Table 1 shows that the effect of rim weighting is seriously to distort those cell counts which were originally correct, whilst failing even to correct the ranking of the wrong cells! Further simulation with other examples shows that in general, the process tries to preserve the proportions of the original table, whether or not these are correct.

Indeed, this property is easily demonstrated analytically. Suppose that the sample size in row i, column j is n_{ij}; that the row totals are r_i and the column totals c_j; and the overall sample size is n. Now let the target row totals be R_i, the target column totals C_j, and the grand total M (which may be the target sample size, or the population size if grossing-up is required simultaneously with weighting).
We can write the cell sizes as

$$n_{ij} = r_i c_j (1+d_{ij})/n \quad (1)$$

where d_{ij} are the deviations from 'proportionality'. A little algebra shows that a single pass of rim weighting, first to rows and then to columns, will produce weighted cell sizes n'_{ij} given by

$$n'_{ij} = \frac{R_i C_j (1+d_{ij})}{M(1+a_j)} \quad (2)$$

where each a_j is a weighted average of d_{1j}, d_{2j} ...(the d-values for column j), the weights being the corresponding row targets R_1, R_2 ...

ie $a_j = \frac{\Sigma R_i d_{ij}}{M}$ (3)

Equation (2) shows that rim weighting attempts to preserve the underlying 'irregularities' of the matrix, as given by the d_{ij}, but . introduces the divisors $1 + a_j$ in order to meet exactly the marginal column targets (these being applied last in the example). Of course, columns could be weighted before rows and a similar result would still be obtained. Similarly, with three or more axes, the principle carries through, though the form of the weighted average becomes more complex.

Effect on the data
The effect of this mechanism on survey results has been simulated by assuming that Table 6 represents the means of a hypothetical survey variable.

TABLE 6

95	100	105	98.33
100	100	105	100
80	100	110	101.67
93	100	107	100

Assuming that the survey is unbiased in each of the nine cells, the weighted row, column and overall means are as follows:

	Row means			Column means			Overall means
Table 6 (true)	98.33	100	101.67	93	100	107	100
Table 3	101.67	100	91.67	87	100	105	100
Table 4 (raw)	98.33	100	91.67	89.29	100	105	96
Table 5 (weighted)	99.63	100	95.62	83.05	100	105.25	98.10

Table 3 delivers wrongly ranked row means. Rim weighting does not change these; only cell weighting would do so.

In Table 4, the worst figure is the 3rd row mean, which lies the wrong side of 100. Rim weighting increases this figure, but it still lies below 100; indeed it is still below the overall mean of 98.1

Problems with empty cells
If any of the original cells are empty, the sum-of-squares as defined above does not exist, so there is no least-squares solution. In these circumstances the iterative procedure may still converge, and usually does. Table 7 is a modified version of Table 2 with an empty cell. Iteration to the marginal targets in Table 1 as before gives rise to

Table 8.

TABLE 7

20	10	0	30
20	20	20	60
10	20	30	60
50	50	50	150

TABLE 8

36.3	23.7	0	60
7.2	9.4	13.4	30
6.5	16.9	38.6	60
50	50	50	150

This does not appear to be a 'bad' solution against Table 1, though note that the empty cell remains empty because of the multiplicative nature of the weighting procedure. Whether this solution is in any sense 'best' remains unclear .

A large number of empty cells can cause different problems, as shown by Table 9.

TABLE 9

60	0	0	60
0	10	20	30
0	30	30	60
60	40	50	150

Despite the fact that the row totals are already satisfied, this can never converge owing to the conflicting constraints imposed by Table 1 on the top left-hand cell. Similar cases do occur in practice if there are large numbers of cells, relative to the sample size. The iterative method eventually converges only in the sense that the cycles become identical, but within a cycle each marginal constraint imposes its own solution. Here the solution after row weighting is Table 10A and that after column weighting is Table 10B.

TABLE 10A

60	0	0	60
0	11.6	18.4	30
0	33.4	26.6	60
60	45	45	150

TABLE 10B

50	0	0	50
0	12.9	20.5	33.3
0	37.1	29.5	66.7
50	50	50	150

Whilst this latter problem rarely occurs in practice, the presence of empty cells is not uncommon. For instance, if two of the axes employed are say:

(i) ISBA areas
(ii) Electricity Board areas

most of the implied cells are empty (eg London ISBA x Yorkshire EBA). In practice, we have found that rim weighting still converges, but as remarked above, it is not clear what properties the result has.

A final piece of advice to users of rim-weighting: ensure that the frequency distribution of the weights is produced. This is useful in that the minimum and maximum weights applied can then be determined (lower and upper limits can be set in some proprietary packages to control these). Also, the variance of this distribution plus 1.0, is the weighting effect ('weff') referred to on page 271.

The ethics of weighting

There are those who believe that any (ie all) weighting is an unethical procedure. It should be obvious from this paper that the correct incorporation of weighting in the overall survey design can improve the cost/accuracy equation beyond what could be" achieved by unweighted designs alone. Providing the details of the weighting process are available to user[. of the data, alongside the other relevant information (such as response-rates) it is hard to see why this should be unethical. On the issue of information, the MRS Code of Conduct (1983, section 3.6) states:

"The agency shall provide to the client, whether in the report, proposals or elsewhere ...an adequate description of. ..the size and nature of the sample and details of any weighting methods used (and) weighted and unweighted bases for all conventional tables, clearly distinguishing between the two."

Providing this and other relevant information is also available (much of which is similarly covered by the Code of Conduct) the user should be in a position to judge whether weighting has been properly employed and whether any claimed benefits are in fact achieved. Useful discussion of the latter point may be found in Boyd (1975).

In practice, not all the information may be available, in particular in respect of the effect of the weighting on the precision (standard errors) of final results. That is partly because the calculation of these for a complex sample design is itself a skilled 1ask which may not have been carried out. Also, statistical analysis packages (such as SPSS, SAS and BMD) do not take account of any specific sampling and weighting procedures which may have been employed, and sampling errors or tests of significance provided by such software will usually be in error (and optimistically so).

Summary

This paper has described the major reasons for, and effects of, weighting and compared the two main methods: cell weighting and rim weighting. There are a number of unanswered questions. Firstly, the relative bias and variance of cell and rim weighting are not well understood. A comparative study of the two methods on a common sample with common targets would illuminate this area. Secondly, it is not obvious why (or rather when) the iterative method converges to the least-squares solution. Although Deming and Stephan lay great stress on this result, they do not prove it.[1] Finally, the behaviour of rim weighting with empty cells is not well understood. Further uptake of that interesting technique will hopefully help to answer these questions.

References

BOYD, K T (1975). Sampling large populations. *Market Research Society Conference Papers. 29-35.*

COLLINS, M & HEDGES, B (1977). Variable sampling fractions and the effect of weighting. SCPR Methodological Working Paper No.4. October. .

CONWAY , S (1982). The weighting game. *Market Research Society Conference Papers 193-207.*

DANCE, R & HOPEWELL-SMITH, A (1983). Brief versus budget-need either suffer? A case study of television measurement amongst a transient population. *Market Research Society Papers,* 175-190.

DEMING, W E & STEPHAN, F F (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics, II.

HARTLEY, H P (1946). Comments on paper by F Yates. *JRS*S.A **109**, 37-38.

HOINVILLE, G, JOWELL., R *et al* (1978). *Survey research practice*. London: Heinemann Educational Books.

MOSER, C A & KALTON, G (1972). *Survey methods in social investigation*. London: Heinemann Educational Books.

POLITZ, A & SIMMONS, W (1949). An attempt to get the 'not at homes' into the sample without callbacks. II. Further theoretical considerations for eliminating call-backs. *JASA* **44**,9-31.

[1] I am grateful to James Rothman for pointing this out.