WEIGHTING: WHY, WHEN, AND HOW?*

Leslie Kish, The University of Michigan, Ann Arbor MI 48106-1248

1. Introduction

Fundamental questions about weighting (Fig 1) seem to be the most common during the analysis of survey data and I encounter them almost every week. Yet we "lack a single, reasonably comprehensive, introductory explanation of the process of weighting" [Sharot 1986], readily available to and usable by survey practitioners, who are looking for simple guidance, and this paper aims to meet some of that need. Some partial treatments have appeared in the survey literature [e.g., Kish 1965], but the topic seldom appears even in the indexes. However, we can expect growing interest, as witnessed by six publications since 1987 listed in the references.

Fig. 1 COMMON QUESTIONS ABOUT WEIGHTING

1. WHY SHOULD sample data be weighted?
2. WHEN do sample data NEED weighting?
3. WHEN is it PROPER to weight sample data?
4. WHEN is it IMPORTANT to weight data? When does it make a real difference?
5. HOW to CALCULATE proper and accurate weights?
6. HOW to APPLY weight to cases, tapes, and statistics?
7. HOW to apply weights in FORMULAS and software?

Some kind of weighting is frequently involved in the analysis of many survey reports, and ad hoc explanations appear sometimes, usually hidden in appendices behind project reports. On the other hand, we can also find articles with theoretical discussions that are concentrated only on some single specific aspect of weighting, such as stratification, or post-stratification, or nonresponses, or variance reductions. We can also encounter misleading statements, such as in Fig 2. Among these are some theoretical discussions, based on diverse models, which are opposed to weighting (as in Fig 2B) [discussed by DuMouchel and Duncan 1983]. But I must avoid these arguments in this simple and brief, but general and useful treatment. In order to satisfy those two criteria, to be both simple and general, I had to forsake any attempt at profundity and precision. (Anybody who tries to satisfy all three criteria of simplicity, generality, and profundity is bound to fail, probably on all three, I believe.) We may expect and welcome contributions to overcome my inadequacies.

Fig 2. COMMON MISLEADING STATEMENTS

a. Weighting survey data is usually a SIMPLE process: each response weight ed inversely to its probability of selection, with adjustment for non-response.
b. I cannot find any justification for weighting in sampling theory, except for nonresponse.
c. It is best to select a simple random sample in order to obtain a self-weighting sample.
d. It is unethical to weight sample cases, because the process can be misused to produce biased results.

*Several colleagues were helpful, especially Drs. Paul Flyer and Frank Potter.

2. Self-Weighting Samples

Self-weighting samples are often preferred, because they possess considerable advantages in simplicity, in reduced variances, and in robustness. Statistical theory, from the lowest to the highest, overwhelmingly assumes self-weighting samples in one form or another. Furthermore, selection of elements with equal probabilities, EPSEM, often seems desirable and reasonable when the survey variables are evenly distributed over the population. Voting by all adults springs readily to mind, but there are other behaviors, attitudes, and opinions which are also democratically, evenly distributed, or at least roughly so. In those situations, of the four main sources I list for weights (1 through 4 in Fig 3), unequal allocation (no. 1) for strata and domains may be avoided. With luck and with skill one may also circumvent frame problems (no. 2). Then if nonresponses are small, number 3 may be unnecessary. And often adjustments (no. 4) may not be needed. Hence self-weighting EPSEM samples occur commonly in surveys.

Sources 1 and 2 refer to EPSEM selections, but sources 3 and 4 concern other reasons for weighting. But we shall also present reasons against weighting in Section 5, even for non-epsem selections. Thus EPSEM selections are neither necessary nor sufficient grounds for self-weighting samples, though the two are strongly related, and also confused with each other.

The "robustness" of analysis with self-weighting samples deserves primary emphasis [Kish 1977]. And achieving EPSEM samples during the selection operations is a fundamental skill in the art of survey sampling. This includes complex multistage selections with probabilities proportional (directly and inversely) to measures of sizes. Often it requires clever handling of imperfections in the sampling frame. After selection, achieving acceptable response rates often needs skillful and devoted care also.

Fig. 3 REASONS FOR WEIGHTING

A - Inverse of Selection Probabilities
B - Reduce Selection Bias
C - Reduce Variances
D - Change Population Base

1. DISPROP. ALLOCATIONS — A
   "OPTIMAL" allocations to reduce variances in stratification OVER-SAMPLING of small domains in multipurpose design MULTIPHASE sampling for rare elements
2. FRAME INEQUALITIES — A
   Sampling units of variable sizes: house holds, organizations, institutions of variable sizes
   Replicate listings
3. NONRESPONSES — AB
   Total nonresponses - refusals, not-at-homes
   Differential callbacks
   Item nonresponses
4. STATS'AL ADJUSTMENTS — ABCD
   Post-stratification, ratio and regression estimates
   Raking estimates, iterated fitting, rim weighting, SPREE
   "Optimal" weights $W_i=(1/\sigma_i^2)/(\Sigma 1/\sigma_i^2)$
5. COMBINING SAMPLES — D
   Diverse populations
   Cumulating periodic surveys
   Meta-analysis
6. CONTROL STATISTICS — B
   Random or biased? Total or small cells?
7. NONPROBABILITY SAMPLES — D
   Controlling with check data
   Similarities with quota sampling

3. Reasons for Weighting

Beyond the first four principal and common sources for weighting, note three more sources, which are less general, and refer to special situations. I distinguish these seven separate sources of weights because they usually have very different effects, and also because they need different strategies and treatments.

1. Disproportionate sampling fractions can be introduced deliberately to decrease either variances or costs, often with "optimal" allocations in strata. They are also used to produce larger samples for separate domains, usually for smaller domains. These deliberate differences in the sampling fractions f_h should be large to be effective, by factors from 2 up to 10, but seldom to 100 [Kish 1987, 4.5; 1965, 11.7]. It is often convenient to make the f_h simple and integral multiples (like 2f or 10f) of a basic sampling rate f. They must be compensated with inverse weights (e.g. 2 or 10) in order to avoid bad biases in combined statistics.

2. Inequalities in sampling frames must be measured, if they have not been corrected during the selection. If they affect only a small portion of the sample or if they differ only by a small factor, their total effect may be measured and then considered to be negligible, if either the portion or the factor is small. Otherwise they should be compensated with weights inversely proportional to selection probabilities; but usually these weights tend not to be large, and not for a large portion of the sample [Kish 1965, 2.7]. Weights due to sources 1 and 2 can be called "inverse Π" because they merely compensate for unequal selection probabilities Π_i.

3. Nonresponses present different problems than types 1 and 2, and compensating for them would not be entirely "inverse Π" weighting, except to the degree that the nonresponses can be assumed to arise as random selections from the sample cases. Nonresponse weights involve models or assumptions of some kind, implicit or explicit. Attempts to approximate those assumptions involve differential weights for cells formed with auxiliary variables -- e.g. geography, age, sex etc -- that are both available and hopefully related to survey variables. It is difficult and rare to obtain data either from the sample or from the check statistics that closely satisfy both requirements. Fortunately, with small nonresponses the differences between subclasses tend to be small and to result in small effects. For item nonresponses compensations seem more justified and more common, and they are usually made with imputation (duplication) of responses [Kalton 1983]. Corrections for noncoverage are much more difficult, because coverage rates cannot be obtained from the sample itself.

4. Statistical adjustments can be made with post-stratification weighting, or with ratio or regression estimators; and the technical literature concentrates on reducing variances with controls that were not used in the process of selection. In practice, however, these controls and methods are more important for reducing the biases of nonresponse and especially of noncoverage; and this is done with differential weighting up to some available check-data [USCB 1978, Ch V; Kish 1987, 4.7]. These are akin in principle to the methods of standardization that are discussed under source 6. The data and the software must both be appraised for integrity.

With multiple classifications, the weights for separate adjustment cells might become both unstable (small n) and unknown (lacking data); but multiple marginal statistics may be used with iterated estimators, which are also called raking, marginal, rim, or SPREE. Modern computers and growing bodies of data files make these methods increasingly useful.

The weighting of cases proportional to their precision ($1/\sigma_i^2$) appears in the literature, but this has more theoretical appeal than practical applicability [Kalton 1968].

5. Combining samples is becoming more popular, more important, and more feasible because of increasing numbers of samples that are available for combinations. All combinations concern weighting in some form, and one should always be explicit about the weights and careful about possible differences in measurements. We note also that any national sample combines diverse domains, some like provinces, some like diverse social or demographic classes; and all those domains differ in the distributions of the survey variables. Nowadays, one may also combine standardized national samples from several countries, e.g. African birth rates from separate national samples of the World Fertility Surveys. Similarly to spatial integration, we may also combine periodic samples into rolling samples integrated over a longer time span; e.g. annual averages of influenza, or cancer rates, or unemployment, or incomes from weekly or monthly surveys. Meta-analysis is a growing field for combining statistics, and already foreshadowed in 1924 by Yates and Cochran [1938]. A special and simple method of combining can be the cumulations of individual cases [Kish 1987, 6.6].

6. Adjustments to match controls can have a variety of motivations. Whereas the reasons under 4 for post-stratification and ratio estimation concerned mostly sampling variations, here we refer mostly to adjustments of samples from one frame population to some other target (standard) population(s). For example, a sample from one province (state) may be reweighted to the national population [Kish 1987, 4.5]. Or we may reweight samples from one country or period to another country or period. Generally, the subclasses of the sample are reweighted to the domains of the target population, and these controls must be available both for the sample and for the target population. If there are too many cells, the control data may be unavailable and the sample cases too few for stability, and then marginal adjustments may be used, with iterated fitting.

7. Adjustments for nonprobability samples to check data may be viewed as a special case of the above, to the extent that a population "frame" may be envisioned for the sample, even if not operationally defined. Theoretically this is similar to the motivation for "quota" sampling, but with weighting substituted for the selection. The nonprobability selections use quotas that are constrained to fit the population proportions within cells (by age, sex, urban, occupation). With several variables the fitting can be marginal.

4. Basic Methods for Weighting

Four main procedures for weighting need individual attention because they require different techniques and also because they can have different effects on the variances (Fig 4).

1. Individual case weights (ICW) yield the most common, simple, practical, and flexible procedures, especially with modern computers and programs that can handle them. The other procedures may be compared with and based on ICW, and they tend to increase variances more than ICW. The weights w_j for sample elements ($j = 1,...n$) may reflect a product $p_j r_j$ of the element probabilities p_j from complex

statistical multistage selections with the response rates r_j, which may also include coverage rates. The weights $w_j = 1/p_j r_j$ are inversely proportional to these products. Both p_j and r_j should be available for all elements of probability samples.

Mean statistics must be "normalized" (standardized) with the sum of weights Σw_j, as in $\bar{y}_w = \Sigma w_j y_j / \Sigma w_j$, also in $\Sigma w_j y_j^2 / \Sigma w_j$, and in $\Sigma w_j y_j x_j / \Sigma w_j$. Because of this normalization the weights may be any nonnegative numbers w_j' proportional to the "true" expansion weights; e.g. $w_j' = f w_j$. For EPSEM selections $w_j = 1/f = N/n$ and $\Sigma w_j = N$; whereas $w_j' = f/f = 1$ and $\Sigma w_j' = n$. The simple expansion totals $\hat{Y} = \Sigma w_j y_j$ are seldom used in practice, only in literature.

For means and complex statistics this weighting produces "consistent" (not strictly unbiased) estimators. But their variances can be increased by the unequal weights, as noted in 5.2 below.

If we use standardized relative case weights, $W_j = w_j / \Sigma w_j$, then $\bar{y}_w = \Sigma W_j y_j$ simply. However a single standardizer will not work comfortably for the many subclasses of complex survey analyses, because the bases change.

Fig. 4 BASIC METHODS FOR WEIGHTING

1. INDIVIDUAL CASE WEIGHTS w_j on the data tape, and used for all statistics, is the most common and practical procedure, especially with modern computers. The weights may reflect a product of probabilities P_j from complex multistage selections and response rates r_j in $w_j = 1/P_j r_j$. Mean statistics must be "normalized" with the sum of weights as in $\bar{y}_w = \Sigma w_j y_j / \Sigma w_j$ and in $\Sigma w_j y_j x_j / \Sigma w_j$ and $\Sigma w_j y_j^2 / \Sigma w_j$.
2. WEIGHTED STATISTICS, e.g. $\Sigma W_h \bar{y}_h$, may be preferred over individual weights, for combining a few strata, or classes, or populations; especially for self-weighted statistics $\bar{y}_h$ within them; and especially for simple statistics like means or totals.
3. REPLICATION (DUPLICATION) may be used to prepare self-weighting tapes. This may be more convenient than weights, especially for item non responses; also for complex, analytical statistics. It increases variances over individual weights, but those increases can be almost eliminated with "multiple replications." But don't count the number of cards m, only the number of genuine cases n, without the m-n duplicates. Even better $n' = n/(1+L)$.
4. ELIMINATION of cases is seldom desirable. But it increases variances only little more than duplication for small percentages. Combinations of replication and replication in different calls are possible.

Example of 4.1

W_i	N_h	f_h	r_h	n_h	w_j	w_j	$w_j n_j$
.900	90,000	.01	.90	810	.009	111.111	90,000
.09	9,000	.10	.85	765	.085	11.765	9,000
.01	1,000	1.0	.8	800	.80	1.25	1,000
1.00	100,000 = N					$\Sigma w_j n_j$ =	100,000

The sampling fractions n_h/N_h in domains h can be used for $w_j = N_j/n_j$, and this is justifiable when the elements j are selected with actual equal probabilities within the domains h. However, it is misleading (or misled) to confuse a mere sampling fraction with a sampling probability; e.g. that a sample of n_h was selected from a population of N_h may perhaps be called a sampling fraction of n_h/N_h but does not mean a sampling probability of $p_h = n_h/N_h$. Probabilities of selection must be justified with probability operations. Otherwise we are faced with judgment samples or model dependent sampling. On the other hand, in other situations the selection probability f_h may be applied without finding appropriate, unbiased, and dependable values of N_h for the population.

2. Weighted statistics, e.g. $\bar{y}_w = \Sigma W_h \bar{y}_h$, combine separate subpopulation statistics $\bar{y}_h$ with appropriate relative weights W_h, with $\Sigma W_h = 1$. This method may be preferred over ICW for: a) combining published statistics when individual cases are not available; b) combining a few strata based on disparate selection procedures; and for c) relatively simple statistics, like means or totals. But they are not as useful for complex analyses of single surveys. Dependable weights W_h are needed from justifiable sources. These can also be used for ICW as above.

3. Duplication of cases may be used instead of ICW in order to prepare self-weighting tapes for convenience in some situations. It is especially convenient for item nonresponses. Also for complex analytical statistics. Some compromise between random selection and "closest" matching to reduce bias is generally used for duplication within subclass cells. If the response rate is r_h in cell h, $(1\text{-}r_h)$ cases can be duplicated to fabricate $(1\text{-}r_h)$ pseudo cases; either randomly selecting with probability $(1\text{-}r_h)/r_h$ or by finding the "closest" matching $(1\text{-}r_h)$ fraction of cases. Duplication increases variances over individual weighting, but those increases are not great for duplicating only a small proportion of the samples. Furthermore these increases of variances can be almost eliminated with procedures of "multiple replications" [Kalton 1983].

We must caution against the embarrassing mistake of accepting from the computing programs the tape counts (or card counts) m, which contain (m-n) replicates as well as n genuine cases. These n genuine cases can be "tagged" for counting. But the "effective number" may be further diminished by duplication to $n' = n/(1+L)$, as noted in 6.2.

4. Elimination of cases can be justified in some situations, although throwing away information may appear statistically criminal. Nevertheless, consider three justifiable situations. a) Large samples have been selected with different sampling rates for a nation's several provinces; then an epsem sample is designated for complex national analysis, with rates suited to the lowest provincial rates. b) A small domain has been greatly oversampled for separate analysis, but a proportionate sample has been "tagged" from it for joint complex analysis, which could be difficult and not much more precise with the extra cases from the small domain. c) Eliminating a small proportion of cases (say $<.05$) increases the variance only little more than duplication of a similar proportion. This counterintuitive result has been used for compromise adjustment for differential nonresponses between strata [Kish 1965, 11.7B].

5. Reasons Against Weighting

1. Complications often arise from weighting, even when good computing programs are available (Fig 5). Some are due to mistakes in the man-machine sys-

Fig. 5 REASONS AGAINST WEIGHTING

1. COMPLICATIONS usually arise from weighting; less from simple aggregates and means, but more from complex, analytical and inferential statistics. Complications lead to mistakes.
2. INCREASED VARIANCES result from random (haphazard unplanned) weighting. The RELATIVE increases can be estimated with $(\Sigma W_h k_h)(\Sigma W_h / k_h) > 1$ with stratum sizes W_h and relative weights k_h. Or with $(1+C_k^2)$, the relvariance of the relative weights in the sample. These increases tend to persist for all statistics, as if increasing the element variance σ^2 or decreasing the effective sample size.
3. SMALL BIASES may be compared to increased variances in the MSE = $S^2+B^2=S^2(1+B^2/S^2)$. The "bias ratio" = B/S, and the B and S are specific to and differ greatly between statistics. Nevertheless a uniform decision about weighting is usually more convenient. The bias ratio for means $(\bar{y}-\bar{y}_w)/ste(y)$ may be used.
4. "MODEL DEPENDENT" theoretical arguments exist, such as 2B.
5. PUBLIC RELATIONS or ETHICS may also hinder weighting; the combined mean $y_w = \Sigma W_h Y_h$ can be made to approach any of the components y_h with extreme weights w_h.

tem and these tend to increase for more complex analyses. Furthermore, for complex, analytical statistics, and for inferential statistics, such as tests of significance, adequate theory may not be available for weighted estimators or for their sampling errors.

2. Increased variances can result from weighting that arises from random, or haphazard (or irregular) differences in selection probabilities (when these are not "optimal" at all). For example, the inequalities due to frame problems or to nonresponses are generally of this kind. Furthermore, these increases of variances (unlike those due to clustering) tend to persist undiminished for subclasses and for all statistics, as if they were to increase the element variances from σ^2 to $(1+L)\sigma^2$, or to decrease the number of elements from n to $n/(1+L)$.

These relative variances can be estimated in the design stage with $(1+L) = (\Sigma W_h k_h)(\Sigma W_h/k_h)$, where W_h are stratum sizes in the population with weights k_h; for k_h we may also use their inverses $1/k_h$ or proportionate numbers ck_h or c/k_h, because of their inverted appearances. From $n = \Sigma n_j$ sample cases with ICW weights k_j we can compute $(1+L) = \Sigma n_j \Sigma n_j k_j^2/(\Sigma n_j k_j)^2$, or from n_h cases in weight classes h, compute $(1+L) = \Sigma n_h \Sigma n_h k_h^2/(\Sigma n_h k_h)^2$. The relative variance increase (or loss) L may be viewed as the relvariance cv^2 = variance/mean^2 of the relative weights $n_j k_j$; because $(1+L) - (\Sigma n_j k_j)^2/(\Sigma n_j k_j)^2 = cv^2$. Thus the factor $(1+cv^2) = (1+L)$ depends on the relative variances of case weights. It serves as a good precautionary measure to print out a cumulative distribution of (relative) weights, in order to appraise possible increases of variances from using haphazard weights.

3. Comparisons of biases B^2 with variances S^2 may reveal that B^2 is smaller than S^2 in the MSE = $S^2 + B^2 = S^2(1+B^2/S^2)$. The bias ratio B/S due to weighting may be estimated from the differences $(\bar{y} - \bar{y}_w)/ste(\bar{y})$ between weighted and unweighted means from the same survey. The bias ratios B/S (and B^2/S^2, and B and S) are specific to all statistics and can differ greatly between various statistics of the same survey (most of which are multipurpose). Note especially that for subclasses the S^2 are

(much) higher, and therefore the bias ratios B^2/S^2 are much lower [Kish 1987, 2.3].

Nonetheless a uniform decision about weighting is often judged more convenient. We should compare the biased unweighted MSE = $S^2(1+B^2/S^2)$ with the weighted MSE = $(1+L)S^2$ without that bias. Curtailing (trimming) the extreme weights, particularly the large weights, may be used to bring about a reasonable compromise solution. This may be found from the cumulated (relative) weights. Where and how to trim? We need theory and methods, but they have been hard to find until recently [Potter 1989, 1990].

4. Model dependent arguments have been advanced by "modellers" against correction for selection biases as unneeded (as in 2b).

5. Public relations or ethics may also hinder overt and differential weighting, because it is possible to misuse it to produce subjectively desired, prejudiced results [Sharot 1986]. Explicit weights expose data to a public screening that biased selection often escapes (alas). For example, the combined mean $\bar{y}_w = \Sigma W_h \bar{y}_h$ could be made to approach any of the components $\bar{y}_h$ with extreme weights W_h. (Journalists, alas, do this commonly, by using the cost of either automobiles and TV sets, or housing and health care as indices of the cost of living.)

6. Diverse Effects for Different Statistics

Users must be warned about these differences, because many may be misled by the phrasing of "THE Bias." The sizes of biases and their bias ratios, B/S, vary a great deal, depending on the variables, also on the subclasses used for bases, and also on the statistical function being estimated (e.g. B/S for medians and means may differ).

1. Expansion totals $\hat{Y} = y/f$ are most sensitive to biases from weights; for example, even uniform and random nonresponses can result in bad underestimates, if not adjusted. Also expansions like $\hat{Y} = \Sigma N_h(y_h/n_h)$ can be very sensitive to biases in the borrowed values of the N_h. But differences or ratios of totals from periodic studies would be less sensitive.

2. Means are usually less affected than totals. Sample surveys survive the terrible nonresponse rates now prevailing in the USA only because nonrespondents do not differ much from responses for most survey variables. Large biases result only from combinations of differences in both weights and survey variables within subclasses. If either of these is uniform over subclasses the net bias tends to be small.

3. For subclass means the variances increase in proportion to the decrease of the sample bases. Then the bias ratios decrease with the decreasing sample bias. For differences between subclass means the biases often tend in the same direction and the bias ratios are often reduced drastically [Kish 1987, 2.7].

4. Analytical statistics, e.g. regression coefficients, pose computational, methodological, and philosophical problems for weighted estimates. I find it impossible to make a general statement, other than skepticism about the kind of avoidance expressed in Fig 2B.

5. Sampling errors, inferential statistics, tests of significance can also pose severe problems of computation, methodology and interpretation for weighted estimates.

7. Conclusion: To Weight or Not to Weight?

Suppose that we are not so prejudiced that we would always weight in order to preserve unbiasedness against unequal selection probabilities, especially against nonresponses and frame problems. Nor are we such confident modellers that we never see reasons for weighting. There are statisticians who preach both extremes, and many who practice them, alas. But I believe in balance, in tradeoffs. We may remember that for weighted estimates the reduction in bias may also bring increases in the variances due to weighting, hence $S^2(1+B^2/S^2)$ must be balanced against $(1+L)S^2$.

Suppose then that we wish to balance biases against the increases of the variances of estimates, probably in attempts to reduce mean-square errors as chief criterion. (But one may adopt other formulations for reducing the errors of our inferential statements.)

These balancing relations differ between variables because B^2 and S^2 differ greatly for different variables and for various statistics. The bias can be less important for subclasses than for the overall sample, and for comparisons of subclass means much less important. However a consistent, uniform decision may be needed for the analysis. And a compromise with curtailing or trimming may be best.

Finally, it is clear that this panoramic view raises questions that I cannot answer, but I hope to have stimulated some others to search for answers.

Fig. 6 - DIVERSE EFFECTS FOR DIFFERENT STATISTICS

1. EXPANSION TOTALS $\hat{Y}=y/f$ are most sensitive to biases of weights; for example even modest uniform and random nonresponses can result in bad underestimates, if unadjusted. Differences $\hat{Y}_m - \hat{Y}_n$ or ratios $\hat{Y}_m/\hat{Y}_n$, in periodic studies, would be less affected. Similarly for stratum weights.
2. MEANS y seem to be less affected; sample surveys survive the prevailing terrible nonresponse rates only because these do not differ greatly from responses for most survey variables. Also for weights: large biases result only from combinations of differences in both weights and survey variables within subclasses; if either of these is uniform over subclasses the net bias tends to be small; especially compared to variances of small samples.
3. DIFFERENCES $(\bar{y}_c - \bar{y}_b)$ between small subclasses tend to be less affected than the means.
4. ANALYTICAL STATISTICS, e.g. regression coefficients, pose computational, methodological, and philosophical problems for weighted estimates.
5. SAMPLING ERRORS AND INFERENTIAL STATISTICS for weighted estimates pose similar difficulties.

References

DuMouchel WH and Duncan GS [1983]. Using sample survey weights in multiple regression analyses of stratified samples, Jour. Am. Stat. Assn 78, 535-43.

Flyer P, Rust K and Morganstein D [1989], Complex survey estimators and contingency table analysis using replication, Proc. Survey Research Section, Am. Stat. Assn.

Kalton G [1968], Standardization: A technique to control for extraneous variables, Applied Statistics, 17, 118-136.

Kalton G [1983], Compensating for Missing Survey Data, Ann Arbor: Institute for Social Research.

Kish L [1965], Survey Sampling, New York: Wiley, 11.7.

Kish L [1977], Robustness in survey sampling, Bulletin of the International Statistical Institute, 47 (3), 515-28.

Kish L [1987], Statistical Design for Research, New York: Wiley, 7.4.

Kish L [1989], Sampling Methods for Agricultural Surveys, Rome: FAO, 12.5-12.6. Available FREE by writing to Statistical Division, FAO, I-00100 Roma, Italy.

Potter FJ [1989], Survey of procedures to control extreme sampling weights, pp. 453-58, Proceedings of the Section on Survey Research Methods, ASA.

Potter FJ [1990], A study of procedures to identify and trim extreme sampling weights, Proceedings of the Section on Survey Research Methods, ASA.

Sharot T [1986], Weighting survey results, Jour. of Market Research Society, 1986, 269-84.

US Census Bureau [1978], The Current Population Survey: Design and Methodology, Technical Report 40, Ch V

Yates F and Cochran WG [1938], The analysis of groups of experiments, Journal of Agricultural Science, 28, 556-80.