DESIGN EFFECTS AND SURVEY PLANNING

Inho Park, Marianne Winglee, Jay Clark, Keith Rust, Andrea Sedlak, David Morganstein
Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Design Effect Decomposition Models; Stratification; Unequal Weighting; Clustering

1. Introduction

The design effects of survey estimates can be used as tools for measuring sample efficiency and for survey planning. Kish (1965) defined the design effect as the ratio of the variance of an estimate under the complex sample design to the variance of the same estimate that would apply with a simple random sample (SRS) of the same size. Complex sample designs typically involve a number of design features, such as stratification, clustering, and unequal weighting. The efficiency of the complex sample design can be evaluated for each design feature through decompositions of design effects. If the loss of precision for survey estimates due to variable weights is found to be notable, then we may review the reasons for the variable weights and consider options to reduce that variation. If the design effect from clustering is very large for some survey estimates, then we may also consider options to reduce its effect. If the gain in precision of estimates due to stratification is negligible, then we may further enhance the current stratification scheme to obtain greater benefit from the stratification of the design. The ultimate goal is to lower design effects of key estimates and maintain sample precision with smaller sample size.

This paper discusses a method of design effect decomposition analyses applicable to data from multistage stratified samples. Section 2 reviews Kish's two-factor design effect decomposition model (Kish, 1987) and discusses an option for a 3-factor model that takes into consideration the relative contribution from implicit stratification. Section 3 discusses a decomposition model developed for the Third National Incidence Study of Child Abuse and Neglect (NIS-3). The key statistics of interest for the NIS are estimates of total countable children. Section 4 discusses the method to compute the design effects for estimates of population totals. Section 5 describes the results of our estimation for ten key NIS-3 estimates. Section 6 presents our discussion.

2. Design Effect Decomposition Models

2.1 Kish's Production Model

Models for design effect decomposition are discussed in Kish (1965), Verma, Scott, O'Muircheartaigh (1980), and Lê, Brick, and Kalton (2001). A direct approach to stating the overall design effect is relatively complicated under a complex sample design with more than one complex feature. As an indirect approach, Kish (1987) proposed a production model of the overall design effect as a function of independent components associated with clustering and unequal weighting. This approach is well-accepted by many survey samplers. For example, Gabler, Haeder, and Lahiri (1999) provided a justification for Kish's model in their sample size determination in the European Social Surveys.

For the sample proportion $\hat{P}$, Kish's 2-factor production model is given as

$$deft^2\left(\hat{P}\right) = deft_c^2\left(\hat{P}\right) \times deft_w^2\left(\hat{P}\right), \qquad (2\text{-}1)$$

where $deft_c^2\left(\hat{P}\right)$ and $deft_w^2\left(\hat{P}\right)$ denote the design effect components for clustering and unequal weighting. These components can be evaluated as follows:

$$deft_c^2\left(\hat{P}\right) = 1 + \left(\bar{n} - 1\right)\rho\,, \qquad (2\text{-}2a)$$

and

$$deft_w^2\left(\hat{P}\right) = 1 + cv_w^2, \qquad (2\text{-}3)$$

where $\bar{n} = \sum_{i=1}^{a} n_i \,/\, a$ is the average cluster sample size, a is the number of sampled clusters, n_i is the cluster sample size for the i-th cluster, ρ is the intracluster correlation coefficient, and cv_w^2 is the relative variance of the sample weights. The intracluster correlation ρ is a measure of homogeneity among elements within PSUs.

Model (2-2a) for the effect of clustering is intended for self-weighting samples. When there are large cluster sample size variations, model (2-2a) tends to underestimate the clustering effect. Holt (1980) derived the following model for unequal size clusters:

$$deft_c^2\left(\hat{P}\right) = 1 + \left(\bar{n}' - 1\right)\rho'\,, \qquad (2\text{-}2b)$$

where $\bar{n}' = \sum_{i=1}^{a} n_i^2 \Big/ \sum_{i=1}^{a} n_i$ is an adjusted (or weighted) average of cluster sample sizes n_i. The ratio $(\bar{n}'-1)/(\bar{n}-1)$ is the inflation factor for the clustering effect due to the variable cluster sample sizes. Model (2-2b) leads to an estimate of ρ' that is smaller than that provided by model (2-2a) by the inverse of the above rate. The intra-cluster correlation ρ' is estimated under model (2-2b) as

$$\hat{\rho}' = \frac{deft_c^2(\hat{P})-1}{\bar{n}'-1}. \qquad (2\text{-}4)$$

Model (2-3) specifies the design effect associated with the departure from self-weighting due to weight adjustments that are independent of the sample variation of the survey characteristic (Kish, 1965, 1992).

2.2 Extension of Kish's Production Model

Multistage sample designs often include a stratified selection of PSUs. Kish's model (2-1) is applicable under some special sample designs with stratified selection of PSUs. For example, Kalton (1979) showed that model (2-2a) can approximate the design effect due to clustering for an overall self-weighting sample with a proportionate stratified selection of PSUs. In an effort to factor out explicitly the design effect component due to stratification, we attempted an extension of Kish's production model as follows:

$$deft^2\left(\hat{P}\right) = deft_s^2\left(\hat{P}\right) \times deft_c^2\left(\hat{P}\right) \times deft_w^2\left(\hat{P}\right) \qquad (2\text{-}5)$$

where $deft_s^2(\hat{P})$ is the design effect component incorporating the stratification effect, $deft_c^2\left(\hat{P}\right)$ is given in (2-2a) or (2-2b), and $deft_w^2\left(\hat{P}\right)$ is given in (2-3). Each component is positive and is expected to be around one when there is neither gain nor loss in sample efficiency as compared to simple random sampling with replacement (SRSWR). Under this model, the increase in the design effect due to clustering and unequal weighting can be compensated by the gain in precision due to efficient stratification.

3. Decomposition Model for NIS-3

This section discusses a method to model for the design effect component $deft_s^2\left(\hat{P}\right)$ for the NIS-3 sample design.

3.1 NIS-3 Sampling and Variance Estimation

The Third National Incidence Study of Child Abuse and Neglect (NIS-3) is based on a complex multistage and multiframe design. In this paper, we focus only on the selection of a list sample from the Child Protective Services (CPS) agency frame. Very briefly, the NIS-3 CPS sample used a multistage probability sample design with complex features including stratification, clustering, and unequal weighting. The first stage sampling units were primary sampling units (PSUs) selected through implicitly stratified probability sampling. The NIS-3 PSUs were single counties or adjacent small county groupings. Prior to sample selection, the PSU frame was sorted by the number of counties within PSUs, by census region, and by the degree of urbanicity. A sample of 40 PSUs was then systematically selected from the presorted PSU frame with probability proportional to size (the child population aged 0-17 in the 1990 Census). Two self-representing (SR) PSUs were selected with certainty and the remaining 38 were nonself-representing (NSR) PSUs.

Variance estimation for the NIS-3 used the jackknife replication method known as JK2 (Westat, 2001), where PSUs and their consecutive pairs are the variance units and variance strata, respectively. The JK2 method assumes that PSUs are grouped into variance strata and that within each stratum two PSUs are selected as variance units, that is, a two-unit-per-stratum design. A total of 21 variance strata were defined, 19 of which were formed by pairing the 38 NSR PSUs in their order of selection from the PSU frame. Two strata were formed for the two SR PSUs, providing no between-PSU component of variation, only a within-PSU component of variation. The between-PSU variance component is reflected only in the 19 NSR PSU strata. We used the variation among the 38 PSUs to indicate the efficiency of implicit stratification in the PSU selection along with the framework of the JK2 variance estimation method.

3.2 Modeling Component for Stratification Effect

This section presents a design effect component model to include the effect of implicit stratification. One approach is to interpret the stratification effect as the efficiency of the adopted systematic PPS selection of PSUs under the JK2 method, as compared to a naïve PS selection (that is, systematic PPS selection with a random sort order) of PSUs.

We denote N as the total number of children countable as abused or neglected under the endangerment definitional standard and denote Y as the total number of children possessing attribute y (e.g., physically abused) among N total countable children.

We also denote w_k as the sample weight and y_k as an indicator variable of attribute *y* for a sampled child *k* (that is, $y_k = 1$ if child *k* has attribute and $y_k = 0$ if child *k* does not have attribute). A customary estimate of Y is the weighted sample total given as

$$\hat{Y} = \sum_{k=1}^{n} w_k y_k \ . \qquad (3\text{-}1)$$

The population proportion $P = Y / N$ is then estimated by the ratio $\hat{P} = \hat{Y} / \hat{N}$, where $\hat{N} = \sum_{k=1}^{n} w_k$ is an estimate of *N*. Note here that $\hat{P}$ is the proportion of countable children with a specific attribute (not the proportion of children with attribute). The summations are over n children in the sample who were evaluatively coded as being abused or neglected under the endangerment countability standard.

Using the Taylor approximation, we can rewrite $\hat{P}$ as a linear sum:

$$\hat{P} - P \quad \sum_{k=1}^{n} w_k z_k \qquad (3\text{-}2)$$

where $z_k = (y_k - P) / N$. Assume that the 38 NSR PSUs are listed in the sort order of the PSU frame. Letting $w_{k|i}$ denote within-PSU weights, the estimated PSU totals of $\hat{z}_k$ for NSR PSU *i* are $\hat{Z}_i = \sum_k w_{k|i} \hat{z}_k$, where $\hat{z}_k = (y_k - \hat{P}) / \hat{N}$ are sample estimates of z_k and the summation is over *k* within PSU i. Letting p_i denote the PSU selection probabilities and w_i the associated weights, expression (3-2) can be rewritten as $\hat{P} - P \quad \sum_{i=1}^{38} w_i \hat{Z}_i$. Assume that the $\hat{Z}_i$ are fixed quantities. If two $\hat{Z}_i$ were selected with unequal probabilities, p_i, from within each of 19 equal-sized strata, then the variance of $\hat{P}$ is estimated by

$$v_{stpps}(\hat{P}) = \frac{1}{38^2} \sum_{i=1}^{19} \left(\frac{\hat{Z}_{2i-1}}{p_{2i-1}} - \frac{\hat{Z}_{2i}}{p_{2i}} \right)^2 .$$

If the $\hat{Z}_i$ were selected with unequal probabilities, p_i, with replacement, then the variance of $\hat{P}$ is estimated by

$$v_{pps}(\hat{P}) = \frac{1}{38 \times 37} \sum_{i=1}^{38} \left(\frac{\hat{Z}_i}{p_i} - \hat{Z} \right)^2 ,$$

where $\hat{Z} = \sum_{k=1}^{n} w_k \hat{z}_k$. (See, for example, Wolter 1985, pp. 287.) The subscripts *STPPS* and *PPS* represent the variances estimated under the sample designs with implicitly stratified and naïve PPS, respectively. We used the ratio

$$deft_s^2(\hat{P}) = \frac{v_{stpps}(\hat{P})}{v_{pps}(\hat{P})} \qquad (3\text{-}3)$$

as a measure of the efficiency of the implicit stratification scheme with the JK2 variance estimation method. To calculate $deft_s^2(\hat{P})$, we let $\hat{N}_i = \sum_k w_{k|i}$ and $\hat{Y}_i = \sum_k w_{k|i} y_k$ denote estimated total numbers of countable children and those with attribute *y* within PSU *i*, respectively, and let $\hat{P}_i = \hat{Y}_i \big/ \hat{N}_i$ denote the corresponding PSU proportion estimates. We can rewrite (3-3) approximately as

$$1 - deft_s^2(\hat{P}) \quad \frac{2\sum_{i=1}^{19} \hat{\hat{N}}_{2i-1} \left(\hat{P}_{2i-1} - \hat{P} \right) \hat{\hat{N}}_{2i} \left(\hat{P}_{2i} - \hat{P} \right)}{\sum_{i=1}^{38} \left\{ \hat{\hat{N}}_i \left(\hat{P}_i - \hat{P} \right) \right\}^2} \qquad (3\text{-}4)$$

where $\hat{\hat{N}}_i = \hat{N}_i / p_i$ is the weighted total number of countable children based on the ith PSU sample. If PSU proportion estimates, $\hat{P}_{2i-1}$ and $\hat{P}_{2i}$ are close to each other (both greater than or less than $\hat{P}$ together), then the right hand side of expression (3-4) can be positive and $deft_s^2(\hat{P})$ can be less than 1. For the opposite case (when $\hat{P}_{2i-1}$ and $\hat{P}_{2i}$ are opposite relative to $\hat{P}$), the right hand side of expression (3-4) can be negative, and thus $deft_s^2(\hat{P})$ can be larger than 1. That is, if the implicit stratified selection of the PSU sample leads to consecutive PSU pairs in the sample being more homogeneous, the ratio $deft_s^2(\hat{P})$ should be less than 1, indicating a gain due to implicit stratification. If the ratio is greater than 1, then implicit stratification did not improve sample efficiency. The magnitude of $deft_s^2(\hat{P})$ is determined by the quantities $\hat{\hat{N}}_i$ and $\hat{P}_i - \hat{P}$.

4. Design Effects for Estimates of Totals

The design effect models discussed in Section 2.1 and Section 3.2 were defined for estimates of population means and proportions. While the same concept can be applied to other statistics, the magnitude of design effects for different statistics can vary. For example, Barron and Finch (1978), in their study for the Survey of Low Income Aged and Disabled, observed that the design effects for $\hat{P}$, estimates of population proportions, are less variable than those for $\hat{Y}$, estimates of population totals. In this section, we compare the overall design effects for $\hat{P}$ and $\hat{Y}$.

Denote $r(\hat{P},\hat{N})$ as the correlation coefficient of $\hat{P}$ and $\hat{N}$ under the complex sample design. Using the Taylor approximation (3-2), the relationship between two overall design effects for $\hat{P}$ and $\hat{Y}$ under a complex sample survey is given as

$$deft^2(\hat{Y}) = deft^2(\hat{P}) + \Delta_y, \qquad (4\text{-}1)$$

where

$$\Delta_y = Q_y \times \left\{ Q_y + 2r(\hat{P},\hat{N})deft(\hat{P}) \right\}$$

is the difference between the design effects for $\hat{Y}$ and $\hat{P}$ and $Q_y = cv(\hat{N}) \big/ cv_{srswr}(\hat{P})$ is the ratio of the coefficients of variation of $\hat{N}$ under the complex sample design and of $\hat{P}$ under SRSWR, respectively (Park and Lee, 2003). Recalling that y_k is binary (0,1), we have $cv^2_{srswr}(\hat{P}) = cv^2_{srswr}(\hat{Y}) = (1-\hat{P})/(n\hat{P})$. In this situation, Q_y can be arbitrarily large as $\hat{P}$ approaches one and small as $\hat{P}$ approaches zero, assuming $cv(\hat{N}) \neq 0$.[1] This means that the difference Δ_y is most influenced by the magnitude of $\hat{P}$, although it is affected by the other factors $r(\hat{P},\hat{N})$ and $deft(\hat{P})$ to a certain degree. In general, $\hat{Y}$ has a larger design effect than $\hat{P}$, unless $Q_y < -2r(\hat{P},\hat{N})deft(\hat{P})$. See Park and Lee (2003) for more discussion.

For the stratification effect, the design effect component model $deft_s^2(\hat{Y})$ is analogous to model (3-3) for $\hat{P}$, defined by replacing $\hat{Z}$ and $\hat{Z}_i$ by $\hat{Y}$ and $\hat{Y}_i$, respectively. That is,

$$deft_s^2(\hat{Y}) = \frac{v_{stpps}(\hat{Y})}{v_{pps}(\hat{Y})}, \qquad (4\text{-}2)$$

from which we have

$$1 - deft_s^2(\hat{Y}) \quad \frac{2\sum_{i=1}^{19}\left(\hat{\hat{Y}}_{2i-1} - \hat{Y}\right)\left(\hat{\hat{Y}}_{2i} - \hat{Y}\right)}{\sum_{i=1}^{38}\left(\hat{\hat{Y}}_i - \hat{Y}\right)^2}, \qquad (4\text{-}3)$$

where $\hat{\hat{Y}}_i = \hat{Y}_i / p_i$ is the contribution from the estimated total number of countable children possessing attribute y based on the ith PSU sample. Given $\hat{\hat{N}}_i\left(\hat{P}_i - \hat{P}\right) = \hat{\hat{Y}}_i - \left(\hat{\hat{N}}_i / \hat{N}\right)\hat{Y}$, expression (3-4) indicates that $deft_s^2(\hat{P})$ is influenced by whether $\hat{\hat{Y}}_i$ and $\hat{\hat{N}}_i$ are close to $\hat{Y}$ and $\hat{N}$, respectively. In contrast, $deft_s^2(\hat{Y})$ depends only on $\hat{\hat{Y}}_i$ compared to $\hat{Y}$. We expect ratios $\hat{\hat{N}}_i / \hat{N}$ to be around one. Hence, $deft_s^2(\hat{Y})$ will not differ much in magnitude from $deft_s^2(\hat{P})$.

For the effect of unequal weighting, the same model (2-3) can be applied for both $\hat{P}$ and $\hat{Y}$, provided that the underlying assumptions of the model hold at least approximately. This is because the model assumption for expression (3-2) should hold also for expression (3-1). No good alternative is yet available except for the cases where the selection probabilities are correlated with a continuous characteristic of survey interest under a single-stage sample design with unequal selection probability (Spencer, 2000, and Park and Lee, 2001, 2002).

[1] When $cv(\hat{N}) = 0$ or $\hat{N} = N$, the two design effects are equivalent, that is, $deft^2(\hat{P}) = deft^2(\hat{Y})$.

For the clustering effect, Kish (1987) notices that model (2-2a) for $\hat{P}$ is not transferable to $\hat{Y}$. Särndal, Swensson, and Wretman (1992, Section 8.7) show that under a one-stage simple random cluster sample design, the difference between two design effects for $\hat{Y}$ and $\hat{P}$ can be quite large, provided cluster sizes N_i are variable. Park and Lee (2003) show that the differences under a two-stage unequal probability sample design can be attributable to the respective sources of variation from each sampling stage. The first stage contribution arises from the efficiency of selection probabilities as compared to the relative sizes of clusters. The second stage contribution arises from variations of estimates in different magnitudes, say $\mathrm{var}(\hat{Y}_i)$ and $\mathrm{var}(\hat{Y}_i - \hat{N}_i P)$. If PSU proportion estimates $\hat{P}_i = \hat{Y}_i / \hat{N}_i$ do not deviate much from the population proportion P, then $\hat{Y}_i$ tend to be more variable than $\hat{Y}_i - \hat{N}_i P = \hat{N}_i(\hat{P}_i - P)$, yielding $\mathrm{var}(\hat{Y}_i) \quad \mathrm{var}(\hat{Y}_i - \hat{N}_i P)$ in most cases. Thus, the second stage contribution can produce a larger design effect for $\hat{Y}$, unless the first stage sampling is dominant in determining the difference Δ_y and favors $\hat{P}$ to a larger extent. This indicates that, in general, $\hat{Y}$ tends to have a larger design effect than $\hat{P}$. The clustering component of the design effect for estimates of population totals may be modeled in the following form:

$$deft_c^2(\hat{Y}) = deft_c^2(\hat{P}) + \frac{b_y}{cv_{srswr}^2(\hat{Y})} \qquad (4\text{-}4)$$

where $deft_c^2(\hat{P})$ can be defined by either (2-2a) or (2-2b), $cv_{srswr}^2(\hat{Y})$ is the relative variance of $\hat{Y}$ under SRSWR, and b_y is a factor that differentiates the two design effects. The way to express the quantity b_y can vary over different sample designs (Park and Lee, 2003). In fact, the quantity Δ_y in expression (4-1) is almost parallel to the ratio $b_y / cv_{srswr}^2(\hat{Y})$. This is because the difference Δ_y is not much influenced by the effects due to unequal weighting and stratification, as discussed above. Thus, $b_y / cv_{srswr}^2(\hat{Y})$ may be understood as parallel to Δ_y given in expression (4-1).

5. Application to the Third National Incidence Study of Child Abuse and Neglect (NIS-3)

We used the NIS-3 Public Use File (PUF) for analyses. Table 5-1 shows estimates of proportions and design factors (*deft*, i.e., square root of design effect) for estimates of proportions and totals for ten selected attributes by the child abuse endangerment standard definition. The first six attributes pertain to overlapping subclasses of children who are countable under the endangerment standard (such as educationally neglected, emotionally abused, etc.) Consecutive pairs of the last four attributes are exhaustive socio-demographic categories. Note that the denominator for these proportions is relative to the entire population of endangerment standard children. For example, the first type of abuse or neglect applies to an estimated 28 percent of the estimated total number of endangerment standard children investigated by CPS agencies ($\hat{N}$ =814,331). The design factor for these estimates of proportions ranged from 1.30 to 3.82. The design factors for the estimates of totals, as expected, are larger than those for the estimates of proportions ($deft(\hat{Y})$ ranged from 2.21 to 7.61.) The only exception is for attribute 10, where the estimated total has a slightly smaller design factor than the estimated proportion. Note that the difference between $deft(\hat{P})$ and $deft(\hat{Y})$ tends to be influenced by the magnitude of $\hat{P}$ to a great extent, as discussed in Section 4.1.

Table 5-1. Proportion estimates $\hat{P}$ and overall design factors for ten selected attributes

Attribute	$\hat{P}$	$deft(\hat{P})$	$deft(\hat{Y})$
1	0.28	1.70	2.21
2	0.14	1.72	2.26
3	0.15	1.91	2.10
4	0.52	1.30	4.28
5	0.13	2.01	2.62
6	0.03	2.59	2.76
7	0.21	3.34	4.00
8	0.79	3.34	7.61
9	0.53	3.82	6.58
10	0.47	3.82	3.75

Note: $\hat{N}$ = 814,331.

We applied the multiplicative decomposition model (2-5) to evaluate the design effect components. The design factor for unequal weighting calculated from expression (2-3) was 1.20. The relative variance of sampling weights (cv_w^2) was 44 percent.

Table 5-2. Design factors due to implicit stratification for ten selected attributes

Attribute	$deft(\hat{P})$	$deft(\hat{Y})$
1	1.11	0.93
2	0.95	0.88
3	1.06	0.83
4	0.92	1.04
5	0.80	0.81
6	1.09	1.10
7	0.85	0.95
8	0.85	0.87
9	1.04	1.10
10	1.04	0.79
Mean	0.97	0.93

The design factors due to stratification for estimated proportions and totals were calculated from expressions (3-3) and (4-2), respectively. Table 5-2 shows the estimates of the design factors due to implicit stratification for ten attributes. They varied from 0.80 to 1.11 for $\hat{P}$ and from 0.79 to 1.10 for $\hat{Y}$. Their respective average design factors were 0.97 and 0.93.

The remainder of the total design effect was attributed to the clustering effect, from which the intraclass correlation for each attribute was then estimated. Table 5-3 shows the resulting estimates of the intracluster correlation for the ten attributes. Expression (2-4) was used to take the large variation in cluster sample sizes into consideration in calculating the intracluster correlation. They were typically small, ranging from 0.004 to 0.102. The adjusted (or weighted) average of cluster sample size was $\overline{n}' = 94.7$. It is larger than the unadjusted average of cluster size ($\overline{n}$ =46.1) by more than two-fold. That is, the variable cluster sample size inflated the impact of the intracluster correlation on the design factor due to clustering. Table 5-3 presents estimates of P, the factor b_y , and $cv_{srswr}^2(\hat{Y})$ by attribute. The quantities b_y were estimated from expression (4-4). As expected (see Section 4) the ratio of $b_y / cv_{srswr}^2(\hat{Y})$ was, in general, close to Δ_y. This suggests that the difference between the two overall design effects for $\hat{Y}$ and P can be mostly attributed to the difference between the associated cluster effects.

Table 5-3. $\hat{\rho}'$, b_y, $cv_{srswr}^2(\hat{Y})$

Attribute	$\hat{\rho}'$	b_y	$cv_{srswr}^2(\hat{Y})$
1	0.006	0.0032	0.0014
2	0.013	0.0075	0.0033
3	0.013	0.0066	0.0031
4	0.004	0.0053	0.0005
5	0.036	0.0099	0.0035
6	0.031	0.0063	0.0165
7	0.102	0.0033	0.0020
8	0.102	0.0061	0.0001
9	0.088	0.0074	0.0005
10	0.088	0.0040	0.0006

Note: The unadjusted and adjusted average cluster sample sizes are $\overline{n}$ =46.1 and $\overline{n}'$ =94.7. Also, the inflation factor is $(\overline{n}'-1)/(\overline{n}-1)$ =2.08.

Figure 5-1 presents comparisons of design factor components for the proportion estimates across attributes. Figure 5-1 shows that unequal weighting was not a dominant contributing factor for the overall design factor. The clustering effect, by contrast, did determine the pattern of the design factor across attributes in general. The only exceptions were attributes 1 and 4. Stratification had a beneficial impact on the design effect for some attributes. For example, the overall design factor for attribute 5 was even less than its design factor component for clustering because the design factor for stratification was less than 1.

Figure 5-2 presents the design factor components for estimated totals across attributes. Unequal weighting has a negligible impact on the overall design factor. Clustering parallels the overall design factor in its pattern across attributes. In contrast to the proportion estimates, its effect was always larger than the weighting effect in the size of the design factor component. In addition, the stratification was much more desirable than it was for the proportion estimates.

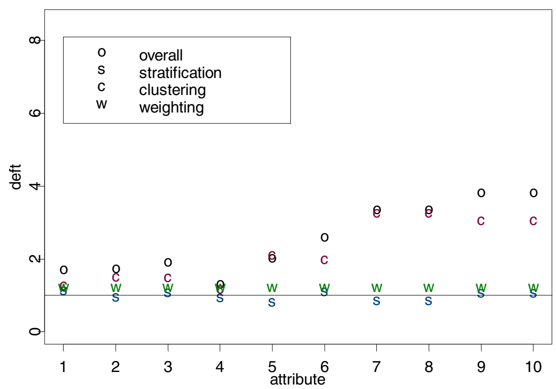

Figure 5-1. Comparisons of Design Factor Components for Proportion Estimates Across Attributes

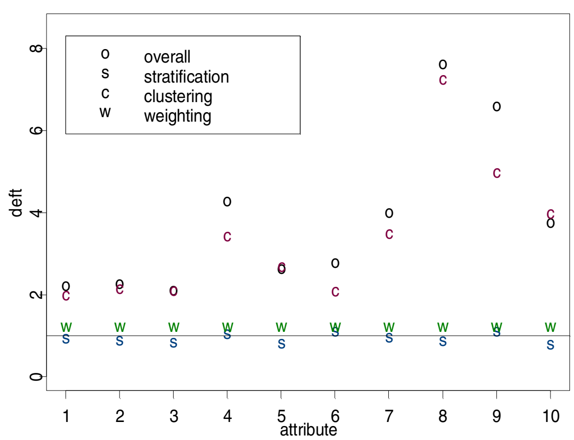

Figure 5-2. Comparisons of Design Factor Components for Total Estimates Across Attributes

6. Discussion

Sample design often includes complex features such as (implicit) stratification, unequal weighting, and clustering. Each design feature has a different impact on sample efficiency. By examining their relative contributions, we seek alternatives to lower the overall design effect and improve the effective sample size in the next survey. In this paper, we explored a three-factor design effect decomposition model suitable for the NIS-3 and estimated the design effects for each individual design component for ten key attributes. We observed relatively minor gain from the implicit stratification of NIS-3 PSUs and moderate loss from unequal weighting. The next NIS may benefit from adopting design options that would improve on both of these design features.

7. References

Barron, E. W., and Finch, R. H. (1978). Design Effects in a complex multistage sample: The Survey of Low Income Aged and Disabled (SLIAD), *Proceedings of the Section on Survey Research Methods, American Statistical Association,* pp. 400-405.

Gabler, S., Haeder, S., and Lahiri, P. (1999). A Model Based Justification of Kish's Formula for Design Effects for Weighting and Clustering. *Survey Methodology*, **25**, pp.105-106.

Holt, D. (1980). Discussion of the Paper by Verma, Scott, and O'Muircheartaigh. *Journal of the Royal Statistical Society A*, 143, Part 4, pp. 468-469.

Kalton, G. (1979). Ultimate cluster sampling. *Journal of the Royal Statistical Society A,* 142, pp. 210-222.

Kish. L. (1965). *Survey Sampling*. (Section 11.7) Wiley and Sons, New York.

Kish, L. (1987). Weighting in Deft^2. *The Survey Statistician*. The Newsletter of the International Association of Survey Statisticians, June.

Kish, L. (1992). Weighting for Unequal p_i. *Journal of Official Statistics*, **8**, pp.183-200.

Lê, T., Brick, J.M., and Kalton, G. (2001). Decomposing Design Effects. *Proceedings of the Section on Survey Research Methods, American Statistical Association,,* Alexandra VA. [CD-ROM].

Park, I., and Lee, H. (2001). The Design Effect: Do We Know All About It? *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Alexandra VA. [CD-ROM].

Park, I., and Lee, H. (2002). A Revisit of Design Effects Under Unequal Probability Sampling. *The Survey Statistician*. The Newsletter of the International Association of Survey Statisticians, No. 46, pp. 23-26.

Park, I., and Lee, H. (2003). Design Effects for the Weighted Mean and Total Estimators Under Complex Survey Sampling. A forthcoming working paper.

Spencer, B. D. (2000). An Approximate Design Effect for Unequal Weighting When Measurements May Correlate With Selection Probabilities. *Survey Methodology,* **26**, pp. 137-138.

Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Verma, V., Scott, C., and O'Muircheartaigh, C. (1980). Sample Designs and Sampling Errors for the World Fertility Survey. *Journal of the Royal Statistical Society A*, 143, Part 4, pp. 468-469.

Westat (2001). *WesVar 4.0 User's Guide.* Rockville, MD: Westat, Inc.

Wolter, K. (1985). *Introduction to Variance Estimation*. Springer-Verlag: New York.