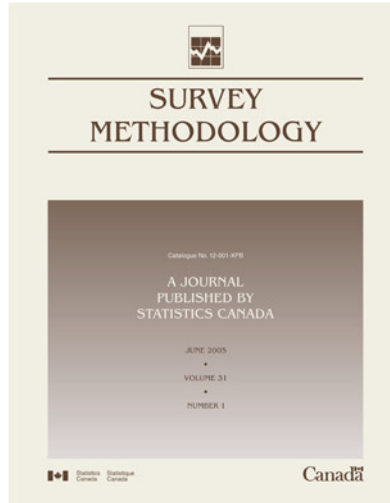
Catalogue no. 12-001-XIE

Survey Methodology

December 2004



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

December 2004

Published by authority of the Minister responsible for Statistics Canada

April 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 30, No. 2, pp. 183-193
Statistics Canada, Catalogue No. 12-001

Design Effects for the Weighted Mean and Total Estimators Under Complex Survey Sampling

Inho Park and Hyunshik Lee [1]

Abstract

We revisit the relationship between the design effects for the weighted total estimator and the weighted mean estimator under complex survey sampling. Examples are provided under various cases. Furthermore, some of the misconceptions surrounding design effects will be clarified with examples.

Key Words: Simple random sample; pps sampling; Multistage sampling; Self-weighting; Poststratification; Intracluster correlation coefficient.

1. Introduction

The design effect is widely used in survey sampling for developing a sampling design and for reporting the effect of the sampling design in estimation and analysis. It is defined as the ratio of the variance of an estimator under a complex sampling design to that of the estimator under simple random sampling with the same sample size. An estimated design effect is routinely produced by computer software packages for complex surveys such as WesVar and SUDAAN. It was originally intended and defined for the weighted (ratio) estimator of the population mean (Kish 1995). However, a common practice has been to apply this concept for other statistics such as the weighted total estimator often with success but at times with confusion and misunderstanding. The latter situation occurs particularly when simple but useful results derived under a relatively simple sampling design are applied to more complex problems. In this paper, we examine the relationship between the design effects for the weighted total estimator and the weighted mean estimator under various complex survey sampling designs. In section 2, we briefly review the definition of the design effect and its practical usage while discussing some of the misconceptions surrounding design effects for the weighted total and mean estimators. Subsequently, in section 3, we analyze the difference between the design effect for the weighted total estimator and that for the weighted mean estimator under a two-stage sampling design followed by a discussion regarding the design effects under various two-stage sampling designs and some more general cases in section 4. We try to clarify some of the misconceptions with these examples. Finally, we summarize our discussion in section 5.

2. A Brief Review on Definition and Use of Design Effect in Practice

A precursor of the design effect that has been popularized by Kish (1965) was used by Cornfield (1951). He defined the efficiency of a complex sampling design for estimating a population proportion as the ratio of the variance of the proportion estimator under simple random sampling with replacement (srswr) to the corresponding variance under a simple random cluster sampling design with the same sample size. The inverse of the ratio defined by Cornfield (1951) was also used by others. For example, Hansen, Hurwitz and Madow (1953, Vol. I, pages $259-270$) discussed the increase of the relative variance of a ratio estimator due to the clustering effect of cluster sampling over simple random sampling without replacement (srswor). The name, design effect, or Deff in short, however, was coined and defined formally by Kish (1965, section 8.2, page 258) as "the ratio of the actual variance of a sample to the variance of a simple random sample of the same number of elements" (for more history, see also Kish 1995, page 73 and references cited therein).

Suppose that we are interested in estimating the population mean $(\bar{Y})$ of a variable y from a sample of size m drawn by a complex sampling design denoted by p from a population of size M. Kish's Deff for an estimate $(\bar{y}_p)$ is given by

$$\text{Deff} = \frac{V_p(\bar{y}_p)}{(1-f)\,S_y^2 \big/ m} \tag{2.1}$$

where V_p denotes variance with respect to $p, f = m/M$ is the overall sampling fraction, and $S_y^2 = (M-1)^{-1} \sum_{k=1}^{M} (y_k - \bar{Y})^2$ is the population element variance of the

1. Inho Park and Hyunshik Lee, Westat, Inc. 1650 Research Blvd., Rockville, MD 20850, U.S.A. E-mail: InhoPark@westat.com.

y-variable. Although the design effect was originally intended and defined for an estimator of the population mean (Kish 1995), it can be defined for any meaningful statistic computed from a sample selected by a complex sampling design.

The Deff is a population quantity that depends on the sampling design and refers to a particular statistic estimating a particular population parameter of interest. Different estimators can estimate the same parameter and their design effects are different even under the same design. Therefore, the design effect includes not only the efficiency of the design but also the efficiency of the estimator. Särndal, Swensson, and Wretman (1992, page 54) made this point clear by defining it as a function of the design (p) and the estimator ($\hat{\theta}$) for the population parameter ($\theta = \theta(y)$). Thus, we may write it as

$$\mathrm{Deff}_p(\hat{\theta}) = \frac{V_p(\hat{\theta})}{V_{\mathrm{srswor}}(\hat{\theta}')}$$

where $\hat{\theta}'$ is the usual form of an estimator for θ under srswor, which is normally different from $\hat{\theta}$. For example, to estimate the population mean, one may use the weighted (ratio) mean $\hat{\theta} = \sum_s w_k y_k / \sum_s w_k$ with sampling weights w_k but $\hat{\theta}'$ would be the simple sample mean $\sum_s y_k / m$, where the summation is over the sample s. We will see the effect of particular estimators $\hat{\theta}$ on the design effect in the later sections.

Kish (1995) later advocated using a somewhat different definition, which is called Deft and uses the srswr variance in the denominator on the ground that without-replacement sampling is a part of the design and should be captured in the definition. He also reasoned that Deft is easier to use for making inferences and that it is better to define the design effect without the finite population correction factor $(1-f)$ because the factor is difficult to compute in some situations. The new definition is given by

$$\mathrm{Deft}_p(\hat{\theta}) = \sqrt{\frac{V_p(\hat{\theta})}{V_{\mathrm{srswr}}(\hat{\theta}')}}$$

or $\mathrm{Deft}_p^2(\hat{\theta}) = V_p(\hat{\theta}) / V_{\mathrm{srswr}}(\hat{\theta}')$. Survey data software such as WesVar and SUDAAN produce Deft^2 instead of Deff. We will use this definition in this paper.

When the population parameter is the total (Y), the unbiased estimator is the weighted sample total, namely, $\hat{Y} = \sum_s w_k y_k$. When the population mean is the parameter of interest, it is usually estimated by the weighted mean, that is, $\hat{\bar{Y}} = \sum_s w_k y_k / \sum_s w_k$. It is a special case of the ratio estimator, $\sum_s w_k y_k / \sum_s w_k x_k$, where $x_k \equiv 1$ for all $k \in s$.

One common misconception about the design effects for $\hat{Y}$ and $\hat{\bar{Y}}$ is that they are similar in values. However, it has been observed that the design effect for $\hat{Y}$, $\mathrm{Deft}_p^2(\hat{Y})$, tends to be much larger than that for $\hat{\bar{Y}}$, $\mathrm{Deft}_p^2(\hat{\bar{Y}})$. This was also noted in, for example, Kish (1987) and Barron and Finch (1978). Some explanation can be found in Hansen *et al.* (1953, Vol. I, pages 336 – 340) who showed that the difference arises from the relative variance of the cluster sizes. More recently Särndal *et al.* (1992, pages 315 – 318) showed that contrary to the case of $\hat{\bar{Y}}$, the design effect for $\hat{Y}$ depends on the (relative) variation of the y-variable. In fact, even the design effect for $\hat{\bar{Y}}$ may depend on the (relative) variation of the y-variable, which we will discuss in section 4. This dependence contradicts what the design effect is intended to measure as Kish (1995) explicitly described:

> "Deft are used to express the effects of sample design beyond the elemental variability (S_y^2/m), removing both the units of measurement and sample size as nuisance parameters. With the removal of S_y, the units, and the sample size m, the design effects on the sampling errors are made generalizable (transferable) to other statistics and to other variables, within the same survey, and even to other surveys."

His statement may be loosely true for the weighted mean $\hat{\bar{Y}}$ as expressed in the frequently used sample approximate formula for $\mathrm{Deft}_p^2(p, \hat{\bar{Y}})$ given by Kish (1987):

$$\mathrm{Deft}_p^2(\hat{\bar{Y}}) = \{1 + \rho(\bar{m} - 1)\}\left(1 + \mathrm{cv}_w^2\right) \qquad (2.2)$$

where the sample design p contains complex features such as unequal weighting and cluster sampling, $\rho = \rho_{\mathrm{p}}(y)$ is the intraclass correlation coefficient (often called within cluster homogeneity measure), $\bar{m}$ is the average cluster sample size, and cv_w^2 is the sample relative variance of the weights. Strictly speaking, this formula is not independent of the y-variable because ρ is dependent on the y-variable. Also, the design effect may not be free of the unit of measurement unless $V_p(\hat{\bar{Y}})$ is expressed in a factorial form of S_y^2/m. See Park and Lee (2002). This formula (2.2) is valid only when there is no correlation between the sampling weights and the survey variable y. However, if the correlation is present, the formula may need to be modified as studied by Spencer (2000) and Park and Lee (2001). In the following section, we elaborate this aspect in detail for two-stage sampling and we will also examine this point further in section 4.1.

3. Decomposition of the Design Effect Under Two-Stage Sampling

We consider a sampling design conducted in two stages. Suppose that a population $U = \{k : k = 1, \ldots, M\}$ with M elements is grouped into N clusters of size M_i such that

$M=\sum_{i=1}^{N} M_i$. The first stage sample $s_a=\{i: i=1,\ldots,n\}$ of n clusters (primary sampling units, or PSUs in abbreviation) is selected with replacement from N clusters with probabilities p_i, where $\sum_{i=1}^{N} p_i=1$. Let $p_a=\Pr(s_a)$ denote the first stage sampling design. The second stage sample $s_{bi}=\{j: j=1,\ldots,m_i\}$ of m_i elements (secondary sampling units or SSUs in abbreviation) is then selected independently from each PSU i selected at the first stage according to some arbitrary sampling design, say $p_{bi}=\Pr(s_{bi}\,|\,s_a)$ where $i\in s_a$. Denote the total sample of elements and the overall sampling design by $s=\cup_{i\in s_a} s_{bi}$ and $p=\Pr(s)$, respectively. Associated with the j^{th} element in the i^{th} cluster is a survey characteristic $y_{ij}, j=1,\ldots,M_i$, $i=1,\ldots,N$. For a given $i\in s_a$, let $w_{j|i}$ be the second stage sampling weights such that an estimator of the form $\hat{Y}_i=\sum_{j=1}^{m_i} w_{j|i}\, y_{ij}$ is unbiased for the cluster total $Y_i=\sum_{j=1}^{M_i} y_{ij}$, that is, $E_b(\hat{Y}_i)=Y_i$, where E_b represents the expectation with respect to the second stage sampling. Let $w_i=1/(np_i)$ be the first stage sampling weights and let $Y=\sum_{i=1}^{N} Y_i$ be the population total. It is easy to show that $E_a(Y_i/p_i)=Y$. Assuming that Y_i are known for $i\in s_a$, $\sum_{i=1}^{n} w_i Y_i$ is the average of n unbiased estimators of Y so that $E_a(\sum_{i=1}^{n} w_i Y_i)=Y$, where E_a denotes the expectation with respect to the first stage sampling design. Note that both stages are sampling with replacement. Accordingly, it is possible that the same sampling unit (either cluster or element) is selected more than once but they are treated differently. Define the overall sampling weights by $w_{ij}=w_i\, w_{j|i}$. Clearly, $\hat{Y}=\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}\, y_{ij}$ is unbiased for Y, that is, $E_p(\hat{Y})=E_a\,E_b\,(\hat{Y})=E_a\,(\sum_{i=1}^{n} w_i\, Y_i)=Y$, where E_p represents the expectation with respect to p. The variance of $\hat{Y}$ can be written as

$$\begin{aligned} V_p(\hat{Y}) &= V_a E_b(\hat{Y}) + E_a V_b(\hat{Y}) \\ &= \sum_{i=1}^{N} w_i (Y_i - p_i Y)^2 + \sum_{i=1}^{N} w_i V_b(\hat{Y}_i) \qquad (3.1) \end{aligned}$$

where V_p, V_a and V_b represent variances defined with respect to the overall, the first stage, and the second stage sampling. See Särndal *et al.* (1992, pages 151 – 152).

A commonly used estimator for the population mean $\bar{Y}=Y/M$ is the weighted (ratio) estimator given by $\hat{\bar{Y}}=\hat{Y}/\hat{M}$ where $\hat{M}=\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}$. Using Taylor linearization, as shown in Särndal *et al.* (1992, pages 176 – 178), $\hat{\bar{Y}}$ can be approximated as

$$\hat{\bar{Y}} \cong \bar{Y} + M^{-1}\hat{D} \qquad (3.2)$$

where $\hat{D}=\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}\, d_{ij}$ is an unbiased estimator of the population total $D=\sum_{i=1}^{N}\sum_{j=1}^{M_i} d_{ij}$ of $d_{ij}=y_{ij}-\bar{Y}$, which represents the deviation of y_{ij} from the population mean $\bar{Y}$. Note that $D=0$. Denoting $D_i=\sum_{j=1}^{M_i} d_{ij}=Y_i-M_i\,\bar{Y}$ and $\hat{D}_i=\sum_{j=1}^{m_i} w_{j|i}\, d_{ij}$, we obtain the approximate variance of $\hat{\bar{Y}}$ from expression (3.2) as

$$\mathrm{AV}_p(\hat{\bar{Y}}) = \frac{1}{M^2}\left[\sum_{i=1}^{N} w_i\left(Y_i - \frac{M_i}{M}Y\right)^2 + \sum_{i=1}^{N} w_i V_b(\hat{D}_i)\right]. \qquad (3.3)$$

If a simple random sample of size $m=\sum_{i=1}^{n} m_i$ is selected with replacement from the population U, then a sample mean $\bar{y}_{\text{srs}}=\sum_s y_k/m$ and its expansion

$$\hat{Y}_{\text{srs}} = M\,\bar{y}_{\text{srs}} = \frac{1}{f}\sum\nolimits_s y_k \qquad (3.4)$$

would serve as the estimators of the population mean $\bar{Y}$ and total Y, respectively, under srswr, where $f=m/M$ is the overall sampling fraction. Their variances under this sampling design are given as $V_{\text{srswr}}(\hat{Y}_{\text{srs}})=M^2\,V_{\text{srswr}}(\bar{y}_{\text{srs}})$, where $V_{\text{srswr}}(\bar{y}_{\text{srs}})=m^{-1}\,S_y^2$ and $S_y^2=(M-1)^{-1}\sum_s(y_k-\bar{Y})^2$. We note that m is the achieved sample size, which is a random quantity in general. From (3.1), (3.3), and above expressions with m replaced by its expected value m' with respect to the overall sampling design p, *i.e.*, $m'=E_p(m)$, the design effects for $\hat{Y}$ and $\hat{\bar{Y}}$ can be written as

$$\mathrm{Deft}_p^2(\hat{Y}) = \frac{m'}{\mathrm{CV}_y^2}\left\{\sum_{i=1}^{N} w_i\left(\frac{Y_i}{Y}-p_i\right)^2 + \sum_{i=1}^{N} w_i V_b\left(\frac{\hat{Y}_i}{Y}\right)\right\} \qquad (3.5)$$

and

$$\mathrm{Deft}_p^2(\hat{\bar{Y}}) \cong \frac{m'}{\mathrm{CV}_y^2}\left\{\sum_{i=1}^{N} w_i\left(\frac{Y_i}{Y}-\frac{M_i}{M}\right)^2 + \sum_{i=1}^{N} w_i V_b\left(\frac{\hat{D}_i}{Y}\right)\right\} \qquad (3.6)$$

where $\mathrm{CV}_y^2=S_y^2/\bar{Y}^2$ represents the population relative variance of the y-variable. From these expressions, the difference in design effects for $\hat{Y}$ and $\hat{\bar{Y}}$ can be written as follows.

$$\mathrm{Deft}_p^2(\hat{Y}) - \mathrm{Deft}_p^2(\hat{\bar{Y}}) \cong \Delta_a + \Delta_b, \qquad (3.7)$$

where

$$\Delta_a = \frac{m'}{\mathrm{CV}_y^2}\left\{\sum_{i=1}^{N} w_i\left[\left(\frac{Y_i}{Y}-p_i\right)^2 - \left(\frac{Y_i}{Y}-\frac{M_i}{M}\right)^2\right]\right\}$$

and

$$\Delta_b = \frac{m'}{\mathrm{CV}_y^2}\left\{\sum_{i=1}^{N} w_i\left[V_b\left(\frac{\hat{Y}_i}{Y}\right) - V_b\left(\frac{\hat{D}_i}{Y}\right)\right]\right\}.$$

The two components Δ_a and Δ_b in expression (3.7) reflect the differences arising from the respective sources of variation from the first and second stages of sampling. Of course, the second component disappears if all the elements in selected clusters are observed since it becomes a single-stage design or if a simple random sample is selected in the second stage. This is because both variances $V_b(\hat{Y}_i)$ and $V_b(\hat{D}_i)$ are equivalent under the aforementioned conditions, that is, 1) $V_b(\hat{Y}_i)=V_b(\hat{D}_i)=0$ if $w_{j|i}=1$ for all i and j, and

2) $V_b(\hat{Y}_i) = V_b(\hat{D}_i) \geq 0$ if $w_{j|i} = M_i/m_i$ for all i and j. In other words,

$$\Delta_b = 0 \quad \text{if} \quad w_{j|i} = c_i \text{ for all } i \text{ and } j, \tag{3.8}$$

where c_i are nonnegative constants and not necessarily equal for different clusters. Meanwhile, we can show that

$$\Delta_a = \begin{cases} 0 & \text{if} \quad p_i \propto M_i, \\ A_p(y) & \text{if} \quad Y_i \propto M_i, \\ -A_p(y) & \text{if} \quad p_i \propto Y_i, \end{cases} \tag{3.9}$$

for all i, where $A_p(y) = (m'/\text{CV}_y^2)\sum_{i=1}^N w_i(p_i - M_i/M)^2$. Note that $A_p(y)$ is a nonnegative quantity and also that the conditions in expression (3.9) can be restated, respectively, as $p_i = M_i/M$, $\overline{Y}_i = \overline{Y}$, and $p_i = Y_i/Y$, where $\overline{Y}_i = Y_i/M_i$ for all $i = 1, \ldots, N$. This result reveals the effect of cluster sampling on the precision of the two estimators. For example, if $p_i = M_i/M$, cluster sampling makes no difference in the precision of the two estimators. On the other hand, if $p_i = Y_i/Y$, $\hat{Y}$ becomes more efficient than $\hat{\overline{Y}}$ in precision under cluster sampling, whereas the cluster sampling favors $\hat{\overline{Y}}$ over $\hat{Y}$ in terms of precision if $\overline{Y}_i = \overline{Y}$ for all i.

Now, let us consider some examples of the conditions of (3.8) and (3.9).

Example 3.1 For one or two-stage cluster design with pps cluster sampling using $p_i = M_i/M$ and $w_{j|i} = c_i$ for all $i = 1, \ldots, N$, we have from (3.8) and (3.9) that $\Delta_a = \Delta_b = 0$, that is, there is no difference in the design effects for $\hat{\overline{Y}}$ and $\hat{Y}$.

The same result as given in example 3.1 can be achieved by $\hat{Y} = M\overline{Y}$. This estimator is the ratio estimator, which can be used if M is known. The case that overall sampling weights are a constant for all the elements (*i.e.*, self-weighting sampling design) is a well known special case. We will come back to this in section 4.

Example 3.2 One-stage simple random cluster sampling or two-stage sample design with srs for both stages. Under these designs, we have $w_{j|i} = c_i$ and $p_i = 1/N$ for all i and j and thus, it follows from (3.8) and (3.9) that $\Delta_b = 0$ and

$$\Delta_a = \begin{cases} 0 & \text{if} \quad M_i \equiv M_0 \text{ for all } i, \\ \overline{m}' \dfrac{\text{CV}_M^2}{\text{CV}_y^2} & \text{if} \quad \overline{Y}_i \text{ are all equal}, \\ -\overline{m}' \dfrac{\text{CV}_M^2}{\text{CV}_y^2} & \text{if} \quad Y_i \text{ are all equal}, \end{cases} \tag{3.10}$$

where $\overline{m}' = m'/n$, $\text{CV}_M^2 = \overline{M}^{-2}\sum_{i=1}^N (M_i - \overline{M})^2/N$ denotes the relative variance of cluster sizes M_i, and $\overline{M} = M/N$ denotes the average size of clusters. The conditions in (3.10) also satisfy the conditions in (3.9) and therefore, (3.10) is a special case of (3.9). Note that the quantity $A_p(y)$ in expression (3.9) approximately reduces to $\overline{m}' \cdot \text{CV}_M^2/\text{CV}_y^2$ when $p_i = 1/N$ for all i.

Example 3.2 shows that when unequal cluster sizes are not reflected in the sampling design, the relative efficiency of $\hat{Y}$ over $\hat{\overline{Y}}$ depends in part on the relative variability of cluster sizes. If the cluster means are all equal, then cluster sampling makes $\hat{\overline{Y}}$ more efficient than $\hat{Y}$, vice versa if all the cluster totals are equal. On the other hand, if all clusters are equal in size, no difference in the design effects arises by simple random sampling of clusters.

In section 4, we utilize the results derived in this section to discuss other examples used in the sampling literature.

4. Examples on the Design Effect in the Sampling Literature

4.1 Unequal Probability Element Sampling

Consider an unequal probability element sampling design without clustering. The discussion in section 3 applies to this example with $M_i \equiv 1$ for all $i = 1, \ldots, N$ and thus, $m = n$. For brevity's sake, we use lower cases y_i to denote the value of the y-variable, and we also assume that N is large so that $N/(N-1) \cong 1$. Due to the absence of the second stage sampling variation, the design effects for $\hat{Y}$ and $\hat{\overline{Y}}$ given in expressions (3.5) and (3.6) reduce to

$$\text{Deft}_p^2(\hat{Y}) \cong \frac{\sum_{i=1}^N p_i^{-1}(y_i - p_iY)^2}{\sum_{i=1}^N N(y_i - \overline{Y})^2} \tag{4.1}$$

and

$$\text{Deft}_p^2(\hat{\overline{Y}}) \cong \frac{\sum_{i=1}^N p_i^{-1}(y_i - \overline{Y})^2}{\sum_{i=1}^N N(y_i - \overline{Y})^2}. \tag{4.2}$$

Further let us consider an example where the survey variable y is not correlated with the selection probability p_i.

Example 4.1 Unequal probability element sampling with no correlation between y_i and p_i. When y_i and p_i are not correlated, we can approximate $\sum_{i=1}^N p_i^{-1}(y_i - \overline{Y})^2$ by $n\overline{W}\sum_{i=1}^N (y_i - \overline{Y})^2$, where $\overline{W} = N^{-1}\sum_{i=1}^N w_i$. Note that $E_p(n^{-1}\sum_{i=1}^n w_i) = N/n$, $E_p(n^{-1}\sum_{i=1}^n w_i^2) = N\overline{W}/n$ and $E_p(n^{-1}\sum_{i=1}^n w_i^2)/E_p^2(n^{-1}\sum_{i=1}^n w_i) = n\overline{W}/N$. Thus,

$$\begin{aligned} \text{Deft}_p^2(\hat{\overline{Y}}) &\cong n\overline{W}/N \\ &= E_p\left(n^{-1}\sum_{i=1}^n w_i^2\right) \Big/ E_p^2\left(n^{-1}\sum_{i=1}^n w_i\right). \end{aligned} \tag{4.3}$$

It is easy to show that $n\overline{W}/N \geq 1$ using the Cauchy-Schwarz inequality (Apostol 1974, page 14). In addition, routine calculations show from (4.1) and (4.2) that

$$\begin{aligned}&\mathrm{Deft}_p^2(\hat{Y})-\mathrm{Deft}_p^2(\hat{\bar{Y}})\\&\cong \mathrm{CV}_y^{-2}\left\{\sum_{i=1}^N p_i^{-1}(p_i-\bar{p})^2-2Y^{-1}\sum_{i=1}^N p_i(y_i-\bar{Y})(p_i-\bar{p})\right\}\\&=\mathrm{CV}_y^{-2}(n\overline{W}/N-1),\end{aligned}$$

where $\bar{p}=N^{-1}\sum_{i=1}^N p_i=1/N$. The latter expression is obtained from $\sum_{i=1}^N p_i^{-1}(p_i-\bar{p})^2=n\overline{W}/N-1$ and $\sum_{i=1}^N p_i^{-1}(y_i-\bar{Y})(p_i-\bar{p})\cong 0$ because y_i and p_i are uncorrelated. Consequently,

$$\mathrm{Deft}_p^2(\hat{Y})-\mathrm{Deft}_p^2(\hat{\bar{Y}})\cong \mathrm{CV}_y^{-2}\left\{\mathrm{Deft}_p^2(\hat{\bar{Y}})-1\right\}$$

or

$$\mathrm{Deft}_p^2(\hat{Y})\cong(1+\mathrm{CV}_y^{-2})\,\mathrm{Deft}_p^2(\hat{\bar{Y}})-\mathrm{CV}_y^{-2}. \qquad (4.4)$$

From (4.4), it is clear that $\mathrm{Deft}_p^2(\hat{Y})\geq \mathrm{Deft}_p^2(\hat{\bar{Y}})$ if $\mathrm{Deft}_p^2(\hat{\bar{Y}})\geq 1$ and the equality holds if $\mathrm{Deft}_p^2(\hat{\bar{Y}})=1$ or $\overline{W}=N/n$. Also, $\mathrm{Deft}_p^2(\hat{Y})<\mathrm{Deft}_p^2(\hat{\bar{Y}})$ if $1/(1+\mathrm{CV}_y^2)<\mathrm{Deft}_p^2(\hat{\bar{Y}})<1$.

Example 4.1 shows that $\hat{Y}$ tends to have a larger design effect than $\hat{\bar{Y}}$ if the correlation between y_i and p_i is weak and $\mathrm{Deft}_p^2(\hat{\bar{Y}})\geq 1$.

The customary quantification of the effect of unequal weights on the design efficiency shown in (2.2) is due to Kish (1965, 11.7). He considered cases where the unequal weights arise from "haphazard" or "random" sources such as frame problems or non-response adjustments. Assuming that (1) a random sample of size *n* selected with replacement is divided into *G* weighting classes such that the same weight w_g is assigned to n_g sampling units within class *g* and $n=\sum_{g=1}^G n_g$, and that (2) all *G* weighting class variances are equal to the unit variance of *y*, *i.e.*, $S_{yg}^2=S_y^2$ for all $g=1,\ldots,G$, he proposed a quantity given as

$$\mathrm{Deft}_{\mathrm{Kish}}^2(\hat{\bar{Y}})=n\sum_{g=1}^G n_g\,w_g^2\bigg/\left(\sum_{g=1}^G n_g\,w_g\right)^2, \qquad (4.5)$$

to measure the increment in the variance of $\hat{\bar{Y}}$ in comparison with the hypothesized variance under srswr of size *n*. The rationale behind the above derivation is that the loss in precision of $\hat{\bar{Y}}$ due to haphazard unequal weighting can be approximated by the ratio of the variance under disproportionate stratified sampling to that under the proportionate stratified sampling.

In (4.5), letting $n_g=1$ for all *g* and thus, $n=G$, Kish (1992) later proposed a well-known approximate formula given as

$$\mathrm{Deft}_{\mathrm{Kish}}^2(\hat{\bar{Y}})=n\sum_{i=1}^n w_i^2\bigg/\left(\sum_{i=1}^n w_i\right)^2=1+\mathrm{cv}_w^2, \qquad (4.6)$$

where $\mathrm{cv}_w^2=n^{-1}\sum_{i=1}^n(w_i-\bar{w})^2/\bar{w}^2$ is the sample relative variance and $\bar{w}$ is the sample mean of w_i. Note that (4.6) is a sample approximate of (4.3). For a sampling design which is inefficient for estimation of *Y*, the inefficiency diminishes with the ratio estimation. Next, we consider the opposite case where the *y*-variable is correlated with the selection probability p_i, where the efficiency of $\hat{Y}$ increases.

Example 4.2 Unequal probability element sampling where y_i is correlated with p_i. Suppose that y_i is linearly related with p_i by $y_i=A+Bp_i+e_i$, where *A* and *B* are the least-square regression coefficients of the model for the (finite) population and e_i is the corresponding residual. Furthermore, assume that the regression model fits well to the population data and the error variance is roughly homogeneous so that $R_{ew}\cong 0$ and $R_{e^2w}\cong 0$, where R_{ew} and R_{e^2w} denote the population correlations of pairs (e_i,w_i) and (e_i^2,w_i), respectively. For example, $R_{ew}=\sum_{i=1}^N(e_i-\bar{E})(w_i-\overline{W})/\{(N-1)S_eS_w\}$, where $\bar{E}=\sum_{i=1}^N e_i/N$; S_e and S_w are the population standard deviations of e_i and w_i, respectively. Then the design effects given by (4.1) and (4.2) reduce to

$$\begin{aligned}\mathrm{Deft}_p^2(\hat{Y})&\cong(n\overline{W}/N)(1-R_{yp}^2)\\&\quad+(n\overline{W}/N-1)\left(\frac{R_{yp}}{\mathrm{CV}_p}-\frac{1}{\mathrm{CV}_y}\right)^2\end{aligned} \qquad (4.7)$$

and

$$\begin{aligned}\mathrm{Deft}_p^2(\hat{\bar{Y}})&\cong(n\overline{W}/N)(1-R_{yp}^2)\\&\quad+(n\overline{W}/N-1)\left(\frac{R_{yp}}{\mathrm{CV}_p}\right)^2,\end{aligned} \qquad (4.8)$$

respectively, where R_{yp} is the population correlation between y_i and p_i and CV_p is the population coefficient of variation of p_i (see Park and Lee (2001) for proof). It follows from (4.7) and (4.8) that $\mathrm{Deft}_p^2(\hat{Y})\geq \mathrm{Deft}_p^2(\hat{\bar{Y}})$ if and only if

$$2R_{yp}\leq \mathrm{CV}_p/\mathrm{CV}_y, \qquad (4.9)$$

where the equality holds if and only if $2R_{yp}=\mathrm{CV}_p/\mathrm{CV}_y$. Also, the inequality is reversed when the inequality in (4.9) becomes opposite.

The condition (4.9) indicates that $\hat{Y}$ tends to be less efficient in terms of precision than $\hat{\bar{Y}}$ whenever R_{yp} is small. Thus, we see that R_{yp} plays an important role in determining the design efficiency of unequal probability sampling on $\hat{Y}$ and $\hat{\bar{Y}}$ and their relative efficiency.

In an attempt to develop an approximate expression to the design effect when y_i is correlated with p_i, Spencer (2000) proposed a sample approximate formula for $\hat{Y}$ and compared it with Kish's approximate formula (4.6) for the special case of $R_{yp}=0$. As seen in example 4.2, the two design effects (4.7) and (4.8) are not equal unless $\overline{W}=N/n$ (see Park and Lee (2001) for more discussion and some numerical examples). In addition, this special case provides the same condition as for example 4.1 and thus, the two approximate design effect formulae (4.7) and (4.8) are equivalent to (4.4) and (4.3), respectively.

4.2 One-Stage Cluster Sampling

Consider a one-stage cluster sampling, where every element in a sampled cluster is included in the sample, *i.e.*, $m_i \equiv M_i$ for all $i \in s_a$. Due to the absence of the second stage sampling variation, the variance of $\hat{Y}$ takes only the first term of expression (3.1) and it can be decomposed as

$$\sum_{i=1}^{N} w_i (Y_i - p_i Y)^2 = \frac{M(N-1)}{n} S_{yB}^2 + \sum_{i=1}^{N} w_i Q_i \overline{Y}_i^2, \quad (4.10)$$

where $S_{yB}^2 = (N-1)^{-1} \sum_{i=1}^{N} M_i (\overline{Y}_i - \overline{Y})^2$ and $Q_i = M_i (M_i - p_i M)$ for $i = 1, \ldots, N$. Note that $Q_i = 0$ if $p_i = M_i / M$, that is, p_i is proportional to the cluster size M_i. Also, note that S_{yB}^2 is the between-cluster mean square deviation in an analysis of variance. Denoting the within-cluster mean square deviation as $S_{yW}^2 = (M-N)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{M_i} (y_{ij} - \overline{Y}_i)^2$, write $S_{yB}^2 = S_y^2 \{1 + \delta (M-N)/(N-1)\}$ with $\delta = 1 - S_{yW}^2 / S_y^2$. Since the expected sample size is $m' = n\overline{M}$, the design effect for $\hat{Y}$ can be written from (4.10) as

$$\text{Deft}_p^2(\hat{Y}) = \left(\frac{N-1}{N}\right)\left(1 + \frac{M-N}{N-1}\delta\right) + \frac{n\overline{M}}{\text{CV}_y^2} \sum_{i=1}^{N} \frac{w_i Q_i}{M_i^2} \left(\frac{Y_i}{Y}\right)^2. \quad (4.11)$$

Similarly, the design effect for $\hat{\hat{Y}}$ can be expressed as

$$\text{Deft}_p^2(\hat{\hat{Y}}) \cong \left(\frac{N-1}{N}\right)\left(1 + \frac{M-N}{N-1}\delta\right) + \frac{n\overline{M}}{\text{CV}_y^2} \sum_{i=1}^{N} \frac{w_i Q_i}{M_i^2} \left(\frac{D_i}{Y}\right)^2. \quad (4.12)$$

We observe that the design effect for $\hat{\hat{Y}}$ differs from that for $\hat{Y}$ in the second term containing $D_i = \sum_{i=1}^{M_i} (y_{ij} - \overline{Y})$ instead of Y_i. In addition, we note that the quantity $\delta = \delta_p(y)$ is the adjusted coefficient of determination (R_{adj}^2) in the regression analysis context. It may be called a homogeneity measure. For more discussion on δ, see Särndal *et al.* (1992, pages 130–131) and Lohr (1999, page 140).

Example 4.3 One-stage simple random sampling of clusters. In this example, if $p_i = 1/N$ for all $i = 1, \ldots, N$, the two design effects in (4.11) and (4.12) reduce, respectively, to

$$\text{Deft}_p^2(\hat{Y}) = \left(\frac{N-1}{N}\right)\left(1 + \frac{M-N}{N-1}\delta\right) + \frac{1}{N \cdot \text{CV}_y^2} \sum_{i=1}^{N} (M_i - \overline{M}) \left(\frac{M_i}{\overline{M}}\right)\left(\frac{\overline{Y}_i}{\overline{Y}}\right)^2 \quad (4.13)$$

and

$$\text{Deft}_p^2(\hat{\hat{Y}}) \cong \left(\frac{N-1}{N}\right)\left(1 + \frac{M-N}{N-1}\delta\right) + \frac{1}{N \cdot \text{CV}_y^2} \sum_{i=1}^{N} (M_i - \overline{M}) \left(\frac{M_i}{\overline{M}}\right)\left(\frac{\overline{D}_i}{\overline{Y}}\right)^2, \quad (4.14)$$

where $\overline{M} = M/N$. Since $\text{Deft}_p^2(\hat{Y}) - \text{Deft}_p^2(\hat{\hat{Y}}) \propto \sum_{i=1}^{N} M_i (M_i - \overline{M})(2\overline{Y}_i - \overline{Y})$, the inequality between design effects for $\hat{Y}$ and $\hat{\hat{Y}}$ depends on the joint distribution of $\overline{Y}_i$ and M_i.

Example 4.4 One-stage simple random sampling of clusters of equal-size. In this case, we have $M_i \equiv M_0$ and $p_i = 1/N$ for all $i = 1, \ldots, N$ and both design effects in (4.13) and (4.14) can be approximated by the same quantity given as

$$\left(\frac{N-1}{N}\right)\left[1 + \frac{N(M_0 - 1)}{N-1}\delta\right], \quad (4.15a)$$

since $M_i - \overline{M} = 0$ for all $i = 1, \ldots, N$.

To introduce the clustering effect on variance estimation, one often uses the simplest form of one-stage simple random cluster sampling as in example 4.4. For example, see Cochran (1977, section 9.4), Lehtonen and Pahkinen (1995, page 91), and Lohr (1999, section 5.2.2). Although these authors adopted a without-replacement sampling scheme, we compare their formulae with our formulae with the with – replacement sampling assumption for the sake of both simplicity and consistency. Furthermore, the comparison is valid because their formulae are defined with the finite population correction incorporated in both numerator and denominator so that its effect is basically cancelled out. Cochran (1977, section 9.4) derived

$$\text{Deft}_p^2(\hat{\hat{Y}}) = \frac{NM_0 - 1}{M_0(N-1)}\left[1 + (M_0 - 1)\rho\right] \cong 1 + (M_0 - 1)\rho, \quad (4.15b)$$

where ρ is called the intracluster correlation coefficient defined by

$$\rho = \frac{2\sum_{i=1}^{N} \sum_{j>k=1}^{M_0} (y_{ij} - \overline{Y})(y_{ij} - \overline{Y})}{(M_0 - 1)\sum_{i=1}^{N} \sum_{j=1}^{M_0} (y_{ij} - \overline{Y})^2}. \quad (4.15c)$$

Rewriting $\sum_{i=1}^{N}[\sum_{j=1}^{M_0}(y_{ij}-\overline{Y})]^2 = M_0(N-1)S_{yB}^2$ and $\sum_{i=1}^{N}\sum_{j=1}^{M_0}(y_{ij}-\overline{Y})^2 = (NM_0-1)S_y^2 = (N-1)S_{yB}^2 + N(M_0-1)S_{yW}^2$, it is easy to show that

$$\begin{aligned} &2\sum_{i=1}^{N}\sum_{j>k=1}^{M_0}(y_{ij}-\overline{Y})(y_{ik}-\overline{Y}) \\ &= \sum_{i=1}^{N}\left[\sum_{j=1}^{M_0}(y_{ij}-\overline{Y})\right]^2 - \sum_{i=1}^{N}\sum_{j=1}^{M_0}(y_{ij}-\overline{Y})^2 \\ &= (M_0-1)\left[(NM_0-1)S_y^2 - NM_0S_y^2w\right] \end{aligned}$$

and, thus, from (4.15c), $\rho = 1-\{NM_0/(NM_0-1)\}(S_{yW}^2/S_y^2) \cong \delta$ assuming $M_i \equiv M_0$ for all $i = 1, \ldots, N$, $NM_0/(NM_0-1) \cong 1$. Therefore, further assuming $(N-1)/N \cong 1$ and $(NM_0-1)M_0^{-1}(N-1)^{-1} \cong 1$, both design effect formulae (4.15a) and (4.15b) are approximately equivalent to $1+(M_0-1)\delta$. Other authors arrived at the same approximate formula. This is because δ and ρ essentially measure the same thing, which is the cluster homogeneity. Under this situation, two estimators $\hat{Y}$ and $\hat{\overline{Y}}$ have the same design effect as discussed in example 3.2. Note that this is a simple case of a self-weighting sampling design.

Särndal *et al.* (1992, section 8.7) compared the design effects for the two estimators under the setting of example 4.3. They also derived a simplified expression $1+(\overline{M}-1)\delta$ for (4.13) and (4.14), assuming the covariances of M_i with $M_i\overline{Y}_i^2$ and $M_i\overline{D}_i^2$ are ignorable. Their discussion on the difference between total and mean estimators boils down to Δ_a in example 3.2. They also noted that the design effect can be much more severe for the population total than for the population mean because more is lost through sampling of clusters when the total is estimated than when the mean is estimated.

A common practice to handle unequal cluster sizes is to use a more efficient sampling method that incorporates the size difference such as pps sampling of clusters. Expressions (4.11) and (4.12) can be applied to arbitrary selection probabilities p_i, where p_i are set to be proportional to some size measures $Z_i \geq 0$. The difference between the design effects for $\hat{Y}$ and $\hat{\overline{Y}}$ is explained by Δ_a in (3.9), or alternatively

$$\Delta_a = \frac{m'}{\mathrm{CV}_y^2}\sum_{i=1}^{N}\frac{w_iQ_i}{M^2}\left[\left(\frac{\overline{Y}_i}{\overline{Y}}\right)^2 - \left(\frac{\overline{D}_i}{\overline{Y}}\right)^2\right]. \qquad (4.16)$$

The term Q_i in (4.16) represents the effect of p_i on the variance estimation when size measures other than the actual cluster sizes M_i are used. Thomsen, Tesfu, and Binder (1986) considered the effect of an out-dated size measure among other factors under two-stage sampling with simple random sample of element at the second stage. We will come back to this in section 4.4.

4.3 Self-Weighting Designs

In a self-weighting sample, every sample element has the same weight. This leads to simple forms for both total and mean estimators. They are given by $\hat{Y} = y/f$ and $\hat{\overline{Y}} = y/m$, where $f = m/M$ is the overall sampling fraction and $y = \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}$ is the sample total. Then just like simple random sampling as shown in (3.4), the two estimators have the same design effect.

A self-weighting sampling design can be implemented in various ways by synchronizing the first stage sampling method with the second stage sampling method (*e.g.*, Kish 1965, section 7.2). For example, if equal probability sampling is used for the first stage sampling, then the second stage should be sampled by an equal probability sampling method with a uniform sampling fraction for all PSUs. As a special case of this, where an srs of PSUs of equal size (*i.e.*, $M_i = M_0$ for all i) is selected, Hansen *et al.* (1953, Vol. II, pages 162 – 163) showed

$$\mathrm{CV}_p^2(\hat{\overline{Y}}) \cong \frac{1}{m}\mathrm{CV}_y^2\left[1+\rho(\overline{m}-1)\right], \qquad (4.17)$$

where $\mathrm{CV}_p^2(\hat{\overline{Y}}) = V_p(\hat{\overline{Y}})/\overline{Y}^2$ is the relative variance of $\hat{\overline{Y}}$ under the sampling design p and ρ is the intracluster correlation coefficient as defined in (4.15c). Since the relative variance of $\hat{\overline{Y}}$ under srswr is $m^{-1}\mathrm{CV}_y^2$, the well known approximate design effect formula for $\hat{\overline{Y}}$ under a self-weighting design follows immediately as

$$\mathrm{Deft}_p^2(\hat{\overline{Y}}) = 1+\rho(\overline{m}-1). \qquad (4.18)$$

For one-stage cluster designs, we showed similar forms given in (4.15a) and (4.15b) (see also Yamane 1967, section 8.7). Hansen *et al.* (1953, Vol. II. page 204) further showed $\mathrm{CV}_p^2(\hat{Y}) = \mathrm{CV}_p^2(\hat{\overline{Y}})$ for a sample design that employs simple random sampling at both stages. This implies that $\hat{Y}$ and $\hat{\overline{Y}}$ have the same design effect.

4.4 Two-Stage Unequal Probability Sampling

Let us first consider the following example.

Example 4.5 A two-stage sampling design where n PSUs are selected with replacement with probability p_i and an equal size simple random sample of $m_0 \geq 2$ elements is selected with replacement from each selected PSU. With routine calculations and simplification, we can show that

$$\mathrm{Deft}_p^2(\hat{Y}) \cong 1+(m_0-1)\tau + W_y^*, \qquad (4.19)$$

where

$$\tau = \frac{(N-1)S_{yB}^2 + \sum_{i=1}^{N}(m_0-1)^{-1}S_{yi}^2}{(N-1)S_{yB}^2 + \sum_{i=1}^{N}(M_i-1)S_{yi}^2}, \qquad (4.20)$$

$S_{yi}^2 = (M_i-1)^{-1}\sum_{j=1}^{M_i}(y_{ij}-\overline{Y}_i)^2$, $W_y^* = W_y/V_{\mathrm{srswr}}(\hat{Y}_{\mathrm{srs}}) = (m_0/\mathrm{CV}_y^2)\sum_{i=1}^{N_i}(Q_i/p_iM^2)(\overline{Y}_i/\overline{Y})^2(1+\mathrm{CV}_{yi}^2/m_0)$, and

$\text{CV}_{yi}^2 = S_{yi}^2 / \bar{Y}_i^2$ denotes the within-cluster relative variance of the y-variable. Similarly,

$$\text{Deft}_p^2(\hat{\bar{Y}}) \cong 1 + (m_0 - 1)\tau + W_d^*, \tag{4.21}$$

where $W_d^* = W_d / V_{\text{srswr}}(\hat{Y}_{\text{srs}}) = (m_0 / \text{CV}_y^2)\sum_{i=1}^N (Q_i / p_i M^2)$ $(\bar{D}_i / \bar{Y})^2 (1 + \text{CV}_{di}^2 / m_0)$, and $\bar{D}_i$ and CV_{di}^2 are defined with the transformed variable $d\,(d_{ij} = y_{ij} - \bar{Y})$ analogously to $\bar{Y}_i$ and CV_{yi}^2, respectively. (Detailed derivations of expressions (4.19) and (4.21) are available from the authors.) For the case with $m_i = m_0$ for all i, the difference in the design effects given in (4.19) and (4.21) reduces to (3.7) or (4.16). There is no contribution from the second stage sampling to the difference.

Coming back to Thomsen *et al*. (1986) who studied the effect of using an outdated measure of size on the variance, the above discussion on $\hat{Y}$ parallels with their discussion. The only difference is that they assumed a without-replacement sampling scheme at the second stage. Note, however, that the definition of τ in Thomsen *et al*. (1986) is slightly different from (4.20) and from δ in section 4.2. However, there is a close connection between them. To see this, let us write the τ as a function of some quantities b_i's associated with PSUs as follows:

$$\tau(b_i) = \frac{(N-1)S_{yB}^2 - \sum_{i=1}^N b_i S_{yi}^2}{(N-1)S_{yB}^2 + \sum_{i=1}^N (M_i - 1)S_{yi}^2}.$$

Then the τ in Thomsen *et al*. (1986) is obtained with $b_i = 1$, the τ in example 4.5 with $-1/(m_0 - 1)$, and δ in section 4.2 with $(M_i - 1)/\{\sum_{i=1}^N (M_i - 1)/(N-1)\}$. Equating Kish's formula (4.18) for $\bar{Y}$ to (4.19) for $\hat{Y}$, they obviously overlooked that the design effects for $\hat{Y}$ and $\hat{\bar{Y}}$ can be very different.

For more general cases, Kish (1987) proposed the following popular formula for $\hat{\bar{Y}}$:

$$\text{Deft}_{\text{Kish}}^2(\hat{\bar{Y}}) = \frac{n\sum_{g=1}^G n_g w_g^2}{\left(\sum_{g=1}^G n_g w_g\right)^2}[1 + \rho(\bar{m} - 1)]$$
$$= (1 + \text{cv}_w^2)[1 + \rho(\bar{m} - 1)].$$

This was obtained by applying (4.5) (or (4.6)) and (4.18) recursively to incorporate the effects of both clustering and unequal weights. Gabler, Haeder and Lahiri (1999) justified the above formula for $\hat{\bar{Y}}$ using a superpopulation model defined for the cross-classification of N clusters and G weighting classes. However, the difference between the design effects for $\hat{\bar{Y}}$ and $\hat{Y}$ cannot be exposed by such a model-based approach, since y_k is treated as a random variable while w_k as fixed. Under this approach, $\text{Deft}_p^2(\hat{Y})$ differs from $\text{Deft}_p^2(\hat{\bar{Y}})$ only by a factor of $(\hat{M}/M)^2$, although the actual difference can be much more pronounced as we have showed in this paper (*e.g*., expressions (3.7) and (4.23)).

4.5 More General Cases

Weighting survey data involves not only sampling weights but also various weighting adjustments such as post-stratification, raking, and nonresponse compensation. We consider these general cases here.

We can rewrite the first-order Taylor approximation to the weighted mean estimator $\hat{\bar{Y}} = \hat{Y}/\hat{M}$ given in (3.2) as $(\hat{Y} - Y)/Y \cong (\hat{\bar{Y}} - \bar{Y})/\bar{Y} + (\hat{M} - M)/M$. Taking variance on both sides,

$$\text{CV}_p^2(\hat{Y}) \cong \text{CV}_p^2(\hat{\bar{Y}}) + \text{CV}_p^2(\hat{M})$$
$$+ 2R_p(\hat{\bar{Y}}, \hat{M})\,\text{CV}_p(\hat{\bar{Y}})\text{CV}_p(\hat{M}), \tag{4.22}$$

where $\text{CV}_p^2(\hat{Y}), \text{CV}_p^2(\hat{\bar{Y}}), \text{CV}_p^2(\hat{M})$ are the relative variances of $\hat{Y}, \hat{\bar{Y}}$, and $\hat{M}$ respectively, and $R_p(\hat{\bar{Y}}, \hat{M})$ is the correlation coefficient of $\hat{\bar{Y}}$ and $\hat{M}$ with respect to the complex sampling design p and any weighting adjustments. Since the relative variances of simple sample total and mean $\hat{Y}_{\text{srs}}$ and $\bar{y}_{\text{srs}}$ are $\text{CV}_{\text{srswr}}^2(\hat{Y}_{\text{srs}}) = \text{CV}_{\text{srswr}}^2(\bar{y}_{\text{srs}}) = m^{-1}\text{CV}_y^2$ under srswr of size m, it follows from (4.22) that

$$\text{Deft}_p^2(\hat{Y}) \cong \text{Deft}_p^2(\hat{\bar{Y}})$$
$$+ 2R_p(\hat{\bar{Y}}, \hat{M})\nabla_p(y)\text{Deft}_p(\hat{\bar{Y}}) + \nabla_p^2(y), \tag{4.23}$$

where $\nabla_p(y) = \text{CV}_p(\hat{M})/\text{CV}_{\text{srswr}}(\bar{y}_{\text{srs}})$ is nonnegative. As an illustration, consider a binary variable y, where $\text{CV}_y^2 \cong (1 - \bar{Y})/\bar{Y}$ and, thus, $\nabla_p(y)$ can be arbitrarily large as $\bar{Y}$ approaches 1 or small as $\bar{Y}$ approaches zero assuming $\text{CV}_p(\hat{M}) \neq 0$. When $\nabla_p(y)$ is near zero, the two design effects are nearly equal. Otherwise, one is larger than the other depending on the values of $\nabla_p(y)$ and $R_p(\hat{\bar{Y}}, \hat{M})$. When the sampling weights are benchmarked to the known population size M, $\hat{Y}$ and $\hat{\bar{Y}}$ have the same design effect since $\hat{M} = M$ and $\text{CV}_p(\hat{M}) = 0$. In this case, $\hat{\bar{Y}}$ is not affected by the benchmarking but $\hat{Y} = M\hat{\bar{Y}}$, which is a ratio estimator. Note that poststratification or raking procedures may be used if population size information is available at subpopulation level and we also get equivalent design effects. In general, however, we have $\text{Deft}_p^2(\hat{Y}) \geq \text{Deft}_p^2(\hat{\bar{Y}})$ if

$$R_p(\hat{\bar{Y}}, \hat{M}) \geq -\frac{1}{2}\frac{\nabla_p(y)}{\text{Deft}_p(\hat{\bar{Y}})} \quad \text{or}$$

$$R_p(\hat{\bar{Y}}, \hat{M}) \geq -\frac{1}{2}\frac{\text{CV}_p(\hat{M})}{\text{CV}_p(\hat{\bar{Y}})}, \tag{4.24}$$

and vice versa.

It is illuminating to look at some specific situations. For example, if $R_p(\hat{\bar{Y}}, \hat{M}) \geq 0$, then $\text{Deft}_p^2(\hat{Y}) > \text{Deft}_p^2(\hat{\bar{Y}})$, however, a negative correlation (*i.e.*, $R_p(\hat{\bar{Y}}, \hat{M}) < 0$) doesn't necessarily lead to $\text{Deft}_p^2(\hat{Y}) \leq \text{Deft}_p^2(\hat{\bar{Y}})$. For a special case of $R_p(\hat{\bar{Y}}, \hat{M}) = 0$, the difference is given by

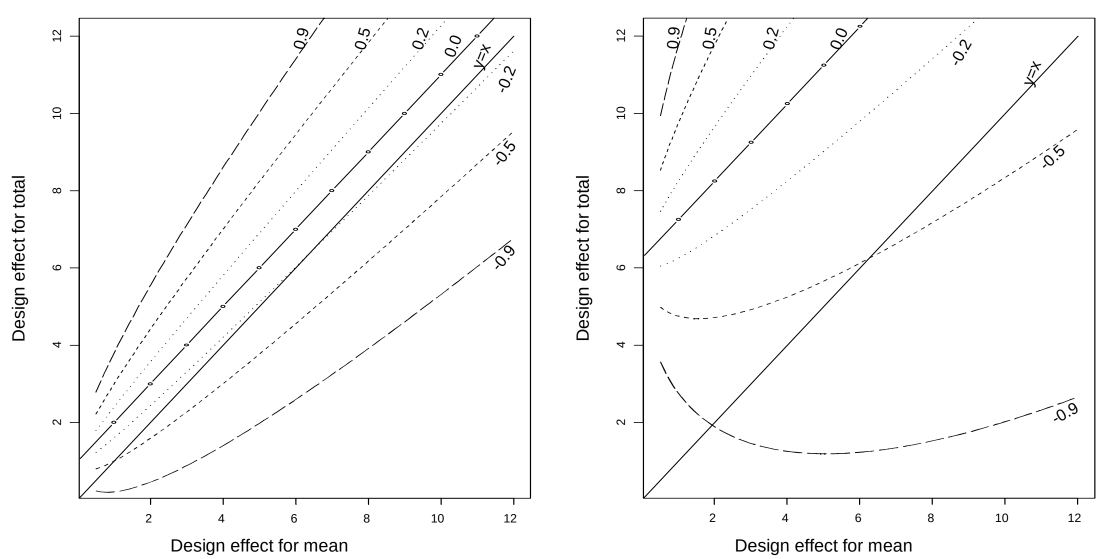
$$\text{Deft}_p^2(\hat{Y}) - \text{Deft}_p^2(\hat{\bar{Y}}) \cong \frac{\text{CV}_p^2(\hat{M})}{\text{CV}_{\text{srswr}}^2(\bar{y}_{\text{srs}})}. \qquad (4.25)$$

Figure 1 shows graphically the relation between the two design effects. The expression in (4.23) is plotted for some fixed values of $R_p(\hat{\bar{Y}}, \hat{M})$ and $\nabla_p(y)$. The solid line passing through the origin which represents equal design effects is the reference line. As the graphs show, the comparison is not clear-cut. When $R_p(\hat{\bar{Y}}, \hat{M}) < 0$, $\text{Deft}_p^2(\hat{Y}) \geq \text{Deft}_p^2(\hat{\bar{Y}})$ for small $\text{Deft}_p^2(\hat{\bar{Y}})$ but the relation flips over as $\text{Deft}_p^2(\hat{\bar{Y}})$ grows larger.

Hansen *et al.* (1953, Vol. I, pages 338 – 339) indicated that $R_p(\hat{\bar{Y}}, \hat{M})$ would often be close to 0. Under this situation, expression (4.25) is also written as $\text{Deft}_p^2(\hat{Y}) \cong \text{Deft}_p^2(\hat{\bar{Y}})\,[1 + \text{CV}_p^2(\hat{M}) / \text{CV}_p^2(\hat{\bar{Y}})]$, from which we get $\text{Deft}_p^2(\hat{Y}) \geq \text{Deft}_p^2(\hat{\bar{Y}})$. This special case was studied by Jang (2001). However, this doesn't seem necessary as can be seen in the following example.

Example 4.6 To illustrate the relationship between the design effects for $\hat{\bar{Y}}$ and $\hat{Y}$, we used a data set for the adults collected from the U.S. Third National Health and Nutrition Examination Survey (NHANES III), which is given as a demo file in WesVar version 4.0. NHANES III is a nationwide large-scale medical examination survey based on a stratified multistage sampling design, for which the Fay's modified balance repeated replication (BRR) method was employed for variance estimation. (See Judkins 1990 for more details on Fay's method.) We used only 19,793 records with complete responses to those characteristics listed in Table 1. Note that the weight in the demo file is different from the NHANES III final weight that was obtained by poststratification. For more detailed information on the demo file, see Westat (2001).

Table 1 presents the design effects for $\hat{\bar{Y}}$ and $\hat{Y}$, and component terms of (4.23) for the selected characteristics. Note that $\nabla_p(y)$ monotonically decreases in CV_y given that $m = 19{,}793$ and $\text{cv}_p(\hat{M}) = 3.2\%$. Although $\nabla_p(y)$ tends to be the determinant factor in the difference of the design effects, $R_p(\hat{\bar{Y}}, \hat{M})$ can be important when it is negative. For example, for two race/ethnicity characteristics, African American and Hispanic, the negative values, –0.67 and –0.24 of $R_p(\hat{\bar{Y}}, \hat{M})$ were responsible for $\text{Deft}_p^2(\hat{Y}) < \text{Deft}_p^2(\hat{\bar{Y}})$. Some design effects for $\hat{Y}$ are huge. This is not the case with the NHANES III poststratified final weights, with which $\hat{Y}$ and $\hat{\bar{Y}}$ have the same design effect. This illustrates the importance of benchmarking weight adjustments for total estimates.



(a) $\nabla_p(y) = 1.0$ (b) $\nabla_p(y) = 2.5$

Figure 1. Plots of $\text{Deft}_p^2(\hat{Y})$ versus $\text{Deft}_p^2(\hat{\bar{Y}})$ for (a) $\nabla_p(y) = 1.0$ (b) $\nabla_p(y) = 2.5$. The solid line corresponds to $\text{Deft}_p^2(\hat{\bar{Y}}) = \text{Deft}_p^2(\hat{Y})$. Other lines correspond to $R_p(\hat{\bar{Y}}, \hat{M}) = -0.9,\ -0.5,\ -0.2,\ 0,\ 0.2,\ 0.5,\ 0.9$, respectively.

Table 1
Comparison of the design effects for the weighted total and mean using a subset of the adult data file from the U.S. Third National Health and Nutrition Examination Survey (NHANES III)

		Mean			Total						
Characteristic		Estimate	Deft^2	cv	Estimate	Deft^2	cv	cv$_y$	$r_p(\hat{\hat{Y}},\hat{M})$	$\nabla_p(y)$	$-\dfrac{\mathrm{cv}_p(\hat{M})}{2\mathrm{cv}_p(\hat{\hat{Y}})}$
Has smoked 100+ cigarettes in life?	Yes	0.53	4.13	0.014	98,397,795	31.31	0.038	0.944	0.20	4.83	−0.58
Has diabetes?	Yes	0.05	1.75	0.040	9,783,307	1.92	0.042	4.246	−0.34	1.07	−0.31
	No	0.95	1.75	0.002	176,341,218	393.47	0.033	0.236	0.34	19.35	−5.53
Has hypertension/	Yes	0.23	3.42	0.024	42,939,866	7.96	0.037	1.826	−0.18	2.50	−0.37
high blood pressure?	No	0.77	3.42	0.007	143,184,660	78.44	0.034	0.548	0.18	8.32	−1.22
Race/Ethnicity	AFRICAN AMERICAN*	0.12	7.64	0.054	21,567,028	4.21	0.040	2.762	−0.67	1.65	−0.11
	HISPANIC*	0.05	6.70	0.079	9,550,326	6.48	0.078	4.300	−0.24	1.06	−0.08
Gender	MALE	0.48	1.40	0.009	88,725,967	19.18	0.033	1.048	−0.11	4.35	−1.55
	FEMALE	0.52	1.40	0.008	97,398,559	25.39	0.034	0.954	0.11	4.77	−1.70
Number of cigarettes smoked per day	–	5.25	6.42	0.037	977,225,826	10.51	0.047	2.044	−0.09	2.23	−0.17
Population Size	–	–	–	–	186,124,526	–	0.032	–	–	–	–

Note: * denotes the cases where the design effect for $\hat{Y}$ is smaller than that for $\hat{\hat{Y}}$.

5. Conclusion

We studied the design effects of the two most widely used estimators for the population mean and total in sample surveys under various with-replacement sampling schemes. We do not think the employment of with-replacement sampling is necessarily a serious limitation because we can see things more clearly without muddling the math with probably unnecessary complications with without-replacement sampling schemes. Furthermore, the effect of the finite population correction is largely canceled out in our formulation of the design effect and so the results are quite comparable with traditional design effects for without-replacement sampling. Therefore, our findings should be useful in practice. We summarize our key findings below.

Kish's well-known approximate formulae for the design effect for (ratio type) weighted mean estimators are not easily generalized in their form and concepts to more general problems, especially weighted total estimators contrary to what many people would perceive. In fact, $\hat{\hat{Y}}$ and $\hat{Y}$ often have very different design effects unless the sampling design is self-weighting or the sampling weights are benchmarked to the known population size. In addition, the design effect is in general not free from the distribution of the study variable even for the mean estimator, let alone the total estimator. Furthermore, the correlation of the study variable with the weights used in estimation can be an important factor in determining the design effect. Therefore, apart from its original intention, the design effect measures not only the effect of a complex sampling design on a particular statistic but also the effects of the distribution of the study variable and its relations to the sampling design on the statistic. As complex survey software packages routinely produce the design effect, it seems appropriate to warn the user of the packages of these rather obscure facts about the design effect.

Acknolwedgement

The authors thank Louis Rizzo at Westat, an associate editor, and two referees for their helpful comments and suggestions on an earlier version of this paper.

References

Apostol, T.M. (1974). *Mathematical Analysis.* 2nd Ed. Reading, MA: Addison-Wesley.

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, 400-405.

Cochran, W.G. (1977). *Sampling Techniques*. 3rd Ed. New York: John Wiley & Sons, Inc.

Cornfield, J. (1951). Modern methods in the sampling of human populations. *American Journal of Public Health*, 41, 654-661.

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, 105-106.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory*. Vol. I, New York: John Wiley & Sons, Inc.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory*, Vol. II, New York: John Wiley & Sons, Inc.

Judkins, D.R. (1990). Fay's method for variance estimation, *Journal of Official Statistics*, 6, 223-239.

Kish, L. (1965). *Survey Sampling.* New York: John Wiley & Sons, Inc.

Kish, L. (1987). Weighting in Deft^2. *The Survey Statistician.* June 1987.

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

Kish, L. (1995). Methods for design effects. *Journal of Official Statistics*, 11, 55-77.

Jang, D. (2001). On procedures to summarize variances for survey estimates. *Proceedings of the Survey Research Methods of the American Statistical Association.* In CD-ROM.

Lehtonen, R., and Pahkinen, E.J. (1995). *Practical Methods for Design and Analysis of Complex Surveys*. New York: John Wiley & Sons, Inc.

Lohr, S.L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Brooks/Cole.

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. In CD-ROM.

Park, I., and Lee, H. (2002). A revisit of design effects under unequal probability sampling. *The Survey Statistician*, 46, 23-26.

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

Spencer, B.D. (2000). An approximate design effect for unequal weighting when measurements may correlate with selection probabilities. *Survey Methodology*, 26, 137-138.

Thomsen, I., Tesfu, D. and Binder, D.A. (1986). Estimation of Design Effects and Intraclass Correlations When Using Outdated Measures of Size. *International Statistical Review*, 54, 343-349.

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

Yamane, T. (1967). *Elementary Sampling Theory*. New Jersey: Prentice-Hall.