Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

On the definition and interpretation of interviewer variability for a complex sampling design

by Siegfried Gabler and Partha Lahiri

June 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2009
Vol. 35, No. 1, pp. 85-99
Statistics Canada, Catalogue No. 12-001-X

On the definition and interpretation of interviewer variability for a complex sampling design

Siegfried Gabler and Partha Lahiri [1]

Abstract

Interviewer variability is a major component of variability of survey statistics. Different strategies related to question formatting, question phrasing, interviewer training, interviewer workload, interviewer experience and interviewer assignment are employed in an effort to reduce interviewer variability. The traditional formula for measuring interviewer variability, commonly referred to as the interviewer effect, is given by $ieff := deff\_int = 1 + (\bar{n}_{\text{int}} - 1)\rho_{\text{int}}$, where ρ_{int} and $\bar{n}_{\text{int}}$ are the intra-interviewer correlation and the simple average of the interviewer workloads, respectively. In this article, we provide a model-assisted justification of this well-known formula for equal probability of selection methods (epsem) with no spatial clustering in the sample and equal interviewer workload. However, spatial clustering and unequal weighting are both very common in large scale surveys. In the context of a complex sampling design, we obtain an appropriate formula for the interviewer variability that takes into consideration unequal probability of selection and spatial clustering. Our formula provides a more accurate assessment of interviewer effects and thus is helpful in allocating more reasonable amount of funds to control the interviewer variability. We also propose a decomposition of the overall effect into effects due to weighting, spatial clustering and interviewers. Such a decomposition is helpful in understanding ways to reduce total variance by different means.

Key Words: Interviewer effect; Interviewer workloads; Intra-interviewer correlation; Spatial clustering; Unequal weighting.

1. Introduction

A major source of measurement errors in surveys is due to the interviewer. This fact was recognized as early as 1929 by Rice and later by many survey researchers. Factors such as the quality of questionnaire design and the interviewer can influence the interviewer effects on survey statistics.

The interviewer can introduce homogeneity in survey data, which generally reduces the effective sample size and thereby increases the total variance of a survey estimator. The within interviewer homogeneity has been traditionally measured by the intra-interviewer correlation coefficient ρ_{int}. The magnitude of the intra-interviewer correlation was studied by many researchers, mostly in the context of telephone surveys without any spatial clustering effects (Kish 1962; Gray 1956; Hanson and Marks 1958; Tucker 1983; Groves and Magilavy 1986; Heeb and Gmel 2001, and others). Researchers have argued that the nature of the survey items may affect the value of ρ_{int}. Attitude items and complex factual items are considered more sensitive to the intra-interviewer correlation than simple factual items are (Collins and Butcher 1982; Feather 1973; Fellegi 1964; Gray 1956; Hansen, Hurwitz and Bershad 1961). According to Groves (1989), values above 0.1 are seldom observed. See Schnell and Kreuter (2005) for further discussion on this issue.

As noted by several researchers, the standard interviewer effect formula $1 + (\bar{n}_{\text{int}} - 1)\rho_{\text{int}}$ suggests that even with a small intra-interviewer correlation, the interviewer effect could be substantial simply due to a high average interviewer workload. For example, when $\rho_{\text{int}} = 0.01$ and $\bar{n}_{\text{int}} = 70$ we have $ieff = 1.69$ (Schnell and Kreuter 2005). Note that a high average interviewer workload (*e.g.*, between 60 and 70) is very common in telephone surveys (Tucker 1983; Groves and Magilavy 1986). For the European Social Survey, Philippens and Loosveldt (2004) provided box plots of the intra-interviewer correlations and the interviewer workloads for 18 participating countries.

The interviewer effect or variance is generally defined as the inflation to the total variance caused solely by the interviewers. For an epsem design with equal interviewer workload, the interviewer variance for the sample mean is simply given by $1 + (n_{\text{int}} - 1)\rho_{\text{int}}$, where n_{int} is the common interviewer workload. For complex surveys with unequal interviewer workload, survey researchers frequently use a simple modification of this formula where the common interviewer workload is replaced by the average interviewer workload, *i.e.*, the formula $1 + (\bar{n}_{\text{int}} - 1)\rho_{\text{int}}$. In Section 2, we argue that this standard formula $1 + (\bar{n}_{\text{int}} - 1)\rho_{\text{int}}$ cannot be interpreted as an inflation to the total variance caused by the interviewers even for an epsem design with unequal interviewer workload. In Sections 2-4, we observe that the interviewer variance definition depends

1. Siegfried Gabler, GESIS, P.O. Box 12 21 55, 68072 Mannheim, Germany. E-mail: siegfried.gabler@gesis.org; Partha Lahiri, University of Maryland, College Park, U.S.A. E-mail: plahiri@survey.umd.edu.

on the nature of the complex sampling design and also on the interviewer workload assignment. In this paper, we provide appropriate definitions of the interviewer variance in different survey scenarios. A reliable definition of the interviewer variance is helpful in determining actions that need to be taken in order to reduce interviewer variability. This paper is foremost applicable to the planning of surveys rather than analyzing survey data. In other words, in this paper we have concentrated on the definitions and interpretation of the interviewer variability and not on estimating it from a given survey.

In Section 2, we consider an epsem design with no spatial clustering and provide a model-assisted interpretation of *ieff*. We show that for the equal interviewer workload *ieff* is simply the ratio of the variances of the sample mean under a correlated model that accounts for the homogeneity of the observations collected by the same interviewer and a simple uncorrelated model that fails to account for such homogeneity. Thus, multiplying the variance of the sample mean for simple random sampling by the *ieff* one can obtain the total variance of the sample mean that incorporates both the sampling and the interviewer variability. This is a very intuitive interpretation of *ieff* and complements the model-assisted justification given earlier by Kish (1962). In this section, we also show that for an epsem design *ieff* is lower than the model-assisted interviewer effect formula if the interviewer workload varies and the intra-interviewer correlation is positive. Thus, the survey designer who uses *ieff* would give less effort to control interviewer variability than is really needed. In this situation, an appropriate interviewer effect formula can be obtained from *ieff* when a weighted average interviewer workload is used in place of the usual simple average.

In Section 3, we entertain the possibility of unequal weighting but no spatial clustering. We obtain a model-assisted interpretation for *ieff* if and only if the respondents interviewed by the same interviewer share the same sampling weight and the interviewer workload is inversely proportional to the square of the common weight for the interviewer. Interestingly, unlike the epsem design, equal interviewer workload does not necessarily guarantee a model-assisted interpretation for *ieff*. When there is an equal interviewer workload and there is at least one interviewer for which the respondents do not all share the same sampling weight, we show that *ieff* is always higher than the model-assisted formula. We also point out the factors that cause the difference between these two formulae. These results have a practical relevance in terms of saving survey costs. To be specific, the survey designer who uses *ieff* is likely to allocate more funds to control interviewer variability than is really needed. We have also cited some situations where *ieff* could have an under-estimation problem and thus survey designers who use *ieff* could give less emphasis to control the interviewer effects. Our formula provides a more accurate assessment of interviewer variability and thus is helpful in the allocation of more reasonable amount of funds to control the interviewer variability. Furthermore, the change in planning formulae will affect the sample size.

In many large scale sample surveys, due to various organizational and financial reasons such as the absence of a general population register or to reduce the overall survey costs, a multi-stage clustered sampling design is considered to be a cost-efficient alternative to simple random sampling. Under a multi-stage clustered sampling design, respondents who live in close spatial proximity of each other get selected. Respondents living in the same spatial cluster tend to share similar attitudes because of their similar socio-economic background and hence increase the internal homogeneity of the survey data. This spatial homogeneity violates the iid (independently identically distributed) assumption frequently used in standard statistical inferential procedures and so does the clustering within the interviewers. This fact has been recognized by many survey researchers and adjustments to various statistical procedures and the related software issues have been addressed in the literature (see Rao and Scott 1984; Skinner, Holt and Smith 1989; Biemer and Trewin 1997; Chambers and Skinner 2003; among others). In Section 4, we present a new definition of the interviewer variability in the presence of unequal weighting and spatial clustering. In the presence of spatial clustering, we argue that *ieff* generally has a tendency to overestimate the interviewer variability. Thus for complex surveys involving spatial clustering, *ieff* may unnecessarily give a false alarm regarding the magnitude of the interviewer variability.

In Section 5, we discuss the effects due to the combined effects of weighting, spatial clustering and the interviewer. The formula for overall effects offers an accurate determination of the sample size at the planning stage. We provide a nice factorization of the overall effects into the effects due to weighting, clustering and interviewer. Such a decomposition of the overall effects can be useful in understanding ways to reduce the total variance by different means. In discussing Verma, Scott and O'Muircheartaigh (1980), Hedges mentioned the need for such an overall effect formula. We generalize a formula earlier proposed by Davis and Scott (1995) to a non-epsem design and for a general correlation model valid for both discrete and continuous data. We present proofs of all the technical results in the Appendix.

2. EPSEM design with no spatial clustering

Let y_{ik} denote the observation obtained from the k^{th} respondent interviewed by the i^{th} interviewer $(i = 1, ..., I;\ k = 1, ...\, n_i)$. Define $n = \sum_{i=1}^{I} n_i$, the total sample size, $\bar{y} = 1/n \sum_{i=1}^{I} \sum_{k=1}^{n_i} y_{ik}$, the unweighted sample mean, and $\bar{n}_{\text{int}}(\mathbf{a}) = \sum_{i=1}^{I} a_i n_i$, a weighted average of the interviewer workload, where a_i is an arbitrary weight attached to the i^{th} interviewer workload and $\mathbf{a} = (a_1, ..., a_I)$.

We shall first provide a model-assisted justification of the traditional interviewer effect formula, *i.e.*, $ieff = 1 + (\bar{n}_{\text{int}} - 1)\rho_{\text{int}}$, where $\bar{n}_{\text{int}}$ is the unweighted average of interviewer workload. Note that $\bar{n}_{\text{int}} = \bar{n}_{\text{int}}(\mathbf{a}_0)$, with $\mathbf{a}_0 = (a_{01}, ..., a_{0I})$, $a_{0i} = 1/I$ and $ieff = ieff(\mathbf{a}_0)$. Using Result 1 given in the Appendix, we get

$$ieff(\mathbf{a}_1) = \frac{\text{Var}_{M_2}(\bar{y})}{\text{Var}_{M_1}(\bar{y})} = 1 + [\bar{n}_{\text{int}}(\mathbf{a}_1) - 1]\rho_{\text{int}},$$

where $\mathbf{a}_1 = (a_{11}, ..., a_{1I})$, with $a_{1i} = n_i / n$. In the above, $\text{Var}_{M_1}(\bar{y})$ and $\text{Var}_{M_2}(\bar{y})$ are the variances of $\bar{y}$ under the following two models, respectively,

$$M_1\text{: Cov}(y_{ik},\ y_{i'k'}) = \begin{cases} \sigma^2 & \text{if } i = i',\ k = k', \\ 0 & \text{otherwise,} \end{cases}$$

$$M_2\text{: Cov}(y_{ik},\ y_{i'k'}) = \begin{cases} \sigma^2 & \text{if } i = i',\ k = k', \\ \rho_{\text{int}}\sigma^2 & \text{if } i = i',\ k \neq k', \\ 0 & \text{otherwise.} \end{cases}$$

Note that unlike model M_1, model M_2 introduces homogeneity of the observations collected by the same interviewer.

Remark 2.1: It follows from the corollary to Result 1, given in the Appendix, that for $\rho_{\text{int}} > 0$, $ieff(\mathbf{a}_1) = ieff$ if and only if $n_i = n/I$ for all i, *i.e.*, if and only if each interviewer has the same workload. For the balanced case, Kish (1962) provided a model-assisted justification of $ieff$ using a linear mixed model, which is a special case of M_2. For the unbalanced case, it is interesting to note the similarity between the interviewer variability formula $ieff(\mathbf{a}_1)$ and the design effects formula given in (A3) of Holt in discussing Verma *et al*. (1980).

Remark 2.2: It follows from the corollary to Result 1 that if $\rho_{\text{int}} > 0$ and n_i's are not equal then $ieff(\mathbf{a}_1) > ieff$.

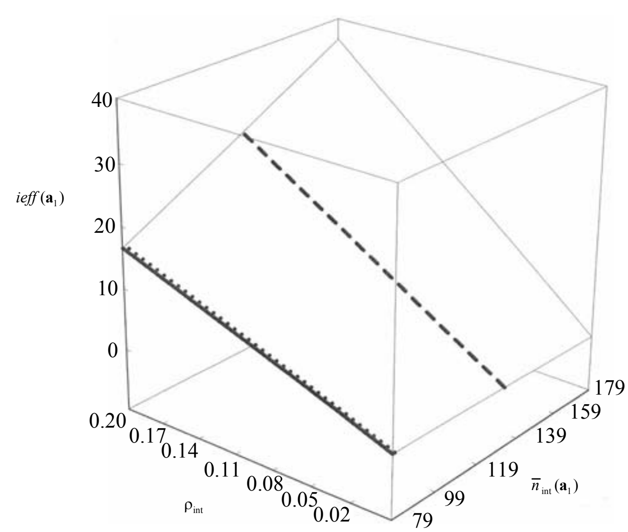
In the following example, we demonstrate the extent to which $ieff(\mathbf{a}_1)$ and $ieff$ could differ for different interviewer workload patterns.

Example 1: In Table 1, we consider three different workload assignments for ten interviewers, each with $n = 790$. Case A) represents the most variable workload assignment with a standard deviation = 68.3; Case B) is nearly balanced with a standard deviation = 9.5; Case C) corresponds to the equal interviewer assignment.

Table 1
Three different interviewer workload assignments (Example 1)

	Interviewer workload pattern		
Interviewer	**A)**	**B)**	**C)**
1	4	70	79
2	10	70	79
3	20	70	79
4	34	70	79
5	52	70	79
6	74	88	79
7	100	88	79
8	130	88	79
9	164	88	79
10	202	88	79
n	790	790	790
$\bar{n}_{\text{int}}(\mathbf{a}_1)$	132	80	79

Let $ieff(\mathbf{a}_{1;A})$, $ieff(\mathbf{a}_{1;B})$, and $ieff(\mathbf{a}_{1;C}) = ieff$ denote $ieff(\mathbf{a}_1)$, the model-assisted interviewer variance formula corresponding to the cases A, B and C, respectively. For $\rho_{\text{int}} > 0$ the function $ieff(\mathbf{a}_1)$ is Schur-convex, which explains the fact $ieff(\mathbf{a}_{1;A}) \geq ieff(\mathbf{a}_{1;B}) \geq ieff(\mathbf{a}_{1;C}) = ieff$. Figure 1 provides the values of the interviewer variance obtained from the standard formula (*i.e.*, $ieff$) and our model-assisted interview variance formula for all combinations of the two influencing factors, *i.e.*, weighted average of interviewer workload and the intra-interviewer correlation. From Figure 1, it is interesting to note that $ieff$ could underreport by about 100%.



Interviewer A: Dashes,
Interviewer B: Spaced dots and
Interviewer C: Solid line

Figure 1 A graph of $ieff(\mathbf{a}_1)$ *vs.* ρ_{int} for different $\bar{n}_{\text{int}}(\mathbf{a}_1)$

3. Unequal weighting with no spatial clustering

In this section, we consider the situation when we have unequal weights. Let w_{ik} be the survey weight attached to the k^{th} respondent interviewed by the i^{th} interviewer. In this situation, a weighted mean $\bar{y}_w = \sum_i \sum_k w_{ik} y_{ik} / \sum_i \sum_k w_{ik}$ is a popular estimator of the finite population mean (See Brewer 1963; Hájek 1971) and the model-assisted interviewer variance formula is given by

$$ieff_w = \frac{\text{Var}_{M_2}(\bar{y}_w)}{\text{Var}_{M_1}(\bar{y}_w)} = 1 + \rho_{\text{int}} \left(\frac{\sum_i \left(\sum_k w_{ik} \right)^2}{\sum_i \sum_k w_{ik}^2} - 1 \right).$$

See Result 1 given in the Appendix.

Define $\bar{w}_i = 1/n_i \sum_{k=1}^{n_i} w_{ik}$, the average survey weight for the i^{th} interviewer and $\sigma_i^2 = 1/n_i \sum_k w_{ik}^2 - \bar{w}_i^2$, the variance of the survey weights for the i^{th} interviewer. It can be shown that

$$ieff_w = 1 + \rho_{\text{int}} (\bar{n}_w - 1),$$

where

$$\bar{n}_w = \frac{\sum_i n_i^2 \bar{w}_i^2}{\sum_i n_i \bar{w}_i^2 + \sum_i n_i \sigma_i^2}.$$

Note that, in general, $ieff_w$ cannot be written in the form $ieff_w = 1 + \rho_{\text{int}} (\bar{n}_{\text{int}}(\mathbf{a}) - 1)$ with $\sum_i a_i = 1$.

Remark 3.1: From Result 2 in the Appendix, we have

$$ieff_w \le ieff(\mathbf{a}_2),$$

where

$$\mathbf{a}_2 = (a_{21}, \ldots, a_{2I}), \text{ with } a_{2i} = \frac{\sum_k w_{ik}^2}{\sum_i \sum_k w_{ik}^2}.$$

In the above, for $\rho_{\text{int}} > 0$, $ieff_w = ieff(\mathbf{a}_2)$ if and only if all σ_i^2 are zero. Thus, $ieff(\mathbf{a}_2)$ can be interpreted as a conservative interviewer variance.

Equality holds if and only if $w_{ik} = \bar{w}_i$ for all i and k in which case

$$ieff_w = ieff(\mathbf{a}_2^*),$$

where

$$\mathbf{a}_2^* = (a_{21}^*, \ldots, a_{2I}^*), \text{ with } a_{2i}^* = \frac{n_i \bar{w}_i^2}{\sum_i n_i \bar{w}_i^2}.$$

Thus, the formulae $ieff_w$ and $ieff(\mathbf{a}_2^*)$ are equivalent if and only if the survey weights are all the same for a given interviewer. One example of such a design is an epsem design for which we have

$$a_{2i}^* = \frac{n_i}{n}$$

and

$$ieff_w = ieff(\mathbf{a}_2^*) = ieff(\mathbf{a}_1).$$

Now we shall try to understand the factors that explain the difference between $ieff_w$ and $ieff$. To this end, define

$\bar{w} = 1/n \sum_{i=1}^{I} \sum_{k=1}^{n_i} w_{ik} = \sum_{i=1}^{I} n_i / n \; \bar{w}_i$, the average survey weight for all interviewers,

$SSB = \sum_{i=1}^{I} n_i (\bar{w}_i - \bar{w})^2$, the between interviewer sum of squares of the survey weights,

$SSW = \sum_{i=1}^{I} \sum_{k=1}^{n_i} (w_{ik} - \bar{w}_i)^2 = \sum_{i=1}^{I} n_i \sigma_i^2$, the within interviewer sum of squares of the survey weights,

$SST = SSB + SSW$, the total sum of squares of the survey weights,

$\tau_w = SSW/SST$, an indicator of the relative contribution of the within interviewer variability of survey weights to the total variability,

$CV_w = \sqrt{SST/n} \,/\, \bar{w}$, the coefficient of variation of the survey weights in the entire sample.

It can be shown that (see Result 4)

$ieff_w - ieff$

$$= \frac{\bar{n}_{\text{int}}}{SST + n\bar{w}^2} \left[\sum_{i=1}^{I} \left(\frac{n_i}{\bar{n}_{\text{int}}} - 1 \right) n_i \bar{w}_i^2 - SSW \right] \rho_{\text{int}} \quad (1)$$

$$= \frac{\bar{n}_{\text{int}}}{(1 + CV_w^{-2}) SST} \left[\sum_{i=1}^{I} \left(\frac{n_i}{\bar{n}_{\text{int}}} - 1 \right) n_i \bar{w}_i^2 - SSW \right] \rho_{\text{int}} \quad (2)$$

$$= \frac{\bar{n}_{\text{int}} \tau_w}{1 + CV_w^{-2}} \left(\frac{\sum_{i=1}^{I} \left(\frac{n_i}{\bar{n}_{\text{int}}} - 1 \right) n_i \bar{w}_i^2}{SSW} - 1 \right) \rho_{\text{int}}. \quad (3)$$

Remark 3.2: We can use formula (1) in any situation. For epsem designs, we have

$$ieff_w - ieff = \rho_{\text{int}} \frac{\bar{n}_{\text{int}}}{n} \sum_{i=1}^{I} \left(\frac{n_i}{\bar{n}_{\text{int}}} - 1 \right) n_i.$$

Note that an application of the Cauchy-Schwarz inequality suggests $ieff_w - ieff \ge 0$ with equality if and only if $n_i = n/I$ for all i.

Remark 3.3: We can use (2) if $SST \ne 0$, *i.e.*, if the design is not epsem. If $\rho_{\text{int}} > 0$, (2) implies

$$ieff_w - ieff \le 0 \quad \text{if and only if } \sum_{i=1}^{I}\left(\frac{n_i}{\overline{n}_{\text{int}}} - 1\right) n_i \overline{w}_i^2 \le SSW.$$

If high interviewer workload tends to be associated with small average survey weights and vice versa and $SSW \ne 0$, we can expect $ieff$ to be a conservative value of the actual interviewer variance $ieff_w$. In Example 2, c) and d), we have such a situation.

Now, we have $ieff_w = ieff$ if and only if $w_{ik} = \overline{w}_i$ (or, equivalently, $SSW = 0$) and $n_i \overline{w}_i^2 / \sum_i n_i \overline{w}_i^2 = 1/I$ for all i and k, *i.e.*, $ieff_w = ieff$ if and only if $w_{ik} = \overline{w}_i$ and $\overline{w}_i \propto 1/\sqrt{n_i}$ for all i and k.

Thus, for a non-epsem design, equal interviewer workload does not necessarily provide us a model-assisted interpretation for $ieff$. For example, if the survey weights vary within at least one interviewer, we will not have a model-assisted interpretation of $ieff$. Obviously, for an epsem design the two formulae are equivalent if and only if we have equal interviewer workload.

Remark 3.4: If the interviewer workload is the same for all interviewers, we have

$$ieff_w - ieff = -\frac{\overline{n}_{\text{int}} \tau_w}{1 + CV_w^{-2}} \rho_{\text{int}}$$

(assume $SST \ne 0$). Thus, $ieff$ is a conservative value of the actual interviewer effect $ieff_w$. Furthermore, $|ieff_w - ieff|$ is an increasing function of the common interviewer workload $\overline{n}_{\text{int}}$ and $\tau_w / (1 + CV_w^{-2})$ (for fixed CV_w^{-2}, the latter is an increasing function of τ_w). The same interviewer workload is given in Example 2 a).

Remark 3.5: We can use formula (3) if $SSW > 0$, *i.e.*, if there is at least one interviewer for which weights are not all equal.

Example 2.

Table 2 presents eight different combinations of $(n_i, \overline{w}_i, \sigma_i^2)$. The first combination assumes equal n_i values but unequal weights. The second combination assumes $\overline{w}_i^2 \propto \sigma_i^2$. The other six combinations show all possible ordering of $\overline{n}_{\text{int}}$, $\overline{n}_{\text{int}}(\mathbf{a}_1)$, $\overline{n}_w$, $\overline{n}_{\text{int}}(\mathbf{a}_2)$ and, therefore, $ieff$, $ieff(\mathbf{a}_1)$, $ieff_w$, $ieff(\mathbf{a}_2)$ taking into consideration that $ieff \le ieff(\mathbf{a}_1)$ and $ieff_w \le ieff(\mathbf{a}_2)$.

Table 2
Ordering of interviewer effects formulae for several parameter combinations (Example 2); in the last column $\rho_{\text{int}} = 0.01$

	n_i	$\overline{w}_i$	σ_i^2	$\overline{n}_{\text{int}}$	$\overline{n}_{\text{int}}(\mathbf{a}_1)$	$\overline{n}_w$	$\overline{n}_{\text{int}}(\mathbf{a}_2)$	Interviewer effects	$ieff / ieff_w$
a)	25	1.022	0.299	25	25	19.20	25	$ieff = ieff(\mathbf{a}_1) = ieff(\mathbf{a}_2) > ieff_w$	1.003
	25	1.036	0.375						
	25	0.998	0.276						
	25	0.945	0.260						
b)	10	1	1	25	30	15	30	$ieff_w < ieff < ieff(\mathbf{a}_1) = ieff(\mathbf{a}_2)$	1.007
	20	1	1						
	30	1	1						
	40	1	1						
c)	10	1	1	25	30	7.5	32.5	$ieff_w < ieff < ieff(\mathbf{a}_1) < ieff(\mathbf{a}_2)$	1.023
	20	1	2						
	30	1	3						
	40	1	4						
d)	10	1	4	25	30	10	26.7	$ieff_w < ieff < ieff(\mathbf{a}_2) < ieff(\mathbf{a}_1)$	1.015
	20	1	3						
	30	1	2						
	40	1	1						
e)	10	4	144	25	30	1.80	11.71	$ieff_w < ieff(\mathbf{a}_2) < ieff < ieff(\mathbf{a}_1)$	0.998
	20	2	9						
	30	0.333	0.555						
	40	0.250	0.125						
f)	10	0.333	0.025	25	30	31.82	35.26	$ieff < ieff(\mathbf{a}_1) < ieff_w < ieff(\mathbf{a}_2)$	1.015
	20	0.666	0.075						
	30	1	0.125						
	40	1.333	0.175						
g)	10	1	0.010	25	30	29.13	30.10	$ieff < ieff_w < ieff(\mathbf{a}_1) < ieff(\mathbf{a}_2)$	0.999
	20	1	0.020						
	30	1	0.030						
	40	1	0.040						
h)	10	1	0.004	25	30	29.94	29.99	$ieff < ieff_w < ieff(\mathbf{a}_2) < ieff(\mathbf{a}_1)$	0.998
	20	1	0.003						
	30	1	0.002						
	40	1	0.001						

In the example, $\sum_i n_i \bar{w}_i = n$. We now explain the eight different patterns.

a) Since all n_i are equal, $ieff = ieff(\mathbf{a}_1) = ieff(\mathbf{a}_2)$. Moreover, $ieff_w$ is smaller than the rest because of the fact that $\sigma_i^2 > 0$.

b) Since σ_i^2 are relatively large, $ieff_w < ieff$. Also, $\sigma_i^2 = c \cdot \bar{w}_i^2$ implies $ieff(\mathbf{a}_1) = ieff(\mathbf{a}_2)$.

c) Since σ_i^2 are relatively large, $ieff_w < ieff$. Moreover, since $\bar{w}_i^2 + \sigma_i^2$ and n_i are both increasing, we have $ieff(\mathbf{a}_1) < ieff(\mathbf{a}_2)$.

d) Since σ_i^2 are relatively large, $ieff_w < ieff$. Since $\bar{w}_i^2 + \sigma_i^2$ is decreasing and n_i is increasing, we have $ieff(\mathbf{a}_2) < ieff(\mathbf{a}_1)$.

e) Since σ_i^2 are relatively large, $ieff_w < ieff$. Also, $\bar{w}_i^2$ and σ_i^2 are decreasing and n_i is increasing implying $ieff(\mathbf{a}_2) < ieff(\mathbf{a}_1)$.

f) The fact that $\bar{w}_i^2$ and n_i are increasing implies that $ieff_w > ieff$; since σ_i^2 and n_i are both increasing, we have $ieff(\mathbf{a}_1) < ieff(\mathbf{a}_2)$.

g) Since $\bar{w}_i^2$ and n_i are increasing, we have $ieff_w > ieff$ and since σ_i^2 is increasing, we have $ieff(\mathbf{a}_1) < ieff(\mathbf{a}_2)$. Moreover, $ieff_w < ieff(\mathbf{a}_1)$ since σ_i^2 is smaller than that in f).

h) Since $\bar{w}_i^2$ and n_i are increasing, we have $ieff_w > ieff$ and since σ_i^2 is decreasing, we have $ieff(\mathbf{a}_2) < ieff(\mathbf{a}_1)$.

4. Unequal weighting and spatial clustering

In this section, we obtain an appropriate interviewer variance formula in the presence of spatial clustering and unequal probability of selection. Consider the situation when more than one interviewer work independently in the same psu and the respondents in each psu are randomly assigned to the interviewers. We shall assume that no interviewer works in more than one psu. Such a design was considered in Biemer and Stokes (1985). Now we shall separate the interviewer effect from psu effect (*i.e.*, spatial clustering) and unequal weighting. Let y_{pik} and w_{pik} be the observation and the associated survey weight for the k^{th} respondent in the p^{th} psu interviewed by the i^{th} interviewer ($p = 1, \ldots, P$; $i = 1, \ldots I_p$; $k = 1, \ldots, n_{pi}$). Let $n_p = \sum_{i=1}^{I_p} n_{pi}$ be the number of sampling units in psu p.

In this case, we use the following weighted average to estimate the finite population mean:

$$\bar{y}_w = \frac{\sum_{p=1}^{P}\sum_{i=1}^{I_P}\sum_{k=1}^{n_{pi}} w_{pik}\, y_{pik}}{\sum_{p=1}^{P}\sum_{i=1}^{I_P}\sum_{k=1}^{n_{pi}} w_{pik}}.$$

Define

$$ieff_{s,w} = \frac{\mathrm{Var}_{M_4}(\bar{y}_w)}{\mathrm{Var}_{M_3}(\bar{y}_w)},$$

where the suffixes s and w signify the presence of spatial clustering and unequal weighting. In the above, $\mathrm{Var}_{M_3}(\bar{y}_w)$ and $\mathrm{Var}_{M_4}(\bar{y}_w)$ are the variances of $\bar{y}_w$ under the following two models respectively

$$M_3\text{: } \mathrm{Cov}(y_{pik},\, y_{p'i'k'}) = \begin{cases} \sigma^2 & \text{if } p=p',\ i=i',\ k=k' \\ \rho_C \sigma^2 & \text{if } p=p',\ k\neq k' \\ 0 & \text{otherwise} \end{cases}$$

$$M_4\text{: } \mathrm{Cov}(y_{pik}, y_{p'i'k'}) = \begin{cases} \sigma^2 & \text{if } p=p',\ i=i',\ k=k' \\ \rho_C \sigma^2 & \text{if } p=p',\ i\neq i' \\ \rho\sigma^2 & \text{if } p=p',\ i=i',\ k\neq k' \\ 0 & \text{if } p\neq p' \end{cases}$$

In the above, ρ_C is the intra-psu correlation and ρ is the combined interviewer and psu intra-class correlation. Define $\rho_{\text{int}} = \rho - \rho_C$, intra-interviewer correlation. Usually, $\rho_{\text{int}} > 0$.

From Result 5, we have

$$ieff_{s,w} = 1 + \rho_{\text{int}} \frac{\bar{n}_{\text{int}}(\mathbf{A}_w) - 1}{1 + \rho_C(\bar{n}_{\text{psu}}(\mathbf{b}_w) - 1)},$$

where

$$\mathbf{A}_w = ((a_{wpi}))_{\substack{i=1,\ldots,I_p \\ p=1,\ldots,P}} \text{ and } a_{wpi} = \frac{n_{pi}\bar{w}_{pi}^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2}$$

with

$$\bar{w}_{pi} = \frac{1}{n_{pi}} \sum_{k}^{n_{pi}} w_{pik},$$

$$\bar{n}_{\text{int}}(\mathbf{A}_w) = \sum_{p=1}^{P}\sum_{i}^{I_P} a_{wpi} n_{pi} = \frac{\sum_{p=1}^{P}\sum_{i}^{I_P}\left(\sum_{k}^{n_{pi}} w_{pik}\right)^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2},$$

and

$$\mathbf{b}_w = (b_{wp})_{p=1,\ldots,P} \text{ and } b_{wp} = \frac{n_p \bar{w}_p^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2},$$

with

$$\bar{w}_p = \frac{1}{n_p}\sum_{i=1}^{I_p}\sum_{k}^{n_{pi}} w_{pik} = \frac{1}{n_p}\sum_{i=1}^{I_p} n_{pi}\bar{w}_{pi},$$

and

$$\bar{n}_{\text{psu}}(\mathbf{b}_w) = \sum_{p=1}^{P} b_{wp} n_p = \frac{\sum_{p=1}^{P}\left(\sum_{i}^{I_P}\sum_{k}^{n_{pi}} w_{pik}\right)^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2}.$$

Note that $\bar{n}_{\text{int}}(\mathbf{A}_w) \le \bar{n}_{\text{psu}}(\mathbf{b}_w)$ with equality if and only if $I_P = 1$. Also note that $\bar{n}_{\text{int}}(\mathbf{A}_w)$ is invariant of the allocation of the interviewers to the psu's while $\bar{n}_{\text{psu}}(\mathbf{b}_w)$ is not.

Remark 4.1: If $\rho_C = 0$ we get

$$ieff_{s,w} = 1 + \rho_{\text{int}}(\bar{n}_{\text{int}}(\mathbf{A}_w) - 1).$$

This formula is similar to $ieff_w$ given in Section 2. Thus, all the comments given in Remark 2.1 apply here. Note that $\bar{n}_{\text{int}}(\mathbf{A}_w)$, just like $\bar{n}_w$, cannot be generally written in the form $\bar{n}_{\text{int}}(\mathbf{A}_w) = \sum_{p=1}^{P}\sum_{i=1}^{I_p} a_{wpi} n_{pi}$ with $\sum_{p=1}^{P}\sum_{i=1}^{I_p} a_{wpi} = 1$; the same comment applies to $\bar{n}_{\text{psu}}(\mathbf{b}_w)$.

Remark 4.2: Define

$$\bar{n}_{\text{int}}(\mathbf{A}) = \sum_{p=1}^{P}\sum_{i}^{I_P} a_{pi} n_{pi}, \text{ where } \mathbf{A} = ((a_{pi})), \text{ with } a_{pi} = \frac{n_{pi}}{n},$$

and

$$\bar{n}_{\text{psu}}(\mathbf{b}) = \sum_{p=1}^{P} b_p n_p, \text{ where } \mathbf{b} = (b_1, \ldots, b_P) \text{ with } b_p = \frac{n_p}{n}.$$

If $\rho_C \ne 0$ but we have an epsem design, then we drop the suffix w in $ieff_{s,w}$. Note that

$$ieff_s = 1 + \rho_{\text{int}} \frac{\bar{n}_{\text{int}}(\mathbf{A}) - 1}{1 + \rho_C[\bar{n}_{\text{psu}}(\mathbf{b}) - 1]}$$

$$= 1 + \frac{\rho_{\text{int}}}{\rho_C} \cdot \frac{\bar{n}_{\text{int}}(\mathbf{A}) - 1}{\bar{n}_{\text{psu}}(\mathbf{b}) - 1} \cdot \frac{\rho_C(\bar{n}_{\text{psu}}(\mathbf{b}) - 1)}{1 + \rho_C(\bar{n}_{\text{psu}}(\mathbf{b}) - 1)}$$

so that

$$ieff_s < 1 + \frac{\rho_{\text{int}}}{\rho_C} \cdot \frac{\bar{n}_{\text{int}}(\mathbf{A}) - 1}{\bar{n}_{\text{psu}}(\mathbf{b}) - 1} < 1 + \frac{\rho_{\text{int}}}{\rho_C} \cdot \frac{\bar{n}_{\text{int}}(\mathbf{A})}{\bar{n}_{\text{psu}}(\mathbf{b})}.$$

It can be readily seen that the right side of the inequality increases with the ratios $\rho_{\text{int}} / \rho_C$ and

$$\frac{\bar{n}_{\text{int}}(\mathbf{A}) - 1}{\bar{n}_{\text{psu}}(\mathbf{b}) - 1}.$$

We have

$$ieff_s - ieff = \rho_{\text{int}} \frac{\frac{\bar{n}_{\text{int}}(\mathbf{A}) - 1}{\bar{n}_{\text{int}} - 1} - [1 + \rho_C(\bar{n}_{\text{psu}}(\mathbf{b}) - 1)]}{1 + \rho_C(\bar{n}_{\text{psu}}(\mathbf{b}) - 1)}(\bar{n}_{\text{int}} - 1).$$

Thus, for $\rho_{\text{int}} > 0$,

$ieff_s < ieff$ if and only if

$$Deff_s := 1 + \rho_C(\bar{n}_{\text{psu}}(\mathbf{b}) - 1) > \frac{\bar{n}_{\text{int}}(\mathbf{A}) - 1}{\bar{n}_{\text{int}} - 1},$$

i.e., if and only if the design effect due to the spatial clustering is larger than the ratio of the weighted average of the interviewer workload –1 and the average interviewer workload –1. If the interviewer workload is the same for all the interviewers, the right hand side of the inequality is 1 and so the inequality is always valid. It is interesting to note that $ieff \approx 4 \cdot ieff_s$ if $\rho_{\text{int}} = 0.1$, $\rho_C = 0.05$, $\bar{n}_{\text{psu}}(b) = 140$, and $\bar{n}_{\text{int}} = 70$.

Remark 4.3: In the general case, we have

$$ieff_{s,w} - ieff = \rho_{\text{int}}\left(\frac{\bar{n}_{\text{int}}(\mathbf{A}_w) - 1}{1 + \rho_C(\bar{n}_{\text{psu}}(\mathbf{b}_w) - 1)} - (\bar{n}_{\text{int}} - 1)\right).$$

Thus, for $\rho_{\text{int}} > 0$,

$ieff_{s,w} < ieff$ if and only if

$$Deff_{s,w} := 1 + \rho_C(\bar{n}_{\text{psu}}(\mathbf{b}_w) - 1) > \frac{\bar{n}_{\text{int}}(\mathbf{A}_w) - 1}{\bar{n}_{\text{int}} - 1},$$

i.e., if and only if

$$\rho_C > \frac{\bar{n}_{\text{int}}(\mathbf{A}_w) - \bar{n}_{\text{int}}}{(\bar{n}_{\text{int}} - 1)(\bar{n}_{\text{psu}}(\mathbf{b}_w) - 1)} =: \rho_C^*, \text{ say.}$$

In Example 2 (see Table 3), $ieff$ is a conservative value for $ieff_{s,w}$ for a) to e) if $\rho_C > 0$. The same holds for f) to h) if $\rho_C > 0.004$.

Table 3
Average interviewer workloads for several parameter combinations (Example 2); $ieff\,/\,ieff_{s,w}$ for $\rho_{\text{int}} = 0.01$ and $\rho_C = 0.02$

					$IA = (1,3)$				$IA = (2,2)$				$IA = (3,1)$			
	n_i	$\bar{w}_i$	σ_i^2	$\bar{n}_{\text{int}}$	$\bar{n}_{\text{int}}(A_w)$	$\bar{n}_{\text{psu}}(\mathbf{b}_w)$	ρ_C^*	$\frac{ieff}{ieff_{s,w}}$	$\bar{n}_{\text{int}}(A_w)$	$\bar{n}_{\text{psu}}(\mathbf{b}_w)$	ρ_C^*	$\frac{ieff}{ieff_{s,w}}$	$\bar{n}_{\text{int}}(A_w)$	$\bar{n}_{\text{psu}}(\mathbf{b}_w)$	ρ_C^*	$\frac{ieff}{ieff_{s,w}}$
a)	25	1.022	0.299	25	19.202	47.528	-0.005	1.133	19.202	38.389	-0.006	1.123	19.202	49.039	-0.005	1.135
	25	1.036	0.375													
	25	0.998	0.276													
	25	0.945	0.260													
b)	10	1	1	25	15	41	-0.010	1.151	15	29	-0.015	1.138	15	26	-0.017	1.134
	20	1	1													
	30	1	1													
	40	1	1													
c)	10	1	1	25	7.5	20.5	-0.037	1.185	7.5	14.5	-0.054	1.180	7.5	13	-0.061	1.178
	20	1	2													
	30	1	3													
	40	1	4													
d)	10	1	4	25	10	27.333	-0.024	1.171	10	19.333	-0.034	1.163	10	17.333	-0.038	1.161
	20	1	3													
	30	1	2													
	40	1	1													
e)	10	4	144	25	1.801	2.755	-0.551	1.230	1.801	3.603	-0.371	1.231	1.801	4.344	-0.289	1.231
	20	2	9													
	30	0.333	0.555													
	40	0.250	0.125													
f)	10	0.333	0.025	25	31.820	75.685	0.004	1.104	31.820	58.427	0.005	1.084	31.820	40.629	0.007	1.058
	20	0.666	0.075													
	30	1	0.125													
	40	1.333	0.175													
g)	10	1	0.010	25	29.126	79.612	0.002	1.118	29.126	56.311	0.003	1.094	29.126	50.485	0.003	1.086
	20	1	0.020													
	30	1	0.030													
	40	1	0.040													
h)	10	1	0.004	25	29.940	81.836	0.003	1.117	29.940	57.884	0.004	1.092	29.940	51.896	0.004	1.084
	20	1	0.003													
	30	1	0.002													
	40	1	0.001													

If a household and a person within the household are selected at random, then the weights are often independent of the psu and the interviewer and depend only on the household sizes. In such a situation, the household sizes form the weighting classes. For weighting classes, we define

m_{pij}: number of sampling units in psu p assigned to interviewer i belonging to weighting class j,

$m_{pj} = \sum_{i=1}^{I_p} m_{pij}$: number of sampling units in psu p belonging to weighting class j,

$m_j = \sum_{p=1}^{P} \sum_{i=1}^{I_p} m_{pij}$: number of sampling units belonging to weighting class j.

Thus,

$n_{pi} = \sum_{j=1}^{J} m_{pij}$: number of sampling units in psu p assigned to interviewer i,

$n_p = \sum_{i=1}^{I_p} \sum_{j=1}^{J} m_{pij}$: number of sampling units in psu p,

$n = \sum_{p=1}^{P} \sum_{i=1}^{I_p} \sum_{j=1}^{J} m_{pij}$: sample size.

Furthermore,

$$\bar{n}_{\text{int}}(\mathbf{A}_w) = \frac{\sum_{p=1}^{P} \sum_{i=1}^{I_P} \left(\sum_{k=1}^{n_{pi}} w_{pik} \right)^2}{\sum_{p=1}^{P} \sum_{i=1}^{I_p} \sum_{k=1}^{n_{pi}} w_{pik}^2} = \frac{\sum_{p=1}^{P} \sum_{i=1}^{I_P} \left(\sum_{j=1}^{J} w_j m_{pij} \right)^2}{\sum_{j=1}^{J} w_j^2 m_j}$$

and

$$\bar{n}_{\text{psu}}(\mathbf{b}_w) = \frac{\sum_{p=1}^{P} \left(\sum_{i}^{I_p} \sum_{k}^{n_{pi}} w_{pik} \right)^2}{\sum_{p=1}^{P} \sum_{i=1}^{I_p} \sum_{k=1}^{n_{pi}} w_{pik}^2} = \frac{\sum_{p=1}^{P} \left(\sum_{j} w_j m_{pj} \right)^2}{\sum_{j=1}^{J} w_j^2 m_j},$$

are ratios of quadratic forms in $\mathbf{w} = (w_1, \ldots, w_J)$.

5. Overall effects

The overall effects take into account unequal weighting, spatial clustering, and the interview effects and can be viewed as a generalization to the traditional design effects. Multiplying the SRS variance for the unweighted sample mean by the overall effects will provide the total variance estimator.

$$eff = \frac{\mathrm{Var}_{M_4}(\bar{y}_w)}{\mathrm{Var}_{M_1^*}(\bar{y})} = eff_w \times eff_s \times eff_{\mathrm{int}},$$

where

$$eff_w = \frac{\mathrm{Var}_{M_1^*}(\bar{y}_w)}{\mathrm{Var}_{M_1^*}(\bar{y})},$$

$$eff_s = \frac{\mathrm{Var}_{M_3}(\bar{y}_w)}{\mathrm{Var}_{M_1^*}(\bar{y}_w)},$$

$$eff_{\mathrm{int}} = ieff_{s,w} = \frac{\mathrm{Var}_{M_4}(\bar{y}_w)}{\mathrm{Var}_{M_3}(\bar{y}_w)}.$$

In the above, $\mathrm{Var}_{M_1^*}$ is with respect to the following model:

$$M_1^*:\ \mathrm{Cov}(y_{pik}, y_{p'i'k'}) = \begin{cases} \sigma^2 & \text{if } p = p', i = i', k = k', \\ 0 & \text{otherwise.} \end{cases}$$

It can be shown that

$$eff = \frac{n\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2}{\left(\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2} \times \left[1+\rho_C\left(\frac{\sum_{p=1}^{P}\left(\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2}-1\right)+\rho_{\mathrm{int}}\left(\frac{\sum_{p=1}^{P}\sum_{i}^{I_P}\left(\sum_{k}^{n_{pi}} w_{pik}\right)^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2}-1\right)\right].$$

The relative contributions of weighting, spatial clustering, and interviewer effects to the overall effects are given by

$$\mathrm{Re}\,eff_w = \frac{\dfrac{n\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2}{\left(\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2}}{eff},$$

$$\mathrm{Re}\,eff_s = \frac{1+\rho_C(\bar{n}_{\mathrm{psu}}(\mathbf{b}_w)-1)}{eff},$$

$$\mathrm{Re}\,eff_I = \frac{1+\rho_{\mathrm{int}}\dfrac{\bar{n}_{\mathrm{psu}}(A_w)-1}{1+\rho_C(\bar{n}_{\mathrm{psu}}(\mathbf{b}_w)-1)}}{eff}.$$

In Figure 2, we present three dimensional graphs of the relative contributions of weighting, spatial clustering, and interviewer effects to the overall effects for different combinations of intra-cluster and intra-interviewer correlations for different patterns of weights given in cases a), f) and h) of Table 3 with $IA = (1, 3)$, where $IA = (a, b)$ indicates that the first a of the four interviewers are in psu 1 and the last b interviewers are in psu 2.

Remark 5.1: From Result 6, we get

$$eff \geq 1 + \rho_C\left(\frac{n}{P}-1\right) + \rho_{\mathrm{int}}\left(\frac{n}{I}-1\right).$$

The right side is the overall effect if the same number of interviewers with equal workload is assigned to each psu. It is interesting to note the similarity between the right hand side of the above inequality and the design effects formula given in (3.1) of Hansen, Hurwitz and Madow (1953, Vol. I, page 370). To claim the similarity, we need to treat the secondary sampling units as the units belonging to an interviewer. In this connection, we also note the formula (3.7) given in Hansen *et al*. (1953, Vol. II, page 292) for the case $I = P$.

Remark 5.2: When we have the same weighting classes across psu × interviewer, we have

$$eff = \frac{n\sum_{j=1}^{J} w_j^2 m_j}{\left(\sum_{j=1}^{J} w_j m_j\right)^2} \times \left[1+\rho_C\left(\frac{\sum_{p=1}^{P}\left(\sum_{j=1}^{J} w_j m_{pj}\right)^2}{\sum_{j=1}^{J} w_j^2 m_j}-1\right)+\rho_{\mathrm{int}}\left(\frac{\sum_{p=1}^{P}\sum_{i=1}^{I_P}\left(\sum_{j=1}^{J} w_j m_{pij}\right)^2}{\sum_{j=1}^{J} w_j^2 m_j}-1\right)\right].$$

Remark 5.3: Consider the special case

$$m_{pij} = \frac{n_{pi} m_j}{n}$$

in which we allow variation in weights within psu × interviewer classes, but we constrain the weights to have the same relative frequency distribution in each class, *i.e.*, the means and the variances of the weights within the classes do not depend on the class (Lynn and Gabler 2004). It is easy to see that in this case

$$eff = \frac{n\sum_{j=1}^{J} w_j^2 m_j}{\left(\sum_{j=1}^{J} w_j m_j\right)^2} \times \left[1+\rho_C\left(\frac{\left(\sum_{j=1}^{J} w_j m_j\right)^2}{\sum_{j=1}^{J} w_j^2 m_j}\frac{\sum_{p=1}^{P} n_p^2}{n^2}-1\right) + \rho_{\mathrm{int}}\left(\frac{\left(\sum_{j=1}^{J} w_j m_j\right)^2}{\sum_{j=1}^{J} w_j^2 m_j}\sum_{p=1}^{P}\sum_{i=1}^{I_P}\frac{n_{pi}^2}{n^2}-1\right)\right].$$

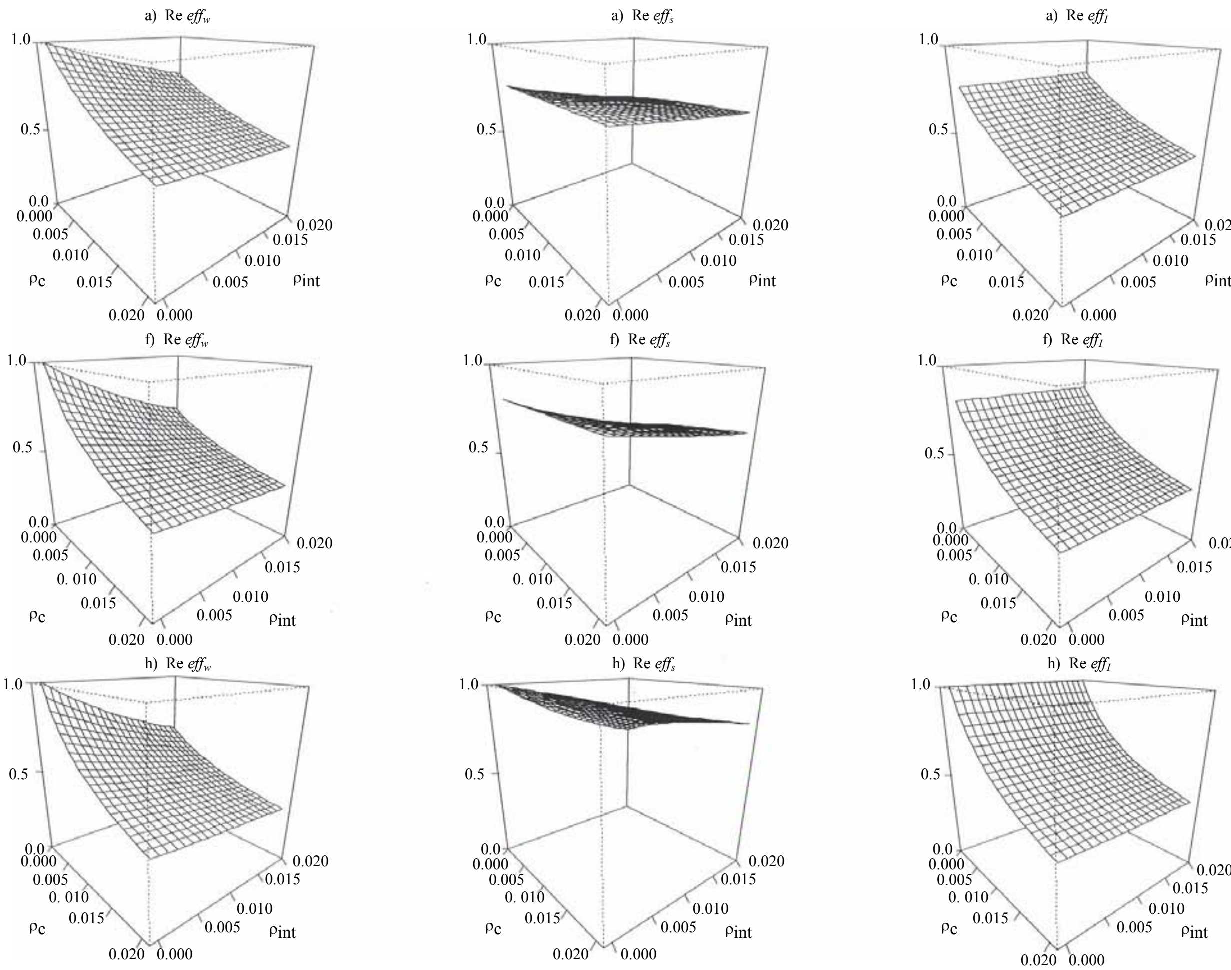


Figure 2 Relative contributions of weighting, design and interviewer effects to the overall effects for cases a), f) and h) in Example 2 for the case *IA* = (1, 3)

Using the same argument given in the proof of Result 6, we get

$$eff \geq 1 + \rho_C \left(\frac{\sum_{p=1}^{P} n_p^2}{n^2} - 1 \right) + \rho_{\text{int}} \left(\sum_{p=1}^{P} \sum_{i=1}^{I_P} \frac{n_{pi}^2}{n} - 1 \right)$$

$$= 1 + \rho_C (\overline{n}_{\text{psu}}(\mathbf{b}) - 1) + \rho_{\text{int}} (\overline{n}_{\text{int}}(\mathbf{A}) - 1).$$

This means that the overall effect is larger than the overall effect for an epsem design (see Remark 5.4).

Remark 5.4: For an epsem design, we have

$$eff = 1 + \rho_C (\overline{n}_{\text{psu}}(\mathbf{b}) - 1) + \rho_{\text{int}} (\overline{n}_{\text{int}}(\mathbf{A}) - 1),$$

where

$$\overline{n}_{\text{psu}}(\mathbf{b}) = \frac{\sum_{p=1}^{P} n_p^2}{n} \text{ and } \overline{n}_{\text{int}}(\mathbf{A}) = \frac{\sum_{p=1}^{P} \sum_{i}^{I_P} n_{pi}^2}{n}.$$

Note that Davis and Scott (1995) obtained this formula for the special case of the following linear mixed model:

$$y_{pik} = \mu + \alpha_i + \beta_p + \varepsilon_{pik},$$

where μ is the overall effect, α_i, β_p are random effects due to the interviewer i, psu p and ε_{pik} is the pure error. They assumed that the random effects are independent with $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_p \sim N(0, \sigma_\beta^2)$ and $\varepsilon_{pik} \sim N(0, \sigma_\varepsilon^2)$.

For the above linear mixed model, it is easy to check that

$$\rho_{\text{int}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\varepsilon^2} \text{ and } \rho_c = \frac{\sigma_\beta^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\varepsilon^2}.$$

However, it is instructive to note that the definition *eff* does not require ρ_{int} and ρ_c to be strictly positive and the definition goes beyond the linear mixed model. For example, the definition applies to the following example:

Example 3: A simple model for binary data.

Assuming $0 < \min(\alpha, \beta) < \theta < 1$, we define the following model:

For all n_{pi} different respondents of interviewer i in psu p.

$P(Y_{pik} = x_1, Y_{pik'} = x_2)$

x_1 \ x_2	1	0	Total
1	α	$\theta - \alpha$	θ
0	$\theta - \alpha$	$1 - 2\theta + \alpha$	$1 - \theta$
Total	θ	$1 - \theta$	1

For all n_{pi} respondents of interviewer i and psu p and all $n_{pi'}$ respondents of interviewer i' and psu p.

$P(Y_{pik} = x_1, Y_{pi'k'} = x_2)$

x_1 \ x_2	1	0	Total
1	β	$\theta - \alpha$	θ
0	$\theta - \alpha$	$1 - 2\theta + \beta$	$1 - \theta$
Total	θ	$1 - \theta$	1

For all n_p respondents of psu p and all $n_{p'}$ respondents of psu p'.

$P(Y_{pik} = x_1, Y_{p'i'k'} = x_2)$

x_1 \ x_2	1	0	Total
1	θ^2	$\theta(1 - \theta)$	θ
0	$\theta(1 - \theta)$	$(1 - \theta)^2$	$1 - \theta$
Total	θ	$1 - \theta$	1

Therefore, we have

$$E(Y_{pik}) = \theta \text{ for all } p, i, k,$$

$$\mathrm{Var}(Y_{pik}) = \theta(1 - \theta) \text{ for all } p, i, k,$$

$$\rho = \frac{\mathrm{Cov}(Y_{pik}, Y_{pik'})}{\sqrt{\mathrm{Var}(Y_{pik})\mathrm{Var}(Y_{pik'})}} = \frac{\alpha - \theta^2}{\theta(1 - \theta)} \text{ for all } p, i \text{ and } k \neq k',$$

$$\rho_C = \frac{\mathrm{Cov}(Y_{pik}, Y_{pi'k'})}{\sqrt{\mathrm{Var}(Y_{pik})\mathrm{Var}(Y_{pi'k'})}} = \frac{\beta - \theta^2}{\theta(1 - \theta)} \text{ for all } p \text{ and } i \neq i',$$

which is a special case of Model M_4 with $\sigma^2 = \mathrm{Var}(Y_{pik}) = \theta(1 - \theta)$. Note that both ρ_C and ρ may be negative and $\rho_{\text{int}} = \rho - \rho_C$ is positive if and only if $\alpha > \beta$.

Remark 5.5: For an epsem design with common psu size $b = n/P$, we have

$$\mathit{eff} = 1 + \rho_C(b - 1) + \rho_{\text{int}}(\overline{n}_{\text{int}}(A) - 1).$$

Remark 5.6: In discussing Verma *et al.* (1980), Holt considered the case when there is no interviewer variability and psu is the weighting class, *i.e.*, the case when $\rho_{\text{int}} = 0$ and $w_{pik} = w_p$ for all p, i, k. In this case eff reduces to

$$\mathit{eff} = \frac{n\sum_{p=1}^{P} n_p w_p^2}{\left(\sum_{p=1}^{P} n_p w_p\right)^2} \times \left[1 + \rho_C\left(\frac{\sum_{p=1}^{P} n_p^2 w_p^2}{\sum_{p=1}^{P} n_p w_p^2} - 1\right)\right].$$

Note that the above formula can be obtained from equation (A4) of Holt in discussing Verma *et al.* (1980), after correcting an obvious typo (*i.e.*, deleting n in the denominator), choosing his choice of survey weight and some algebra. Design effect formulae in the absence of the interviewer effects were considered by many authors. See Kish (1965), Verma *et al.* (1980), Skinner (1986), Valliant (1987), Skinner *et al.* (1989), Gabler, Häder and Lahiri (1999), Lynn and Gabler (2004), Kalton, Brick and Lê (2005) and others.

6. Concluding remarks

We have noticed that the standard interviewer effects formula could have either an overestimation or under-estimation problem depending on the situation. For example, it could severely underestimate the interviewer effects in an epsem sampling design with different interviewer workloads. Interestingly, spatial correlation can turn this underestimation to an overestimation. In the former case, the survey designer who uses the standard interviewer effect formula may pay little attention to control the interviewer effect. In the latter case, a high value of the interviewer effect may unnecessarily raise concerns about the quality of data connected with the interviewer. This may trigger allocation of a higher portion of budget than is necessary to reduce the interviewer effect, which may be already much lower than the value obtained by an application of the standard formula. The paper is an attempt to define and interpret interviewer effects that are appropriate in different complex survey situations.

We have considered the case when an interviewer is assigned only in one psu. The case when an interviewer works in different psu's is also important and will be considered in a later paper. The weights used in the proposed formulae only account for sampling weights as they are planned at the design stage, but do not necessarily reflect the actual weights attached to each case once the data are collected. In other words, our interviewer effect formulae do not incorporate the effects due to nonresponse and post-stratification adjustments. The formulae presented in the paper are mainly useful in the planning and design stage when we have some ideas about the intra-interviewer and spatial correlations.

Reliable estimation of ρ_{int} and ρ_c is important. Although there are some papers that deal with the estimation of ρ_{int} and ρ_c, there is certainly a need to advance research in this important area. In comparing the two sources of homogeneity, Hansen *et al.* (1961) found that the interviewer variability was often larger than the sampling variability. In many surveys, such an evaluation, which requires estimation of the intra-interviewer and intra-cluster correlations, is either difficult or even impossible because the interviewer effects are often confounded with the spatial clustering effects. The use of an interpenetrating design, first proposed by Mahalanobis (1946), where respondents are randomly assigned to the interviewers, is a way to get around the problem. In practice, the implementation of such a design in a large scale sample survey is difficult, but some approximated interpenetrated designs can be applied (Hansen *et al.* 1961, Bailar, Bailey and Stevens 1977, Bailey, Moore and Bailar 1978, Collins and Butcher 1982, O'Muircheartaigh and Campanelli 1998). Multi-level models have been used as a partial remedy to the problem (Hox and De Leeuw 1994, Davis and Scott 1995, O'Muircheartaigh and Campanelli 1998, Scott and Davis 2001). We have not considered the problem of the estimation of the intra-interviewer and intra-cluster correlations. This is an important problem and will be considered in a later paper.

In practice, interviewer or design effects are computed for many items using the same formula and a summary measure such as the median interviewer or design effect is taken for the planning and design of the survey. So far as the issues related to handling multiple items are concerned, one may continue to follow one's own protocol; the only change we may suggest is to use our new definitions for interviewer effects or overall effects whenever applicable. The use of our formula may suggest overall effects, which may be much lower than the standard formula. This, in turn, may suggest lower sample size and hence may save survey costs.

Appendix

Result 1. $$ieff_w = \frac{\text{Var}_{M_2}(\bar{y}_w)}{\text{Var}_{M_1}(\bar{y}_w)} = 1 + \rho_{\text{int}} \left(\frac{\sum_i \left(\sum_k w_{ik} \right)^2}{\sum_i \sum_k w_{ik}^2} - 1 \right).$$

Proof: The result follows by noting

$$\text{Var}_{M_1}(\bar{y}_w) = \text{Var}_{M_1} \left[\frac{\sum_i \sum_k w_{ik} y_{ik}}{\sum_i \sum_k w_{ik}} \right] = \frac{\sigma^2 \sum_i \sum_k w_{ik}^2}{\left(\sum_i \sum_k w_{ik} \right)^2},$$

and

$$\text{Var}_{M_2}(\bar{y}_w) = \frac{\sigma^2 \left[\sum_i \sum_k w_{ik}^2 + \rho_{\text{int}} \sum_i \sum_{k \neq k'} w_{ik} w_{ik'} \right]}{\left(\sum_i \sum_k w_{ik} \right)^2},$$

and some algebra.

Corollary: Assume $\rho_{\text{int}} > 0$ and $w_{ik} = 1/n.$ Using Result 1 and the Cauchy-Schwarz inequality, we get

$$ieff(\mathbf{a}_1) = 1 + \rho_{\text{int}} \left(\frac{\sum_i n_i^2}{n} - 1 \right) \geq 1 + \rho_{\text{int}} \left(\frac{n}{I} - 1 \right) = ieff.$$

Result 2. $ieff_w \leq ieff(\mathbf{a}_2),$ where

$$\mathbf{a}_2 = (a_{21}, \ldots, a_{2I}) \text{ with } a_{2i} = \frac{\sum_k w_{ik}^2}{\sum_i \sum_k w_{ik}^2}.$$

Proof: Using the Cauchy-Schwarz inequality, we have

$$\sum_i \left(\sum_k w_{ik} \right)^2 \leq \sum_i n_i \sum_k w_{ik}^2$$

with equality if and only if $w_{ik} = \bar{w}_i$ for all i and k, where

$$\bar{w}_i = \frac{\sum_{k=1}^{n_i} w_{ik}}{n_i}$$

is the average survey weight for the i^{th} interviewer. Thus, we have $ieff_w \leq 1 + [\bar{n}_{\text{int}}(\mathbf{a}_2) - 1]\rho_{\text{int}} = ieff(\mathbf{a}_2).$

The equality holds if and only if $w_{ik} = \bar{w}_i$ for all i and k in which case $ieff_w = ieff(\mathbf{a}_2^*),$ where

$$\mathbf{a}_2^* = (a_{21}^*, \ldots, a_{2I}^*), \text{ with } a_{2i}^* = \frac{n_i \bar{w}_i^2}{\sum_i n_i \bar{w}_i^2}.$$

If all weights are non-negative, then

$$\sigma_i^2 = \frac{1}{n_i}\sum_k (w_{ik} - \overline{w}_i)^2 \le (n_i - 1)\,\overline{w}_i^2,$$

since σ_i^2 is Schur-convex. Defining

$$x_i = \frac{1 + \dfrac{\sigma_i^2}{\overline{w}_i^2}}{n_i} \text{ implies } \frac{1}{n_i} \le x_i \le 1$$

and

$$\overline{n}_{\text{int}}(\mathbf{a}_2) = \frac{\sum_i n_i \sum_k w_{ik}^2}{\sum_i\sum_k w_{ik}^2} = \frac{\sum_i n_i^2\overline{w}_i^2 + \sum_i n_i^2\sigma_i^2}{\sum_i n_i\,\overline{w}_i^2 + \sum_i n_i\,\sigma_i^2}$$

$$= \frac{\sum_i n_i^3\overline{w}_i^2 - \sum_i n_i^2((n_i - 1)\,\overline{w}_i^2 - \sigma_i^2)}{\sum_i n_i^2\overline{w}_i^2 - \sum_i n_i\,((n_i - 1)\,\overline{w}_i^2 - \sigma_i^2)}$$

$$= \frac{\sum_i n_i^3\overline{w}_i^2 x_i}{\sum_i n_i^2\overline{w}_i^2 x_i} \le \frac{\sum_i n_i^3\overline{w}_i^2}{\sum_i n_i^2\overline{w}_i^2} = \sum_i n_i \frac{n_i^2\overline{w}_i^2}{\sum_i n_i^2\overline{w}_i^2}$$

with equality if and only if $\sigma_i^2 = (n_i - 1)\,\overline{w}_i^2$ for all i or if all n_i are equal.

The inequality follows from the logarithmic concavity of $\overline{n}_{\text{int}}(\mathbf{a}_2)$ as function of $(x_1, ..., x_I)$.

Result 3. For $\mathbf{a}_2^* = (a_{21}^*, ..., a_{2I}^*)$ with $a_{2i}^* = \dfrac{n_i\overline{w}_i^2}{\sum_i n_i\overline{w}_i^2}$

and

$$\mathbf{a}_2 = (a_{21}, ..., a_{2I}) \text{ with } a_{2i} = \frac{\sum_k w_{ik}^2}{\sum_i\sum_k w_{ik}^2},$$

we have

$$ieff(\mathbf{a}_2^*) \underset{\ge}{\overset{\le}{}} ieff(\mathbf{a}_2) \quad \text{if and only if}$$

$$\sum_i n_i\,\sigma_i^2 \sum_i n_i^2\overline{w}_i^2 \underset{\ge}{\overset{\le}{}} \sum_i n_i^2\sigma_i^2 \sum_i n_i\,\overline{w}_i^2.$$

Proof. We have

$$ieff(\mathbf{a}_2^*) - ieff(\mathbf{a}_2) = \frac{\sum_i n_i\,\sigma_i^2\sum_i n_i^2\overline{w}_i^2 - \sum_i n_i^2\sigma_i^2\sum_i n_i\,\overline{w}_i^2}{\left(\sum_i n_i\,\sigma_i^2 + \sum_i n_i\,\overline{w}_i^2\right)\sum_i n_i\,\overline{w}_i^2}.$$

For $n_i = n/I$ for all i, we get

$$ieff(\mathbf{a}_2^*) = ieff(\mathbf{a}_2).$$

For $w_{ik} = \overline{w}_i$ for all i, *i.e.*, $\sigma_i^2 = 0$, we get

$$ieff(\mathbf{a}_2^*) = ieff(\mathbf{a}_2).$$

For $\overline{w}_i = \overline{w}$ for all i and $\sigma_i^2 = \sigma^2 > 0$ for all i, we get

$$ieff(\mathbf{a}_2^*) = ieff(\mathbf{a}_2).$$

For $\overline{w}_i = \overline{w}$ for all i, we get

$$ieff(\mathbf{a}_2^*) \underset{\ge}{\overset{\le}{}} ieff(\mathbf{a}_2) \text{ iff } \sum_i n_i\,\sigma_i^2\sum_i n_i^2 \underset{\ge}{\overset{\le}{}} n\sum_i n_i^2\sigma_i^2.$$

For $\sigma_i^2 = \sigma^2 > 0$ for all i, we get

$$ieff(\mathbf{a}_2^*) \underset{\ge}{\overset{\le}{}} ieff(\mathbf{a}_2) \text{ iff } n\sum_i n_i^2\overline{w}_i^2 \underset{\ge}{\overset{\le}{}} \sum_i n_i\,\overline{w}_i^2\sum_i n_i^2.$$

Result 4. We have

$$ieff_w - ieff = \frac{\overline{n}_{\text{int}}}{SST + n\overline{w}^2}\left[\sum_{i=1}^{I}\left(\frac{n_i}{\overline{n}_{\text{int}}} - 1\right)n_i\overline{w}_i^2 - SSW\right]\rho_{\text{int}}$$

$$= \frac{\overline{n}_{\text{int}}}{(1 + CV_w^{-2})SST}\left[\sum_{i=1}^{I}\left(\frac{n_i}{\overline{n}_{\text{int}}} - 1\right)n_i\overline{w}_i^2 - SSW\right]\rho_{\text{int}}$$

$$= \frac{\overline{n}_{\text{int}}\tau_w}{1 + CV_w^{-2}}\left(\frac{\sum_{i=1}^{I}\left(\frac{n_i}{\overline{n}_{\text{int}}} - 1\right)n_i\overline{w}_i^2}{SSW} - 1\right)\rho_{\text{int}}.$$

Proof.

$$ieff_w - ieff$$

$$= 1 + \left(\frac{\sum_i\left(\sum_k w_{ik}\right)^2}{\sum_i\sum_k w_{ik}^2} - 1\right)\rho_{\text{int}} - 1 - (\overline{n}_{\text{int}} - 1)\,\rho_{\text{int}}$$

$$= \left(\frac{\sum_i\left(\sum_k w_{ik}\right)^2}{\sum_i\sum_k w_{ik}^2} - \overline{n}_{\text{int}}\right)\rho_{\text{int}}$$

$$= \left(\frac{\sum_i n_i^2\overline{w}_i^2}{SST + n\overline{w}^2} - \overline{n}_{\text{int}}\right)\rho_{\text{int}}$$

$$= \frac{\overline{n}_{\text{int}}}{SST + n\overline{w}^2}\left(\frac{\sum_i n_i^2\overline{w}_i^2}{\overline{n}_{\text{int}}} - (SST + n\overline{w}^2)\right)\rho_{\text{int}}$$

$$= \frac{\overline{n}_{\text{int}}}{SST + n\overline{w}^2}\left(\sum_{i=1}^{I}\left(\frac{n_i}{\overline{n}_{\text{int}}} - 1\right)n_i\overline{w}_i^2 + \sum_{i=1}^{I} n_i\overline{w}_i^2 - (SST + n\overline{w}^2)\right)\rho_{\text{int}}.$$

Now the result follows using algebra.

Result 5.

$$ieff_{s,w} = \frac{\mathrm{Var}_{M_4}\left(\bar{y}_w\right)}{\mathrm{Var}_{M_3}\left(\bar{y}_w\right)}$$

$$= 1 + \rho_{\text{int}} \frac{\dfrac{\sum_{p=1}^{P}\sum_{i}^{I_P}\left(\sum_{k}^{n_{pi}} w_{pik}\right)^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2} - 1}{1 + \rho_C\left(\dfrac{\sum_{p=1}^{P}\left(\sum_{i}^{I_P}\sum_{k}^{n_{pi}} w_{pik}\right)^2}{\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2} - 1\right)}.$$

Proof. The result follows by noting that

$$\frac{\mathrm{Var}_{M_4}\left(\bar{y}_w\right)}{\mathrm{Var}_{M_3}\left(\bar{y}_w\right)} =$$

$$\frac{\sum_{p=1}^{P}\sum_{i=1}^{I_P}\sum_{k=1}^{n_{pi}} w_{pik}^2 + \rho_C \sum_{p=1}^{P}\sum_{i\neq i'}\sum_{k}^{n_{pi}}\sum_{k'}^{n_{pi'}} w_{pik}w_{pi'k'} + \rho\sum_{p=1}^{P}\sum_{i}^{I_p}\sum_{k\neq k'}^{n_{pi}} w_{pik}w_{pik'}}{\sum_{p=1}^{P}\left(\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2 + \rho_C\sum_{i,i'}^{I_p}\sum_{k\neq k'} w_{pik}w_{pi'k'}\right)}$$

and some algebra.

Result 6. For $0 < \rho_C < 1$ and $0 < \rho_{\text{int}} < 1$,

$$eff \geq 1 + \rho_C\left(\frac{n}{P} - 1\right) + \rho_{\text{int}}\left(\frac{n}{I} - 1\right),$$

with equality if and only if the weights are all equal and each interviewer has the same workload.

If we have in each psu only one interviewer, then

$$eff \geq 1 + (\rho_C + \rho_{\text{int}})\left(\frac{n}{P} - 1\right).$$

Proof: Using some algebra and the general inequality,

$$\sum_{j}^{J} p_j x_j^2 \geq \left(\sum_{j}^{J} p_j x_j\right)^2$$

with

$$p_j \geq 0 \text{ and } \sum_{j=1}^{J} p_j = 1,$$

we have

$$eff = \frac{n\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}^2}{\left(\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2}(1 - \rho_C - \rho_{\text{int}})$$

$$+ n\rho_C \frac{\sum_{p=1}^{P}\left(\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2}{\left(\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2} + n\rho_{\text{int}} \frac{\sum_{p=1}^{P}\sum_{i}^{I_P}\left(\sum_{k}^{n_{pi}} w_{pik}\right)^2}{\left(\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2}$$

$$\geq 1 - \rho_C - \rho_{\text{int}} + \rho_C\frac{n}{P} + \rho_{\text{int}} n \frac{I\sum_{p=1}^{P}\dfrac{I_p}{I}\left(\dfrac{1}{I_p}\sum_{i}^{I_p}\sum_{k}^{n_{pi}} w_{pik}\right)^2}{\left(\sum_{p=1}^{P}\sum_{i=1}^{I_p}\sum_{k=1}^{n_{pi}} w_{pik}\right)^2}$$

$$\geq 1 - \rho_C - \rho_{\text{int}} + \rho_C\frac{n}{P} + \rho_{\text{int}}\frac{n}{I}$$

$$= 1 + \rho_C\left(\frac{n}{P} - 1\right) + \rho_{\text{int}}\left(\frac{n}{I} - 1\right).$$

Acknowledgements

We thank the editor and referees for constructive comments and suggestions which have improved the original version of this paper.

References

Bailar, B.A., Bailey, L. and Stevens, J. (1977). Measures of interviewer bias and variance. *Journal of Marketing Research*, 14, 337-343.

Bailey, L., Moore, T.F. and Bailar, B.A. (1978). An interviewer variance study for eight impact cities of the National Crime Survey Cities Sample. *Journal of the American Statistical Association*, 73, 16-23.

Biemer, P.P., and Stokes, S.L. (1985). Optimal design of interviewer variance experiments in complex surveys. *Journal of the American Statistical Association*, 80, 369, 158-166.

Biemer, P., and Trewin, D. (1997). A review of measurement error effects on the analysis of survey data. In *Survey Measurement and Process Quality*, (Eds., L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin), New York: John Wiley & Sons, Inc., 603-632.

Brewer, K.R.W. (1963). Ratio estimation and finite populations: Some results deductible from the assumption of an underlying stochastic process, *Australian Journal of Statistics*, 5, 93-105.

Chambers, R.L., and Skinner, C.J. (Eds.) (2003). *Analysis of Survey Data*. Wiley, Chichester.

Collins, M., and Butcher, B. (1982). Interviewer and clustering effects in an attitude survey. *Journal of the Market Research Society*, 25, 39-58.

Davis, P.D., and Scott, A.J. (1995). The effect of interviewer variance on domain comparisons. *Survey Methodology*, 21, 99-106.

Feather, J. (1973). A study of interviewer variance. WHO International Collaborative Study of Medical Care Utilization, Saskatchewan Study Area Reports, Series II, Monograph No. 3.

Fellegi, I.P. (1964). Response variance and its estimation. *Journal of the American Statistical Association*, 59, 1016-1041.

Gabler, S., Häder, 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.

Gray, P.G. (1956). Examples of interviewer variability taken from two sample surveys. *Applied Statistics*, V, 73-85.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.

Groves, R.M., and Magilavy, L.J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, 50, 251-266.

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

Hansen, M.H., Hurwitz, W.N. and Bershad, M.A. (1961). Measurement errors in census and surveys. *Bulletin of the ISI 38*, 2, 351-374.

Hanson, R.H., and Marks, E.S. (1958). Influence of the interviewer on the accuracy of survey results. *Journal of the American Statistical Association*, 53, 635-655.

Hájek, J. (1971). Comments, In Foundations of Statistical Inference, (Eds. V.P. Godambe and D.A. Sprott). Toronto: Holt, Rinchart, and Winston.

Heeb, J.-L., and Gmel, G. (2001). Interviewers and respondents effects on self-reported alcohol consumption in Swiss Health Survey. *Journal of Studies on Alcohol*, 62, 434-442.

Hox, J.J., and De Leeuw, E.D. (1994). A comparison of nonresponse in mail, telephone, and face-to-face surveys. Applying multilevel modeling to meta-analysis. *Quality & Quantity*, 329-344.

Kalton, G., Brick, J.M. and Lê, T.h. (2005). Estimating components of design effects for use in sample design. In: *Household Sample Surveys in Developing and Transition Countries*, Chapter VI. Available from http://unstats.un.org/unsd/hhsurveys/pdf/Chapter_6.pdf.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, 57, 92-115.

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

Lynn, P., and Gabler, S. (2004). Approximations to b^* in the estimation of design effects due to clustering. *Working Papers of the Institute for Social and Economic Research*, paper 2004-07. Colchester: University of Essex. Available from http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-07.pdf.

Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, Serie A, 109, 325-378, reprinted in *Sankhyā* (1958), 1-68.

O'Muircheartaigh, C., and Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society*, Series A, 161, 63-77.

Philippens, M., and Loosveldt, G. (2004). Interviewer-related variance in the European Social Survey. Paper presented at the sixth international conference on social science methodology, August 17-20 in Amsterdam.

Rice, S.A. (1929). Contagious bias in the interview: A methodological note. *American Journal of Sociology*, 35, 420-423.

Rao, J.N.K., and Scott, A.J. (1984). On chi-squared tests for multi-way contingency tables with proportions estimated from survey data. *Annals of Statistics*, 12, 46-60.

Schnell, R., and Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics*, 21, 389-410.

Scott, A.J., and Davis, P.D. (2001). Estimating interviewer effects for binary responses. *Proceedings*: *Symposium 2001*, Achieving Data Quality in a Statistical Agency.

Skinner, C.J. (1986). Design effect of two-stage sampling, *Journal of the Royal Statistical Society*, Serie B, 48, 89-99.

Skinner, C.J., Holt, D. and Smith, T.M.F. (eds.) (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons, Inc.

Tucker, C. (1983). Interviewer effects in telephone surveys. *Public Opinion Quarterly*, 47, 84-95.

Valliant, R.M. (1987). Generalized variance functions in stratified two-stage sampling, 82, 499-508.

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*, Serie A, 143, 431-473.