Molefe, Wilford

Article

Optimal allocation for equal probability two-stage design

Statistics in Transition new series (SiTns)

Provided in Cooperation with:
Polish Statistical Association

Suggested Citation: Molefe, Wilford (2022) : Optimal allocation for equal probability two-stage design, Statistics in Transition new series (SiTns), ISSN 2450-0291, Sciendo, Warsaw, Vol. 23, Iss. 4, pp. 129-148,
https://doi.org/10.2478/stattrans-2022-0046

This Version is available at:
https://hdl.handle.net/10419/301893

STATISTICS IN TRANSITION new series, December 2022
Vol. 23, No. 4, pp. 129–148, DOI 10.2478/stattrans-2022-0046
Received – 14.04.2021; accepted – 09.06.2022

Optimal allocation for equal probability two-stage design

Wilford Molefe[1]

ABSTRACT

This paper develops optimal designs when it is not feasible for every cluster to be represented in a sample as in stratified design, by assuming equal probability two-stage sampling where clusters are small areas. The paper develops allocation methods for two-stage sample surveys where small-area estimates are a priority. We seek efficient allocations where the aim is to minimize the linear combination of the mean squared errors of composite small area estimators and of an estimator of the overall mean. We suggest some alternative allocations with a view to minimizing the same objective. Several alternatives, including the area-only stratified design, are found to perform nearly as well as the optimal allocation but with better practical properties. Designs are evaluated numerically using Switzerland canton data as well as Botswana administrative districts data.

Key words: sample designs, optimal allocation, composite estimation, mean squared error, two-stage sampling, simple random sampling without replacement

1. Introduction

In many situations it is not feasible for every small area to be represented in a sample. In practice, it is not possible to anticipate and plan for all possible areas (or domains) and uses of survey data as "the client will always require more than is specified at the design stage" (Fuller, 1999).

Longford (2006), Molefe (2011), Molefe and Clark (2015) and Molefe, Shangodoyin and Clark (2015) derive optimal allocations for stratified sampling, which minimize weighted sums of the MSEs of small area estimates and a grand mean estimate. In Longford (2006), the MSEs are design-based (that is, based on repeated probability sampling from a fixed population without reference to a model), and in Molefe (2011), Molefe and Clark (2015) and Molefe, Shangodoyin and Clark (2015) anticipated MSEs are used. In all the references above, stratified simple random sampling without replacement is assumed, where strata are small areas. All find that the optimal design could sometimes have zero sample size for the smallest areas. The authors establish numerically that simpler designs with positive stratum sample sizes give near optimal anticipated MSEs. Power allocation (Bankier, 1988) with stratum sample sizes proportional to a numerically optimized exponent of the stratum population performs particularly well.

In this paper we consider the case of equal probability two-stage sampling design where small areas are clusters or primary sampling units (PSUs). Two-stage sampling with equal probabilities of selection for all clusters (at least within broad regions) are used in many large scale sample surveys including the Australian and New Zealand labour force surveys.

[1]University of Botswana, Botswana. E-mail: molefewb@ub.ac.bw. ORCID: https://orcid.org/0000-0001-7674-2244.

It will be assumed that a sample of small areas is selected by SRSWOR, followed by a sample of the second stage units (units) from each selected small area, also by SRSWOR. There are several possible reasons for this approach. There may be a list of the small areas in the population, but not of the population units. Two stage surveys are also useful so that the sample can be made more geographically clustered, which often reduces enumeration costs (Cochran, Chapter 10 1977).

In optimizing this sampling design for small area estimation where small areas are clusters, the fundamental question is how to choose the number of clusters (m) and the number of subunits, referred to as just units (n_d) per cluster subject to fixed cost. One approach is to choose m and n_d to optimize some criteria subject to a cost constraint based on some model for cost.

We adopt the criterion of the weighted sum of the MSE for the small areas in-sample and the MSE of the estimator for the small areas out-of-sample.

A question within the above setup is when it is appropriate to have some sample in every small area. This would only be feasible when there are a relatively small number of small areas (M), or a very large survey budget, and would usually mean that the number of units $\{n_d\}$ in each small area would be fairly small. In this case the design will be a special case of stratified design considered by Longford (2006), Molefe (2011), Molefe and Clark (2015) and Molefe, Shangodoyin and Clark (2015).

In practice, it is not always feasible for every small area to be represented in the sample. This is clear from the fact that zero stratum sample sizes sometimes arise in Longford (2006), Molefe (2011) and Molefe and Clark (2015). In this paper, we explicitly allow for the sampling of small areas. It is assumed that a two-stage design is used, where clusters are small areas. A cluster d may be selected with equal probability $\pi_d = \frac{m}{M}$ and a different sample size n_d to be selected from each selected cluster. In Section 2 we state a two-level model and the resulting anticipated MSE of small area estimates. An objective function which is a linear combination of anticipated MSEs is defined. A linear cost model consisting of per-cluster and per-unit costs is assumed. The aim is then to minimize the objective function with respect to m and n_d subject to fixed expected cost for the survey. In Section 3 we develop an optimal analytical solution when only small area estimates are a priority. Section 4 suggests sensible but ad-hoc designs that include equal allocation, proportional allocation, classical optimal allocation and a combined design made up of the proportional allocation and the classical optimal design. Section 5 is a numerical study based on the Switzerland canton population sizes used by Longford (2006). Section 6 contains conclusions.

2. Methods

From a population of M small areas (clusters) indexed by d, denoted by U^1, a first stage sample of m small areas selected by SRSWOR is denoted s^1. In the second stage of the selection a sample of size n_d elements selected by SRSWOR from area d is denoted by s_d. The set of N_d population units in a particular cluster d is denoted by U_d. Let the sampling variances be $v_d = var_p(\bar{y}_d)$ and $v = var_p(\bar{y})$ respectively for the small area mean estimator and overall mean estimator. The composite estimator is denoted $\tilde{y}_d^{\mathscr{C}}[\phi_{d(opt)}]$.

Let Y_j be the value of the characteristic of interest for the jth unit in the population. The small area population mean is $\bar{Y}_d$ and the national mean is $\bar{Y}$. Auxiliary variables x_j are assumed to be available for the full population $j \in U^1$.

The following two-level linear mixed model ξ will be assumed:

$$\left.\begin{array}{l} Y_j = \beta^T x_j + u_d + \varepsilon_j \\ E_\xi[u_d] = E_\xi[\varepsilon_j] = 0 \\ var_\xi[u_d] = \sigma^2_{ud} \\ var_\xi[\varepsilon_j] = \sigma^2_{\varepsilon d} \end{array}\right\} \tag{1}$$

for $d \in U^1$ and $j \in U_d$ with mutual independence of u_d and ε_j for $d \in U^1$ and $j \in U_d$. This implies $var_\xi[Y_j] = \sigma^2_{ud} + \sigma^2_{\varepsilon d} = \sigma^2$ for all $j \in U$, and that the covariance $cov_\xi[Y_i, Y_j] = \rho_d \sigma^2_d$ for units $i \neq j$ in the same small area and 0 for units from different small areas, where $\rho_d = \sigma^2_{ud}/\sigma^2$. For simplicity it will be assumed that $\rho_d = \rho$.

Following Molefe and Clark (2015), we assume a small-area composite estimator which is a weighted mean of an approximately design unbiased estimator

$$\bar{y}_{dr} \quad = \quad \bar{y}_d + \hat{\beta}^T (\bar{X}_d - \bar{x}_d)$$

recommended by Hidiroglou and Patak (2004) for small domains, and a model-based synthetic estimator $\hat{\bar{Y}}_{d(syn)} = \hat{\beta}^T \bar{X}_d$.

The composite estimator which approximately minimizes the anticipated MSE is

$$\tilde{\bar{y}}^{\mathscr{C}}_d[\phi_{d(opt)}] \quad = \quad (1-\phi_d)\bar{y}_{dr} + \phi_d \hat{\bar{Y}}_{d(syn)} = \hat{\beta}^T \bar{X}_d + (1-\phi_d)\left(\bar{y}_d - \hat{\beta}^T \bar{x}_d\right)$$

where $\phi_{d(opt)} = (1-\rho)\left[1+(n^*_d-1)\rho\right]^{-1}$, assuming that n, N_d and M are all large (Molefe and Clark, 2015). Under the same assumptions, the approximate anticipated MSE of the optimal composite estimator of $\bar{Y}_d$ conditional on n^*_d is

$$\begin{aligned} & E_\xi MSE_p\left(\tilde{\bar{y}}^{\mathscr{C}}_d\left[\phi_{d(opt)}\right]; \bar{Y}_d | n^*_d\right) \\ & \approx \quad \left\{n^*_d \rho\left[1+(n^*_d-1)\rho\right]^{-1}\right\}^2 (n^*_d)^{-1} \sigma^2 (1-\rho) + \left\{(1-\rho)\left[1+(n^*_d-1)\rho\right]^{-1}\right\}^2 \sigma^2 \rho \\ & = \quad \sigma^2 \rho (1-\rho)/\left[1+(n^*_d-1)\rho\right] \end{aligned} \tag{2}$$

See Molefe (2011) for the derivation.

Small areas with no sample would have a direct estimate of zero. For these, a synthetic estimator is used. An indirect estimator, $\tilde{\bar{y}}^{\mathscr{C}}_d = \bar{y}$ is proposed, if cluster $d \notin s^1$. The MSE of $\bar{y}$ is given by $MSE_p\left(\bar{y}; \bar{Y}_d\right) = v + B^2_d$, where B_d is the design bias of using $\bar{y}$ to estimate $\bar{Y}_d$.

The population level mean estimator $\bar{y}$ and area mean $\bar{y}_d$ are assumed to be unbiased for $\bar{Y}$ and $\bar{Y}_d$ respectively. The design variance of the synthetic estimator will be small relative to the design variance of the direct estimator because it depends only on the precision of direct estimators at a larger area level. If the number of small areas in the sample is large, v is negligible and can be ignored. Therefore, we approximate $MSE_p\left(\bar{y}; \bar{Y}_d\right)$ by B^2_d.

For optimal allocation of sample sizes of clusters and subunits, we search for the area-

level sampling design that minimizes the weighted expected value of the sum of the sampling variances (MSEs) for a combination of small area composite estimates for clusters in-sample and out-of-sample and an overall estimator of the mean given by

$$F = \sum_{d\in U^1} \pi_d N_d^q AMSE_d\left\{\tilde{y}_d^{\mathscr{C}}\left[\phi_{d(opt)}\right];\bar{Y}_d\right\} + \sum_{d\in U^1}(1-\pi_d)N_d^q AMSE_d\left[\bar{y};\bar{Y}_d\right] \quad (3)$$

where the first component in (3) is due to the m clusters in-sample and the second component is due to the remaining $(M-m)$ clusters. The small-area population sizes N_d are weights, that is, N_d^q for $0 \leq q \leq 2$, where for $q=0$, inference is equally important for every area. With increasing q, relatively greater importance is ascribed to more populous areas, with $q=2$ corresponding to proportional allocation. $AMSE_d$ is the model assisted mean squared error, that is ξMSE_d.

We can then write the model expectation of the criterion function to be minimized, ignoring the goal of national estimation, as

$$F \approx \frac{m}{M}\sum_{d\in U^1} N_d^q \frac{\sigma^2\rho(1-\rho)}{[1+(n_d-1)\rho]} + \left(1-\frac{m}{M}\right)\sigma^2\rho\sum_{d\in U^1} N_d^q \quad (4)$$

2.1. Cost Models and Cost Estimates

In a two stage sampling scheme the sampling variance of the estimate of the overall population mean ($\bar{y}$) is minimized (for fixed sample size) when $\bar{n}=1$ since this is when the sample is most spread out. However, costs will be minimized when as few first stage units as possible are selected. Hence, some compromise between these two extremes has to be chosen and this is the optimal design problem in multistage sampling. As always costs and variances are pulling in opposite directions and the task of optimal design is to choose the optimal balance of these.

In a two-stage sample, several types of costs can be distinguished (Hansen, 1953; Cochran, 1977):

(a) Overhead costs - costs associated with planning, administration, setting up processing systems, etc. These costs do not depend on the sample sizes used at either stage;

(b) Costs associated with the selection of clusters - these arise from drawing maps, listing units within selected primary stage units, travel between selected primary stage units. These costs increase as the number of clusters selected increases;

(c) Costs associated with the selection of secondary stage units - these mainly arise from the enumeration of selected population units, e.g. the cost of time spent in interviewing people and the cost of processing an individual questionnaire. These costs increase as the number of selected units increases.

Linear cost models are commonly used by official statistics agencies (Hansen, 1953; Sukhatme, 1954; Cochran, 1977; Foreman, 1991; Clark, 2007). A linear cost model is often adequate for sample design, even though it cannot perfectly capture the real cost structure.

A simple cost model for a two-stage sample is given by

$$C_F = c_0 + c_1 m + c_2 \frac{m}{M} \sum_{d \in U^1} n_d$$

where m equals the number of primary sampling units (clusters) in the sample; n_d is the number of secondary sampling units (units), for example, households, in the sample from cluster d; the coefficient c_0 is the fixed costs of conducting the survey, independent of the number of sample clusters and subunits per cluster, including costs for survey planning, development of the survey design, preparatory work, survey management and data processing, analysis and presentation of results; the coefficient c_1 is the average cost of adding a cluster to the sample, consisting of travel by interviewers and supervisors between clusters and home base or between clusters (fuel costs, driver salaries) and interviewer salaries, including the cost of obtaining maps and other material for the cluster, the cost of establishing the survey in the local area, entailing, for example, meeting with and obtaining permission from local authorities, and the cost of listing and sampling of dwelling units within the cluster; the coefficient c_2 is the average cost of including an extra household in the sample, including the costs for locating, contacting and interviewing a household, where the costs consist of interviewer and supervisor salaries and allowance, and also costs for travel by interviewers and supervisors within clusters (Pettersson and Sisouphanthong, 2005).

Costs for the different components of a survey differ from survey to survey and from country to country. The survey manager often has a good idea of the time required for specific survey operations based on information from previous surveys of a similar nature. Experiences from prior surveys (or from pilot surveys) could often be used for reasonable estimates of time per household required for locating and interviewing the household. In these cases, reasonable estimates of c_2 could be compiled.

Computing a reasonable estimate of c_1 is often difficult because it involves determining the effect of additional interviewer travel when a cluster is added to the sample. The travel depends on the size of the area being covered, the number of clusters assigned to each interviewer, and the travel pattern of the interviewers. The travel includes between-cluster and within-cluster travel during a data collection trip.

Cost modelling is mainly used for budgetary purposes and for finding an efficient sample design. In this thesis, our interest is mainly in the use of cost models to find an efficient design. We do not consider the fixed costs (c_0) in trying to work out an efficient design; we only consider the fieldwork costs. The total sampling cost function has two components; the first part depends on how many small areas, $c_1 m$, and the other on the total number of units sampled, namely $c_2 \sum_{d \in s^1} n_d$. The second component will, however, vary from sample to sample of the clusters.

Therefore, the expected total sampling cost function will then be given by

$$C_E = c_1 m + c_2 \frac{m}{M} \sum_{d \in U^1} n_d \tag{5}$$

The aim is to minimise F with respect to n_d and m subject to a cost constraint $C_E = C_F$.

3. Area-Only Simple Two-Stage Optimal Design

The expected criterion function (4), eliminating the common σ^2 and ρ terms, reduces to

$$F \approx -\frac{m}{M}\sum_{d\in U^1}\frac{N_d^q n_d \rho}{1+(n_d-1)\rho} \tag{6}$$

plus constant terms which do not depend on m or n_d.

We minimize (6) subject to the cost constraint (5). The Lagrangian is:

$$L = F+\lambda\left(c_1 m + c_2\frac{m}{M}\sum_{d\in U^1} n_d - C_F\right) \tag{7}$$

To obtain an optimal number of clusters and subunits to take into the sample, we take partial derivatives of (7) with respect to n_d, λ and m.

We use the partial derivatives to derive the optimal design by firstly deriving $\bar{n}_{opt.}$, the optimal average within-cluster sample size. This result will then be used to derive the optimal values of n_d.

We use $\frac{\partial L}{\partial n_d}=0$ to obtain the optimal value for n_d as follows:

$$n_d = N_d^{\frac{q}{2}}\sqrt{(1-\rho)/(\lambda c_2\rho)}-(1-\rho)/\rho \tag{8}$$

This solution for n_d given implies that the average within-cluster sample size is

$$\bar{n} = \bar{N}_d^{\frac{q}{2}}\sqrt{(1-\rho)/(\lambda c_2\rho)}-(1-\rho)/\rho$$

Therefore, we can write

$$\sqrt{(1-\rho)/(\lambda c_2\rho)} = \left(\bar{N}_d^{\frac{q}{2}}\right)^{-1}\{\bar{n}+(1-\rho)/\rho\}$$

Then, the optimal cluster sample sizes can be expressed as

$$n_d = N_d^{\frac{q}{2}}\left(\bar{N}_d^{\frac{q}{2}}\right)^{-1}\bar{n}+(1-\rho)/\rho\left[N_d^{\frac{q}{2}}\left(\bar{N}_d^{\frac{q}{2}}\right)^{-1}-1\right]$$

We can also substitute for n_d given by (8) in $\frac{\partial L}{\partial \lambda}=0$ to obtain

$$c_1 m + c_2\frac{m}{M}\sum_{d\in U^1}\left(N_d^{\frac{q}{2}}\sqrt{(1-\rho)/(\lambda c_2\rho)}-(1-\rho)/\rho\right) = C_F$$

This simplifies to

$$C_F = \gamma m + \sqrt{\frac{c_2}{\lambda}}\frac{m}{M}\sum_{d\in U^1} N_d^{\frac{q}{2}}\sqrt{(1-\rho)/\rho}$$

where $\gamma = c_1 - c_2(1-\rho)/\rho$.

Similarly, we substitute for n_d in $\frac{\partial L}{\partial m} = 0$ and after simplifying we obtain

$$\frac{1}{M}\sum_{d\in U^1} N_d^q = \frac{1}{M}\sum_{d\in U^1} N_d^{\frac{q}{2}}\sqrt{\lambda c_2(1-\rho)/\rho} + \lambda\left(\gamma + \frac{1}{M}\sqrt{\frac{c_2}{\lambda}}\sum_{d\in U^1} N_d^{\frac{q}{2}}\sqrt{(1-\rho)/\rho}\right)$$

Removing the bracket on the right hand size, we obtain

$$\frac{1}{M}\sum_{d\in U^1} N_d^q = 2\frac{1}{M}\sum_{d\in U^1} N_d^{\frac{q}{2}}\sqrt{\lambda c_2(1-\rho)/\rho} + \lambda\gamma$$

The resulting two simultaneous equations in m and λ are:

$$\frac{m}{M}\sqrt{c_2/\lambda(1-\rho)/\rho}\sum_{d\in U^1} N_d^{\frac{q}{2}} + \gamma m = C_F \tag{9}$$

$$2\sqrt{\lambda c_2(1-\rho)/\rho}\sum_{d\in U^1} N_d^{\frac{q}{2}} + \lambda\gamma M = \sum_{d\in U^1} N_d^q \tag{10}$$

We use (9) to write λ in terms of m as follows:

$$\sqrt{\lambda} = \frac{1}{M\left(C_F/m - \gamma\right)}\sum_{d\in U^1} N_d^{\frac{q}{2}}\sqrt{c_2(1-\rho)/\rho}$$

Substituting for λ in (10) we obtain

$$\sum_{d\in U^1} N_d^q = \frac{2c_2(1-\rho)/\rho\left(\sum_{d\in U^1} N_d^{\frac{q}{2}}\right)^2}{M\left(C_F/m-\gamma\right)} + \frac{c_2(1-\rho)/\rho\left(\sum_{d\in U^1} N_d^{\frac{q}{2}}\right)^2}{M\left(C_F/m-\gamma\right)^2}\times\gamma$$

Cross-multiplying and further simplifying we obtain

$$\begin{aligned} 0 = {} & \gamma\left\{c_2(1-\rho)/\rho\left(\sum_{d\in U^1} N_d^{\frac{q}{2}}\right)^2 + \gamma M\sum_{d\in U^1} N_d^q\right\} + M\left(\frac{C_F}{m}\right)^2\sum_{d\in U^1} N_d^q - \\ & 2\frac{C_F}{m}\left\{c_2(1-\rho)/\rho\left(\sum_{d\in U^1} N_d^{\frac{q}{2}}\right)^2 + \gamma M\sum_{d\in U^1} N_d^q\right\} \end{aligned} \tag{11}$$

which is a quadratic in m^{-1} of the form $am^{-2} + bm^{-1} + c = 0$.

Define $C_{q/2}^2$ the relative population variance of $N_d^{\frac{q}{2}}$ given by

$$C_{q/2}^2 = M^{-1}\sum_{d\in U^1}\left(N_d^{\frac{q}{2}} - \bar{N}^{\frac{q}{2}}\right)^2/\left(\bar{N}^{\frac{q}{2}}\right)^2 \tag{12}$$

Then

$$\frac{M^{-1}\sum_{d\in U^1}(N_d^{\frac{q}{2}})^2}{(M^{-1}\sum_{d\in U^1}N_d^{\frac{q}{2}})^2} = \frac{M^{-1}\sum_{d\in U^1}N_d^q}{(M^{-1}\sum_{d\in U^1}N_d^{\frac{q}{2}})^2} = 1+C_{q/2}^2$$

Hence, we write

$$\sum_{d\in U^1}N_d^q = M^{-1}\left(\sum_{d\in U^1}N_d^{\frac{q}{2}}\right)^2\left(1+C_{q/2}^2\right) \tag{13}$$

We substitute for $\sum_{d\in U^1}N_d^q$ into (11) to obtain a reduced quadratic equation in m^{-1}:

$$0 = \left(\frac{C_F}{m}\right)^2(1+C_{q/2}^2)-2\frac{C_F}{m}\left[c_2(1-\rho)/\rho+\gamma\left(1+C_{q/2}^2\right)\right]+\gamma\left[c_2(1-\rho)/\rho+\gamma\left(1+C_{q/2}^2\right)\right] \tag{14}$$

Define $\bar{n}=E\left[\frac{n}{m}\right]=\frac{1}{M}\sum_{d\in U^1}n_d$. There is a one-to-one relationship between m and $\bar{n}$ because $C_F=c_1m+c_2m\bar{n}$ so that $m=C_F/(c_1+c_2\bar{n})$. Hence finding the optimal m is equivalent to finding $\bar{n}$. Substituting for m^{-1} into (14) we obtain

$$0 = (c_1+c_2\bar{n})^2(1+C_{q/2}^2)-2(c_1+c_2\bar{n})\left[c_2(1-\rho)/\rho+\gamma(1+C_{q/2}^2)\right]+\gamma\left[c_2(1-\rho)/\rho+\gamma(1+C_{q/2}^2)\right]$$

which is a quadratic in $\bar{n}$ of the form $a\bar{n}^2+b\bar{n}+c$.

Therefore, the optimum $\bar{n}$ is:

$$\bar{n}_{opt.} = \frac{-c_2(1-\rho)/\rho C_{q/2}^2 \pm \left[c_1c_2(1-\rho)/\rho+\left\{c_2(1-\rho)/\rho\gamma\right\}C_{q/2}^2\right]^{\frac{1}{2}}}{c_2\left(1+C_{q/2}^2\right)} \tag{15}$$

Of primary interest will be to compare the optimal sample size using composite estimation, $\bar{n}_{opt.}$, with the classical two-stage optimal design given by Hansen, Hurwitz and Madow (1953, page 173 equations 10.1 and 10.2) and Cochran (1977, page 281 equation 10.26) as $\bar{n}_{cl.}=\sqrt{c_1/c_2(1-\rho)/\rho}$ for the purpose of drawing general conclusions on whether the two-stage composite optimal is always more clustered or always less clustered than the standard or classical two-stage cluster optimal.

The classical optimal for the two-stage cluster design $\bar{n}_{cl.}$ coincides with $\bar{n}_{opt.}$ when $q=0$.

It is not obvious whether the two-stage general optimal $\bar{n}_{opt.}$ is larger or smaller than the classical two-stage cluster design optimal $\bar{n}_{cl.}$ when $q>0$. In fact, it is not clear that the stationary point for $\bar{n}_{opt.}$ exists at all. If ρ is small enough, then the contents of the square bracket in (15) will become negative, so that the square root will not exist.

Looking at (15) it appears that $\bar{n}_{opt.}$ will usually be less than in the classical design ($\bar{n}_{cl.}$), because in the square root of the discriminant, the coefficient of $C_{q/2}^2$ is $c_2(1-\rho)/\rho\gamma$. Usually, ρ is 0.05 or less, so that γ/c_2 becomes $\{c_1/c_2 - 19\}$. The cost of including a new PSU in the sample (c_1) will always be higher than the cost of including a new household in a selected PSU (c_2), hence the cost ratio will always be well above 1.0. The higher the cost ratio, the more costly it is to select a new PSU compared with selecting more households in selected PSUs; consequently, we should select more households in already selected PSUs. We assume that $c_1/c_2 < 19$, so the coefficient of $C_{q/2}^2$ is negative. In the term $-b$, the coefficient of $C_{q/2}^2$ is negative and in the denominator the coefficient of $C_{q/2}^2$ is positive. Hence, $\bar{n}_{opt.}$ is (usually) a decreasing function of $C_{q/2}^2$, so that for $C_{q/2}^2 > 0$, $\bar{n}_{opt.}$ will be less than the classical design. A sufficient condition for this is that $\gamma/c_2 < 0$, which would usually be satisfied, unless c_1 or ρ are unusually large. When $C_{q/2}^2 = 0$ as is the case when $N_d = \bar{N}$ the optimal sample size reduces to the standard optimal cluster size so that $\bar{n}_{opt.} = \bar{n}_{cl.}$.

Let $n_{tot} = \sum_{d\in U^1} n_d$ (note that $n_{tot} \neq n$, the sample size, since n_{tot} is the sum of n_d over all clusters in-sample and out-of-sample).

We now consider the solution of n_d given by n_d given by (8). Summing over all the clusters and dividing by the total number of clusters M we obtain

$$\frac{n_{tot}}{M} = \frac{1}{M\sqrt{\lambda c_2}} \sum_{d\in U^1} N_d^{\frac{q}{2}} \sqrt{(1-\rho)/\rho} - (1-\rho)/\rho \tag{16}$$

Solving for $\sqrt{\lambda}$ in (16) and substituting in (8) we obtain

$$n_d \quad = \quad n_{tot} P_d^{\frac{q}{2}} + (1-\rho)/\rho(MP_d^{\frac{q}{2}} - 1) \tag{17}$$

where $P_d^{\frac{q}{2}} = N_d^{\frac{q}{2}} / \sum_{d\in U^1} N_d^{\frac{q}{2}}$.

This solution for $\{n_d\}$ is identical to the area-only stratified formula for n_d given by Longford (2006), Molefe (2011) and Molefe and Clark (2015):

$$n_{h,opt.} \quad = \quad nP_h^{\frac{q}{2}} + (1-\rho)/\rho(HP_h^{\frac{q}{2}} - 1) \tag{18}$$

for stratified sampling design, with total sample size n replaced by n_{tot}. This shows that the two-stage allocation for n_d is the same as stratified allocation, given n_{tot}. We can then write the expected cost constraint (5) in terms of n_{tot} as

$$C_E \quad = \quad c_1 m + c_2 \frac{m}{M} n_{tot} \tag{19}$$

For $\frac{c_1}{c_2} = 10$, equation (14) gives a value of m which is greater than M when $q = 1$. When this happens, the optimal value for the number of clusters to take into the sample is $m = M$. As q approaches 2, the discriminant becomes negative so that there is no real-valued solution for m, implying that $m = M$ is optimal.

3.1. Numerical Example

The cost constraint for sampling is set at $C_F = 350$ cost units. The following per cluster to per subunit cost ratios are considered; $\frac{c_1}{c_2} = 10$, 4 and 2 where the cost per second stage unit $c_2 = 1$ cost units.

We used data on the 26 cantons (clusters) of Switzerland (Longford, 2006) to allocate the sample using the various simple two-stage designs. Throughout, we assume that $\rho = \frac{1}{40}$.

We compute the optimum sample sizes for each ratio $\frac{c_1}{c_2}$ by priority exponent q using (15) in Table 1. From the results, it is apparent that $\bar{n}_{opt.}$ is a decreasing function of q. As q increases the discriminant becomes small and eventually negative, resulting in the solution for $\bar{n}_{opt.}$ being negative or even a complex number. When this happens, the optimal sample size is $\bar{n}_{opt.} = 1$. We also observe that the optimum sample size decreases as $\frac{c_1}{c_2}$ decreases.

Therefore, the main finding here is that the general optimal gives a less clustered design when $q > 0$ than the classical two-stage optimal.

Table 1: Area-only simple two-stage optimum sample sizes

Priority	$\frac{c_1}{c_2} = 10$			$\frac{c_1}{c_2} = 4$			$\frac{c_1}{c_2} = 2$		
exponent	m_{opt}	$\bar{n}_{opt}$	$\bar{n}_{cl.}$	m_{opt}	$\bar{n}_{opt.}$	$\bar{n}_{cl.}$	$m_{opt.}$	$\bar{n}_{opt.}$	$\bar{n}_{cl.}$
$q = 0$	12	20	20	21	12	13	26	9	9
$q = \frac{1}{4}$	12	18	20	24	10	13	26	6	9
$q = \frac{1}{2}$	15	14	20	26	4	13	26	1	9
$q = \frac{3}{4}$	20	7	20	26	1	13	26	1	9
$q = 1$	26	1	20	26	1	13	26	1	9
$q = 2$	26	1	20	26	1	13	26	1	9

When $\frac{c_1}{c_2}$ is large, the sample is more clustered hence the CV's of the estimates of the cluster means are relatively smaller. However, the CV of the estimate of the grand mean will be large. When $\frac{c_1}{c_2}$ goes down, the sample becomes less clustered since we can take a larger number of clusters into the sample. When this happens the CV's of the estimates of the cluster means will be relatively larger since the within-cluster sample size is smaller, and the CV of the estimate of the grand mean will be smaller.

In the case of clusters of equal size, the within-cluster sample size is the same for all clusters selected into the sample. Hence, the optimization problem reduces to a singular problem of finding the optimal number of clusters to take into the sample.

The optimal number of clusters, $m_{opt.}$, and the optimal expected sample size of ultimate cluster, $\bar{n}_{opt.}$, subject to a fixed total expenditure, $C_F = c_1 m + c_2 m\bar{n}$, are $m_{opt.} = C_F/(c_1 + c_2\bar{n}_{opt.})$ where $\bar{n}_{opt.} = \sqrt{c_1/c_2(1-\rho)/\rho}$.

4. Other Designs

We consider several sensible but ad-hoc designs that include equal allocation, proportional allocation, classical optimal allocation and a combined design made up of the proportional allocation and the classical optimal design. We consider these ad-hoc designs

because sometimes the optimal design derived in Section 3 has undesirable properties such as negative or complex values for the analytical result for $\bar{n}_{cl.}$ or n_d (implying values of 1 in practice).

4.1. Equal Design

In the cluster equal design we consider the case in which a sample is taken from each and every small area (cluster). An equal number of secondary stage units is taken from each and every cluster. That is, $m = M$ and $n_d = n/M$ for $d = 1, \ldots, M$, where $n = (C_F - c_1 M)/c_2$ is the total sample size.

4.2. Proportional Design

In this design a sample is taken from each and every cluster. The within-cluster sample sizes are proportional to the population sizes of the clusters. The design is $m = M$ and $n_d = nP_d$ for $d = 1, \ldots, M$, where n is the same as in equal design and $P_d = N_d/N$.

4.3. Classical Optimal Design

The number of clusters taken into the sample is determined by the cost constraint. The within-cluster sample size is the standard optimal two-stage cluster design given by $m = C_F/(c_1 + c_2\bar{n}_{cl.})$ and $n_d = \bar{n}_{cl.}$ for $d = 1, \ldots, m$.

4.4. Proportional & Optimal Design

It may also be constructive to propose modifications of existing sampling designs. This design uses a combination of two designs. The within-cluster sample size is proportional to the cluster population size and also optimal for two-stage cluster design: $m = C_F/(c_1 + c_2\bar{n}_{cl.})$ and $n_d = P_d\bar{n}_{cl.}$ for $d = 1, \ldots, m$.

5. Numerical Evaluation

In this section, we compare the efficiency of the ad-hoc designs and the area-only optimum derived in Section 3. We consider the relative efficiency of these designs by calculating the ratios of F given by (6) of the designs using the equal design as the base design. A ratio less than one implies that a design is more efficient than the base design, whilst a ratio greater than one implies a design is less efficient than the base design.

In Table 2 we show the summary statistics of the CV's of the estimates of the cluster and national means for $\frac{c_1}{c_2} = 10$, 4 and 2 under the ad-hoc designs. The results show that for equal and classical optimum allocations, the CV's of the estimates of the small area means are narrowly dispersed by virtue of the design allocations being equal sample sizes. On the other hand, we see that the ranges of the CV's under proportional allocation and proportional & optimum allocation designs are widely dispersed since the clusters receive sample sizes that are proportionate to their population sizes.

Table 2: CV's of the ad-hoc designs

	Equal allocation	Proportional allocation	Classical optimum	Proportional & optimum	Area-only optimum[a]
$\frac{c_1}{c_2} = 10$ CV % (SAE's)					
Minimum	57.01	25.49	22.64	10.77	13.31
1st Quarter	57.01	44.16	22.65	20.50	20.16
Median	57.01	57.01	22.65	26.03	29.77
Mean	57.01	60.42	22.65	34.47	38.08
3rd Quarter	57.01	69.82	22.65	44.16	49.37
Maximum	57.01	98.74	22.65	98.74	98.74
CV % (National)	81.41	52.33	57.67	46.37	34.32
$\frac{c_1}{c_2} = 4$ CV % (SAE's)					
Minimum	32.91	15.24	27.37	13.19	12.34
1st Quarter	32.91	28.82	27.38	25.11	19.31
Median	32.91	37.61	27.39	33.07	27.39
Mean	32.91	49.54	27.38	41.87	36.05
3rd Quarter	32.91	69.82	27.39	57.01	38.82
Maximum	32.91	98.74	27.39	98.74	98.74
CV % (National)	46.75	31.75	45.71	33.75	31.24
$\frac{c_1}{c_2} = 2$ CV % (SAE's)					
Minimum	29.77	13.96	29.76	13.96	12.06
1st Quarter	29.77	26.64	29.77	26.64	18.60
Median	29.77	33.91	29.77	33.91	25.49
Mean	29.77	44.05	29.77	44.05	30.19
3rd Quarter	29.77	57.01	29.77	57.01	33.91
Maximum	29.77	98.74	29.77	98.74	69.82
CV % (National)	42.21	28.55	42.21	28.55	26.95

[a] Area-only optimum when $q = 1$

We observe that the classical optimum is relatively more efficient for estimating cluster means as shown by smaller CV's of the estimates of the cluster means compared to the other allocations. However, for estimating the national mean, the proportional & optimum allocation is relatively more efficient than the other designs. The CV of the estimate of the national mean is however considerably higher for the four ad-hoc designs, possibly showing that these two-stage cluster designs are not well suited for estimating the overall mean.

When $\frac{c_1}{c_2} = 4$, it implies more clusters and within-cluster samples for fixed C_F. The result of increased sample size is that the CV's of the estimates of the cluster means under

equal allocation and classical optimum are considerably lower than when $\frac{c_1}{c_2} = 10$. But proportional allocation and proportional & optimum allocation seem to be relatively better than equal and classical optimum allocations in estimating the national mean.

For $\frac{c_1}{c_2} = 2$ equal design give identical results to classical optimal design. Proportional design on the other hand also gives identical results to proportional & optimal design. These two designs perform better than equal and standard optimal designs for $q \geq \frac{1}{2}$ in terms of CV's and the criterion function F.

In Table 3 we see that proportional allocation and area-only stratified optimum given by (18) is less efficient than the base design. The area-only optimum designs should always be the best since the criterion function is minimized when the largest clusters are included in the sample but they are not when $q = 0$. Classical optimum and proportional & optimum are the only designs that are more efficient than equal allocation at $q = 0$. As the priority exponent q increases all the designs' efficiency against equal designs improve with the exception of classical optimum, whose efficiency is constant. At $q = 2$, the area-only stratified optimum and area-only optimum are nearly twice as efficient as equal design.

Table 3: Relative efficiency of two-stage designs for $\frac{c_1}{c_2} = 10$

		Priority Exponent (q)				
Designs	n_d	$q=0$	$q=\frac{1}{2}$	$q=1$	$q=\frac{3}{2}$	$q=2$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	0.91	0.83	0.77	0.71	0.66
Classical optimum	$\bar{n}_{opt}$	0.92	0.92	0.92	0.92	0.92
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	0.91	0.83	0.77	0.71	0.61
Area-only stratified optimum	${n_d}^1$	1.00	0.87	0.76	0.63	0.53
Area-only optimum	${n_d}^2$	1.09	0.94	0.76	0.63	0.53

Table 4: Relative efficiency of two-stage designs for $\frac{c_1}{c_2} = 4$

		Priority Exponent (q)				
Designs	n_d	$q=0$	$q=\frac{1}{2}$	$q=1$	$q=\frac{3}{2}$	$q=2$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	1.02	0.94	0.86	0.80	0.74
Classical optimum	$\bar{n}_{opt}$	0.98	0.98	0.98	0.98	0.98
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	1.02	0.94	0.87	0.82	0.77
Area-only stratified optimum	${n_d}^1$	1.00	0.92	0.81	0.69	0.58
Area-only optimum	${n_d}^2$	1.05	0.92	0.81	0.69	0.58

In Table 4 we present the relative efficiency for the two-stage designs when $\frac{c_1}{c_2} = 4$. We see that area-only stratified optimum is equally efficient as the base design whilst the area-only optimum is less efficient than the base design when $q = 0$. Also, proportional and proportional & optimum allocations are less efficient than equal allocation. The classical optimum design is the only design that is slightly more efficient than equal allocation at $q = 0$. As the priority exponent q increases all the designs' efficiency against equal designs improve with the exception of classical optimum, whose efficiency is marginal and constant. At $q = 2$, the area-only stratified optimum and the area-only optimum are almost twice as efficient as the equal design.

Table 5: Relative efficiency of two-stage designs for $\frac{c_1}{c_2} = 2$

		Priority Exponent (q)				
Designs	n_d	$q=0$	$q=\frac{1}{2}$	$q=1$	$q=\frac{3}{2}$	$q=2$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	1.03	0.94	0.85	0.78	0.72
Classical optimum	$\bar{n}_{opt}$	1.00	1.00	1.00	1.00	1.00
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	1.03	0.94	0.85	0.78	0.72
Area-only stratified optimum	${n_d}^1$	1.00	0.92	0.81	0.70	0.59
Area-only optimum	${n_d}^2$	1.00	0.92	0.81	0.70	0.59

The area-only stratified optimum and the area-only optimum compare favorably to proportional allocation and proportional & optimum. Their relative efficiency improves as the priority exponent q approaches 2. At $q = 0$, equal allocation is as good as any of these designs, even better than, for example, proportional allocation and proportional & optimum. But at $q = 2$ area-only stratified optimum and the area-only optimum are twice as efficient as equal design, whilst proportional allocation and proportional & optimum are also more efficient than equal design but to a lesser extent.

In Table 5 one can observe that the relative performance of the area-only stratified optimum and the area-only optimum (relative to equal design) are only slightly superior to proportional allocation and proportional & optimum as q approaches 2.

When $\frac{c_1}{c_2} = 2$ the relative performance of the classical optimum is the same as the base design. We observe that the performance of proportional allocation is identical to proportional & optimum design. At $q = 0$ these two designs are less efficient than equal design. The area-only stratified optimum and the area-only optimum on the other hand are more efficient than the base design.

Overall we can see that the designs relative efficiencies improves as the ratio of $\frac{c_1}{c_2}$ goes up and q approaches 2.

[1] $n_d = nP_d^{\frac{q}{2}} + (1-\rho)/\rho(MP_d^{\frac{q}{2}} - 1)$
[2] $n_d = n_{tot}P_d^{\frac{q}{2}} + (1-\rho)/\rho(mP_d^{\frac{q}{2}} - 1)$
[1] $n_d = nP_d^{\frac{q}{2}} + (1-\rho)/\rho(MP_d^{\frac{q}{2}} - 1)$
[2] $n_d = n_{tot}P_d^{\frac{q}{2}} + (1-\rho)/\rho(mP_d^{\frac{q}{2}} - 1)$

6. Sensitivity Analysis

In section 5 the numerical evaluations of the sample designs were based on an assumed value of the intraclass correlation coefficient for the Switzerland canton data (Longford, 2006). In this section selected tables are replicated using Switzerland's cantons data for different values of ρ for the two-stage cluster designs, as well as for data on the population of the administrative districts of Botswana, to investigate how the optimal sample designs are altered as a result. For the two-stage designs we consider varying ρ, and C_F for $q = 1$.

6.1. Switzerland Canton Data

Here the interest is in finding out how the values of $\frac{c_1}{c_2}$, the cost ratio, C_F, the total fixed sampling cost, and ρ, the intraclass correlation coefficient, affect these designs. To investigate this we consider the relative efficiency of these designs by fixing one parameter and varying the others.

Table 6: Relative efficiency of simple two-stage designs for $\rho = \frac{1}{40}$, $\frac{c_1}{c_2} = 10$, $q = 1$

		Sampling cost (C_F)			
Designs	n_d	$C_F = 250$	$C_F = 300$	$C_F = 350$	$C_F = 400$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	1.01	0.97	0.93	0.90
Classical optimum	$\bar{n}_{opt}$	0.88	0.90	0.92	0.94
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	1.01	0.97	0.93	0.90
Area-only stratified optimum	${n_d}^1$	0.73	0.74	0.76	0.76
Area-only optimum	${n_d}^2$	0.73	0.74	0.76	0.76

In Tables 6 - 7 we present the results of the numerical evaluation of the relative efficiency for the simple two-stage designs for $\rho = \frac{1}{4}$ and $q = 1$ when the sampling cost C_F is varied using data on the Switzerland's cantons. The results show that the area-only stratified optimum and the area-only optimum are the best designs and are identically efficient.

Table 7: Relative efficiency of simple two-stage designs for $\rho = \frac{1}{40}$, $\frac{c_1}{c_2} = 5$, $q = 1$

		Sampling cost (C_F)			
Designs	n_d	$C_F = 250$	$C_F = 300$	$C_F = 350$	$C_F = 400$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	0.92	0.89	0.88	0.86
Classical optimum	$\bar{n}_{opt}$	0.97	0.98	0.99	1.00
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	0.92	0.89	0.88	0.86
Area-only stratified optimum	${n_d}^1$	0.80	0.80	0.80	0.80
Area-only optimum	${n_d}^2$	0.80	0.80	0.80	0.80

Table 8: Relative efficiency of simple two-stage designs for $C_F = 350$, $\frac{c_1}{c_2} = 10$, $q = 1$

		Intraclass Correlation (ρ)				
Designs	n_d	$\rho = \frac{1}{1000}$	$\rho = \frac{1}{100}$	$\rho = \frac{1}{4}$	$\rho = \frac{1}{20}$	$\rho = \frac{1}{10}$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	1.00	0.97	0.93	0.89	0.84
Classical optimum	$\bar{n}_{opt}$	0.99	0.95	0.92	0.91	0.91
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	1.00	0.97	0.93	0.89	0.84
Area-only stratified optimum	${n_d}^1$	0.96	0.83	0.76	0.70	0.63
Area-only optimum	${n_d}^2$	0.96	0.83	0.76	0.76	0.76

Table 9: Relative efficiency of simple two-stage designs for $C_F = 350$, $\frac{c_1}{c_2} = 5$, $q = 1$

		Intraclass Correlation (ρ)				
Designs	n_d	$\rho = \frac{1}{1000}$	$\rho = \frac{1}{100}$	$\rho = \frac{1}{4}$	$\rho = \frac{1}{20}$	$\rho = \frac{1}{10}$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	0.99	0.93	0.88	0.83	0.80
Classical optimum	$\bar{n}_{opt}$	1.00	0.99	0.99	0.99	1.03
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	0.99	0.93	0.88	0.83	0.80
Area-only stratified optimum	${n_d}^1$	0.96	0.85	0.80	0.76	0.74
Area-only optimum	${n_d}^2$	0.96	0.85	0.80	0.76	0.74

In Tables 8 - 9 we consider the relative efficiency of the designs for $C_F = 350$ cost units and $q = 1$ when ρ is varied. The results show that the area-only stratified optimum and the area-only optimum with partial coverage are the best designs for small values of ρ. As ρ increases, the area-only stratified optimum is the best design, with the area-only optimum nearly as good when $\rho = \frac{1}{40}$.

6.2. Botswana District Data

In this section we investigate the new sample designs for different data. We use data for the administrative districts of Botswana published by the Central Statistics Office (CSO). The population of Botswana is 1.67 million (Central Statistics Office, 2002). Botswana is divided into 16 administrative districts comprising major cities, towns and tribal territories. The smallest district is a mining town of Sowa with a population of almost 3,000 persons and the largest is Central district with a population of just over half a million inhabitants as per the 1991 population and housing census (Central Statistics Office, 2002).

For the simple two-stage designs we are interested in finding out whether the values of C_F and ρ has any effect on these designs. To investigate this we consider the relative

[1] $n_d = nP_d^{\frac{q}{2}} + (1-\rho)/\rho(MP_d^{\frac{q}{2}} - 1)$
[2] $n_d = n_{tot}P_d^{\frac{q}{2}} + (1-\rho)/\rho(mP_d^{\frac{q}{2}} - 1)$
[1] $n_d = nP_d^{\frac{q}{2}} + (1-\rho)/\rho(MP_d^{\frac{q}{2}} - 1)$
[2] $n_d = n_{tot}P_d^{\frac{q}{2}} + (1-\rho)/\rho(mP_d^{\frac{q}{2}} - 1)$

efficiency of these designs by fixing one parameter and varying the others.

In Tables 10 - 11 we present the results of the numerical evaluation of the relative efficiency for the simple two-stage designs for $\rho = \frac{1}{10}$ and $q = 1$ when the sampling cost C_F is varied using data on Botswana administrative data. The results show that the area-only stratified optimum given by (18) and the area-only optimum given by (17) are the best designs and are identical.

Table 10: Relative efficiency of simple two-stage designs for $\rho = \frac{1}{10}$, $\frac{c_1}{c_2} = 10$, $q = 1$

		Sampling cost (C_F)			
Designs	n_d	$C_F = 250$	$C_F = 300$	$C_F = 350$	$C_F = 400$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	0.78	0.80	0.82	0.78
Classical optimum	$\bar{n}_{opt}$	0.96	1.00	1.00	1.00
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	0.90	0.83	0.77	0.78
Area-only stratified optimum	${n_d}^1$	0.68	0.69	0.71	0.72
Area-only optimum	${n_d}^2$	0.68	0.69	0.71	0.72

Table 11: Relative efficiency of simple two-stage designs for $\rho = \frac{1}{10}$, $\frac{c_1}{c_2} = 5$, $q = 1$

		Sampling cost (C_F)			
Designs	n_d	$C_F = 250$	$C_F = 300$	$C_F = 350$	$C_F = 400$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	0.82	0.85	0.87	0.90
Classical optimum	$\bar{n}_{opt}$	1.00	1.00	1.00	1.00
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	0.77	0.77	0.78	0.78
Area-only stratified optimum	${n_d}^1$	0.72	0.74	0.74	0.75
Area-only optimum	${n_d}^2$	0.72	0.74	0.74	0.75

Table 12: Relative efficiency of simple two-stage designs for $C_F = 350$, $\frac{c_1}{c_2} = 10$, $q = 1$

		Intraclass Correlation (ρ)				
Designs	n_d	$\rho = \frac{1}{1000}$	$\rho = \frac{1}{100}$	$\rho = \frac{1}{4}$	$\rho = \frac{1}{20}$	$\rho = \frac{1}{10}$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N}n$	0.98	0.89	0.83	0.80	0.82
Classical optimum	$\bar{n}_{opt}$	0.99	0.98	0.98	1.00	1.00
Proportional & optimum	$\frac{N_d}{N}\bar{n}_{opt}$	1.01	0.99	0.95	0.85	0.77
Area-only stratified optimum	${n_d}^1$	0.99	0.89	0.81	0.75	0.71
Area-only optimum	${n_d}^2$	0.99	0.89	0.81	0.75	0.71

In Tables 12 - 13 we consider the relative efficiency of the designs for $C_F = 350$ cost units and $q = 1$ when ρ is varied. The results show that the area-only stratified optimum and the area-only optimum with partial coverage are the best designs for all values of ρ. The proportional & optimum design is nearly as good.

Table 13: Relative efficiency of simple two-stage designs for $C_F = 350$, $\frac{c_1}{c_2} = 5$, $q = 1$

		Intraclass Correlation (ρ)				
Designs	n_d	$\rho = \frac{1}{1000}$	$\rho = \frac{1}{100}$	$\rho = \frac{1}{4}$	$\rho = \frac{1}{20}$	$\rho = \frac{1}{10}$
Equal allocation	$\frac{n}{M}$	1.00	1.00	1.00	1.00	1.00
Proportional allocation	$\frac{N_d}{N} n$	0.98	0.87	0.82	0.82	0.87
Classical optimum	$\bar{n}_{opt}$	1.00	1.00	1.00	1.00	1.00
Proportional & optimum	$\frac{N_d}{N} \bar{n}_{opt}$	0.99	0.93	0.80	0.78	0.78
Area-only stratified optimum	${n_d}^1$	0.98	0.89	0.81	0.76	0.74
Area-only optimum	${n_d}^2$	0.98	0.89	0.81	0.76	0.74

In this section we have used Switzerland canton data and Botswana district data. We have replicated the numerical evaluation of the various designs by considering the relative efficiencies of the designs, by computing the values of F for designs under consideration. We considered relative priority exponent $q = 1$ and selected values of the relative priority coefficient. Selected tables are replicated using Switzerland's cantons data for different values of ρ for the stratified designs, as well as for data on the population of the administrative districts for Botswana to investigate how the optimal sample designs are modified as a result. For the two-stage designs we consider varying $\frac{c_1}{c_2}$, ρ, and C_F for fixed q.

To investigate whether the value of ρ, the intraclass correlation, has an effect on the stratified allocations, we consider different values of ρ whilst keeping the priority coefficient and priority exponent fixed, for these designs. When $q = 1$ proportional allocation and optimal power allocation are the best designs when $\rho = \frac{1}{1000}$. As ρ increases, all designs are the best except for proportional allocation and equal allocation.

For the simple two-stage designs we are interested in finding out whether the values of C_F and ρ has any effect on the choice of the within-cluster sample size. The results as in section 5, show that the area-only stratified optimum given by equation 18 and the area-only optimum given by equation 17 are the best designs. When ρ is varied for fixed C_F and $q = 1$, the results show that the area-only stratified optimum and the area-only optimum with partial coverage are the best designs for all values of ρ.

7. Conclusions

An analytical solution for the stationary point exists when the only priority is small area estimation. This optimal design is less clustered than the usual classical two-stage optimal sample size $\bar{n}_{cl.}$ when more priority is given to larger clusters ($q > 0$). The optimal sample size depends on the cost per cluster relative to ($\frac{c_1}{c_2}$), intraclass correlation coefficient (ρ) and the relative variance of $N_d^{\frac{q}{2}}$ denoted by $C^2_{q/2}$. When the only priority is small area

estimation, that is, $q = 0$, or when the N_d's are constant, $C^2_{q/2} = 0$ and the general optimal coincides with the classical optimal. The area-only optimal average sample size is usually a decreasing function of $C^2_{q/2}$, so that when $C^2_{q/2} > 0$, $\bar{n}_{opt.}$ will be less than the classical optimum. A sufficient condition for this is that $\gamma/c_2 < 0$, which would usually be satisfied, unless $\frac{c_1}{c_2}$ or ρ are unusually large.

The area-only stratified optimum and the area-only simple two-stage optimum should always be the best designs in minimizing the objective function but they are not when there is equal priority for each cluster, that is when $q = 0$. These two designs have undesirable properties of allocating zero or even negative sample sizes to smaller clusters. Negative sample sizes are obviously not possible in practice and this anomaly is corrected by setting them to zero and reallocating again.

When the clusters are equally important ($q = 0$), classical optimum and proportional & optimum are the best designs especially when the cost ratio is high, in this case when $\frac{c_1}{c_2} = 10$. When $\frac{c_1}{c_2} = 2$, proportional design and proportional & optimum design are less efficient than equal allocation. Also, the classical optimum is as efficient as equal allocation. All the other designs are better as q approaches 2, with area-only stratified optimum and the area-only optimum being the best.

References

Bankier, M. D., (1988). Power allocations: determining sample sizes for subnational areas. *The American Statistician*, 42(3), pp. 174–177.

Clark, R. G., Steel, D. G., (2007). Sampling within households in households surveys. *Journal of the Royal Statistical Society A*, 170(Issue 1), pp. 63–82.

Cochran, W. G., (1977). Sampling Techniques. Wiley and Sons.

Foreman, E. K., (1991). Survey Sampling Principles. Marcel Dekker, Inc.

Fuller, W. A., (1999). Environmental Surveys Over Time. Journal of Agricultural, Biological and Environmental Statistics, 4, pp. 331–345.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G., (1953). Sample survey methods and theory, volume 1 & 2. Wiley, New York.

Hidiroglou, M. A., Patak, Z., (2004). Domain estimation using linear regression. *Survey Methodology*, 30, pp. 67–78.

Longford, N. T., (2006). Sample size calculation for small-area estimation. Survey Methodology, 32(1), pp. 87–96.

Molefe, W. B., (2011). Sample design for small area estimation. PhD thesis, University of Wollongong, http://ro.uow.edu.au/theses/3495.

Molefe, W. B., Clark, R. G., (2015). Model-assisted optimal allocation for planned domains using composite estimation. *Survey Methodology*, 41(2), pp. 377387.

Molefe, W. B., Shangodoyin, D. K., and Clark, R. G., (2015). An approximation to the optimal subsample allocation for small areas. *Statistics in Transition, new series*, 16(2), pp. 163–182.

Pettersson, H., Sisouphanthong, B., (2005). Household Sample Surveys in Developing and Transition Countries, chapter Cost Model for an Income and Expenditure Survey, pages 267–277. Number 96 in Series F. United Nations: Statistics Division, Department of Economic and Social Affairs.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Asok, C., (1954). Sampling theory of surveys with applications. Iowa State University Press, third edition.