# A design-based procedure for the analysis of experiments embedded in complex sample surveys

## Discussion paper 03007

*J.A. van den Brakel*

Invited paper for XX International Symposium on Survey Methodology
29-31 October 2003, Ottawa, Statistics Canada.


Statistics Netherlands

Voorburg/Heerlen, September 2003

**Explanation of symbols**

| | |
|---|---|
| . | = data not available |
| * | = provisional figure |
| x | = publication prohibited (confidential figure) |
| – | = nil or less than half of unit concerned |
| – | = (between two figures) inclusive |
| 0 (0,0) | = less than half of unit concerned |
| blank | = not applicable |
| 2002–2003 | = 2002 to 2003 inclusive |
| 2002/2003 | = average of 2002 up to and including 2003 |
| 2002/'03 | = crop year, financial year, school year etc. beginning in 2002 and ending in 2003 |

Due to rounding, some totals may not correspond with the sum of the separate figures.

**Publisher**
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg
The Netherlands

**Printed by**
Statistics Netherlands - Facility Services

**Cover design**
WAT ontwerpers, Utrecht

**Information**
E-mail: infoservice@cbs.nl

**Where to order**
E-mail: verkoop@cbs.nl

**Internet**
http://www.cbs.nl



ISSN: 1572-0314
Key figure: X-10
Production code: 6008303007

Statistics Netherlands

# A DESIGN BASED PROCEDURE FOR THE ANALYSIS OF EXPERIMENTS EMBEDDED IN COMPLEX SAMPLE SURVEYS

Jan A. van den Brakel[1]

*Summary: Experiments embedded in ongoing sample surveys are particularly appropriate to test effects of alternative survey methodologies on estimates of finite population parameters. Ignoring the sampling design in the analysis of such experiments, typically leads to design-biased parameter and variance estimates, which makes the analysis results incommensurable with the estimation results of the regular survey. To test hypotheses about differences between sample estimates observed under alternative survey methodologies, a design-based approach for the analysis experiments embedded in complex sample surveys is proposed.*

*Keywords: completely randomized designs, generalized regression estimator, measurement error models, randomized block designs, Wald statistic*

## 1. Introduction

Generally, survey sampling is a very laborious process that potentially invokes many sources of non-sampling errors. Therefore a lot of research in the field of survey methodology is aimed at the improvement of the quality and efficiency of sample survey processes. An important part of this research is to consider and test alternative survey methodologies. Large-scale field experiments embedded in ongoing sample surveys are particularly appropriate to quantify the effect of alternative survey implementations or approaches on response behaviour or sample estimates. In National Statistical Offices, sample surveys are generally kept unchanged as long as possible to construct uninterrupted time series of the estimated population parameters. It remains inevitable, however, that survey processes be adjusted from time to time. In such situations, embedded experiments can be applied to detect and quantify possible trend disruptions in time series due to adjustments of the survey process. Applications can be found in Van den Brakel and Renssen (1998). To compare the effect of $K$ different survey approaches on the main estimates of the finite population parameters of an ongoing survey, a sample drawn from a finite population is randomly divided into $K$ subsamples according to an experimental design. The sampling units of each subsample are assigned to one of the $K$ survey approaches or treatments. Generally there is one large subsample that is assigned to the regular survey, which serves for official publication purposes as well as the control group in the experiment. The purpose of such embedded experiments

[1] Jan A. van den Brakel, Department of Statistical Methods, Statistics Netherlands, P.O. Box 4481, 6401 CZ, Heerlen, The Netherlands, E-mail:JBRL@CBS.NL.

is the estimation of finite population parameters observed under different survey implementations or treatments and to test hypotheses about the differences between these sample estimates. In this paper a design-based analysis procedure for embedded experiments, designed as a completely randomized design (CRD) or a randomized block design (RBD), is proposed. Design-unbiased estimators for the population parameters observed under each of the $K$ treatments as well as design-unbiased estimators for the covariance matrix of the contrasts between these estimates are derived using the Horvitz-Thompson estimator or the generalized regression estimator. These estimators account for the sampling design, the experimental design and the weighting procedure of the ongoing survey and give rise to a design-based Wald statistic to test hypotheses about differences between finite population parameter estimates of the sample survey.

## 2. Embedding experiments in sample surveys

The embedding of experiments in sample surveys is sometimes also referred to as the superimposition of experiments on sample surveys. First a sample is drawn from a finite target population by means of a generally complex sampling design. Next the sample is according to an experimental design randomly divided into $K$ subsamples. The most straightforward approach is to use a CRD. The application of unrestricted randomization, however, is generally not the most efficient design available. Fienberg and Tanur (1987, 1988) argued that the application of an RBD where sampling structures like strata, primary sampling units (PSU's), clusters, interviewers and the like might improve the precision of an experiment considerably. Moreover, unrestricted randomization is not always feasible from a practical point of view. For example in CAPI surveys where interviewers collect data in geographical areas around their place of residence, restricted randomization of sampling units within interviewers or geographical regions which are unions of adjacent interviewer regions might be required to avoid an unacceptable increase of the travel distance of the interviewers. This naturally leads to RBD's with interviewers or regions as block variables, see e.g. Van den Brakel and Van Berkel (2002).

The main advantages of embedded experiments is the random selection of the sampling units from a finite target population, which increases the possibility to generalize the results observed in the sample of the experiment to larger populations (Fienberg and Tanur, 1987, 1988). From a financial point of view the conduction of embedded experiments is efficient since the regular survey is also used as control group. In some situations the data obtained in the subsamples assigned to the alternative treatments can even be used for the regular publication purposes of the ongoing survey, see e.g. Van den Brakel (2002). Furthermore, embedded experiments form a safe transition from an old to a new survey design. Running the old and new approach in parallel by means of an embedded experiment creates the possibility to quantify possible trend disruptions and one can still fall back on the old approach for the regular publication purposes if the new approach turns out to be a failure. Nevertheless it should be realized that in an embedded experiment two more or less competing objectives are combined. According to the purpose of the regular survey, that is the estimation of population parameters as precise as possible, the subsample assigned to regular survey should be maximized. According to the purpose of the experiment, that is the estimation of contrasts between the subsample estimates as precise as possible, the subsample sizes are preferably equal, since balanced designs maximize the power of the tests about treatment effects (see e.g. Montgomery, 1997).

## 3. Analysis of embedded experiments

The main purpose of embedded experiments is the estimation of subsample means observed under different treatments and to test hypotheses about the differences between these subsample estimates. This requires an analysis procedure that accounts for the sampling design, the experimental design and the estimation procedure applied in the regular sample survey. The application of such an analysis procedure is also a natural consequence of the random selection of sampling units from a target population, with the purpose to generalize the observed results to populations larger then the sample included in the experiment.

### 3.1 Measurement error models

Although the analysis procedure proposed in this paper is predominantly design-based, some use is made of measurement error models. Design-based sampling theory is largely based on the traditional notion that observations obtained from sampling units are true fixed values observed without error (e.g. Cochran, 1977). This approach, however, is not tenable if experiments are conducted to test systematic differences between finite population parameter estimates due to different survey implementations or non-sampling errors. Therefore a measurement error model has to be introduced to relate systematic differences between estimates of finite population parameters to different survey implementations or treatments. We will follow the approach to measurement error modeling in surveys of Biemer and Stokes (1991). It is assumed that the observations obtained in the experiment are a realization of the following measurement error model:

$$y_{ikl}^{\alpha} = u_i + \beta_k + \gamma_l^{\alpha} + \varepsilon_{ik}^{\alpha} .$$

Here $y_{ikl}^{\alpha}$ is the response of sampling unit *i* assigned to treatment *k* and interviewer *l* on the $\alpha$ -th occasion, $u_i$ the true intrinsic value of sampling unit *i*, $\beta_k$ an additive effect of treatment *k*, $\gamma_l^{\alpha}$ an effect of interviewer *l* and $\varepsilon_{ik}^{\alpha}$ an error component on the $\alpha$ -th occasion that the intrinsic value of sampling unit *i* is measured under treatment *k*. The model allows for mixed interviewer effects, i.e. $\gamma_l^{\alpha} = \psi_l + \xi_l^{\alpha}$, where $\psi_l$ and $\xi_l^{\alpha}$ denote the fixed and random interviewer effects of interviewer *l*.

The superscript $\alpha$ expresses which variables are random with respect to the measurement error model.

Since for each sampling unit $K$ response variables are defined under $K$ different treatments, the measurement error model can be expressed in matrix notation as

$$\mathbf{y}_{il}^{\alpha} = \mathbf{j}u_i + \boldsymbol{\beta} + \mathbf{j}\gamma_l^{\alpha} + \boldsymbol{\varepsilon}_i^{\alpha}, \tag{3.1}$$

where $\mathbf{y}_{il}^{\alpha} = (y_{i1l}^{\alpha}, \ldots, y_{iKl}^{\alpha})^t$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^t$, $\boldsymbol{\varepsilon}_i^{\alpha} = (\varepsilon_{i1}^{\alpha}, \ldots \varepsilon_{iK}^{\alpha})^t$ and $\mathbf{j}$ a $K$-vector with each element equal to one. Let $E_{\alpha}$ and $Cov_{\alpha}$ denote the expectation and the covariance with respect to the measurement error model. It is assumed that $E_{\alpha}(\xi_l^{\alpha}) = 0$ and that random interviewer effects between interviewers are independent. Furthermore, it is assumed that $E_{\alpha}(\boldsymbol{\varepsilon}_i^{\alpha}) = \mathbf{0}$ and that measurement errors between different sampling units are independent. Hence

$$E_{\alpha}(\mathbf{y}_{il}^{\alpha}) = \mathbf{j}u_i + \boldsymbol{\beta} + \mathbf{j}\psi_l,$$

and

$$Cov_{\alpha}(\mathbf{y}_{il}^{\alpha}, \mathbf{y}_{i'l'}^{\alpha}) = \begin{cases} Var_{\alpha}(\boldsymbol{\varepsilon}_i^{\alpha}) + \mathbf{j}\mathbf{j}^t Var_{\alpha}(\xi_l^{\alpha}) & : i = i' \text{and } l = l' \\ \mathbf{j}\mathbf{j}^t Var_{\alpha}(\xi_l^{\alpha}) & : i \neq i' \text{and } l = l' \\ \mathbf{O} & : i \neq i' \text{and } l \neq l' \end{cases}.$$

Obviously, any correlation between the response of different sampling units assigned to the same interviewer can be modeled by means of the random interviewer effects. Any fixed interviewer effect influences the bias of the response values.

### 3.2 Hypothesis testing

After having defined a measurement error model for the observations obtained in the experiment, we can relate systematic differences between population parameters due to different survey implementations. Suppose that $L$ interviewers are available for the conduction of the fieldwork. The population $U$ of size $N$ can conceptually be divided into $L$ groups $U_l$ of size $N_l$ such that all sampling units within a group are potentially interviewed by the same interviewer. Let $\overline{\mathbf{Y}}^{\alpha} = (\overline{Y}_1^{\alpha}, \ldots, \overline{Y}_K^{\alpha})^t$ denote the $K$ dimensional vector of population means of $\mathbf{y}_i^{\alpha}$, i.e.

$$\overline{\mathbf{Y}}^{\alpha} = \mathbf{j}\frac{1}{N}\sum_{i=1}^{N} u_i + \boldsymbol{\beta} + \mathbf{j}\sum_{l=1}^{L}\frac{N_l}{N}\psi_l + \mathbf{j}\sum_{l=1}^{L}\frac{N_l}{N}\xi_l^{\alpha} + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i^{\alpha}.$$

Since we are interested in systematic differences between population means observed under the $K$ different treatments, random deviations between the components of $\overline{\mathbf{Y}}^{\alpha}$ should not lead to significant differences in the analysis of the experiment. This is accomplished by formulating hypotheses about

$$E_{\alpha}\overline{\mathbf{Y}}^{\alpha} = \mathbf{j}\frac{1}{N}\sum_{i=1}^{N} u_i + \boldsymbol{\beta} + \mathbf{j}\sum_{l=1}^{L}\frac{N_l}{N}\psi_l \equiv \overline{\mathbf{Y}},$$

and consequently results in the formulation of the following hypothesis:

$$\begin{aligned} H_0 &: \mathbf{C}\overline{\mathbf{Y}} = \mathbf{0} \\ H_1 &: \mathbf{C}\overline{\mathbf{Y}} \neq \mathbf{0} \end{aligned} . \qquad (3.2)$$

Here $\mathbf{C} = (\mathbf{j} \mid -\mathbf{I})$ denotes a $(K-1)\times K$ contrast matrix. Since $\mathbf{C}\overline{\mathbf{Y}} = \mathbf{C}\boldsymbol{\beta}$, it follows that hypothesis (3.2) concerns the treatment effects as represented by $\boldsymbol{\beta}$ in the measurement error model. Consequently, the differences between the population means exactly correspond to the treatment effects. Now the hypothesis about treatment effects can be tested by estimating $\overline{\mathbf{Y}}$ instead of $\boldsymbol{\beta}$, where we account for the sampling design, the experimental design and the weighting procedure of the regular sample survey. If $\hat{\overline{\mathbf{Y}}}^{\alpha}$ denote such a design-unbiased estimator for $\overline{\mathbf{Y}}$ and $\mathbf{V}$ the covariance matrix of $\hat{\overline{\mathbf{Y}}}^{\alpha}$, then hypothesis (3.2) can be tested with the Wald statistic $W = \hat{\overline{\mathbf{Y}}}^{\alpha^t}\mathbf{C}^t(\mathbf{C}\mathbf{V}\mathbf{C}^t)^{-1}\mathbf{C}\hat{\overline{\mathbf{Y}}}^{\alpha}$.

### 3.3 Estimation of subsample means

To test hypothesis (3.2), a sample *s*, drawn from a finite population *U* of size *N* is available. Let $\pi_i$ and $\pi_{ii'}$ denote the first and second order inclusion probabilities of the *i*-th and *i,i'*-th sampling unit(s) respectively with respect to this sampling design. In the case of a CRD, *s* is randomly divided into *K* subsamples $s_k$ of size $n_k$. If $n_+ = \sum_{k=1}^{K} n_k$ denotes the number of sampling units of *s*, then the conditional probability that sampling unit *i* is assigned to treatment *k*, given the realization of *s*, equals $n_k / n_+$. In the case of an RBD the sampling units of *s* are deterministically divided in *B* blocks $s_b$. The sampling units within each block, are randomized over the *K* treatments. Let $n_{bk}$ denote the number of sampling units assigned to treatment *k* in block *b*. Then $n_{b+} = \sum_{k=1}^{K} n_{bk}$ denotes the size of $s_b$, $n_{+k} = \sum_{b=1}^{B} n_{bk}$ denotes the size of subsample $s_k$ assigned to treatment *k* and $n_{++} = \sum_{b=1}^{B}\sum_{k=1}^{K} n_{bk}$ denotes the size of sample *s*. The conditional probability that sampling unit *i* is assigned to treatment *k*, given the realization of *s* and that $i \in s_b$, equals $n_{bk} / n_{b+}$. Each subsample $s_k$ can be considered as a two-phase sample, where the first phase corresponds to the sampling design used to draw sample *s* and the second phase to the experimental design used to divide *s* into *K* subsamples $s_k$. As a result the first order inclusion probabilities with respect to the subsamples are given by $\pi_i^* = (n_k / n_+)\pi_i$ in the case of a CRD and $\pi_i^* = (n_{bk} / n_{b+})\pi_i$ in the case of an RBD. Now, the Horvitz-Thompson estimator for $\overline{Y}_k^{\alpha}$ based on the observations obtained in subsample $s_k$ is given by

$$\hat{\overline{Y}}_k^{\alpha} = \frac{1}{N}\sum_{i \in s_k} \frac{y_{ik}^{\alpha}}{\pi_i^*}.$$

To allow for the weighting procedure of the regular survey, the analysis is based on the generalized regression estimator. The use of auxiliary information by means of this estimator has the advantage that it might reduce the design variance of the subsample estimates and that it corrects, at least partially, for the bias due to selective nonresponse (Bethlehem, 1988). In the present context the generalized regression estimator represents a design-based analogy of covariance analysis in standard experimental design methodology. Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{iH})^t$ a vector of order *H* with each element $x_{ih}$ an auxiliary variable of sampling unit *i*. According to the model assisted approach of Särndal et al. (1992), the intrinsic values $u_i$ in the measurement error model for each unit in the population are assumed to be an independent realization of the linear regression model:

$$u_i = \mathbf{b}^t \mathbf{x}_i + e_i, \tag{3.3}$$

where $\mathbf{b}$ denotes a vector of order $H$ with regression coefficients and $e_i$ the residuals of the regression model. Let $\omega_i^2$ denote the variance of $e_i$. It is required that all $\omega_i^2$ be known up to a common scale factor; that is $\omega_i^2 = v_i \omega^2$, with $v_i$ known. The finite population means of these variables are denoted by the vector $\overline{\mathbf{X}}$. Let

$$\mathbf{b}_k^\alpha = \left( \sum_{i \in U} \frac{\mathbf{x}_i \mathbf{x}_i^t}{\omega_i^2} \right)^{-1} \sum_{i \in U} \frac{\mathbf{x}_i y_{ik}^\alpha}{\omega_i^2}$$

denote the population regression coefficients of $\mathbf{b}$ in (3.3) observed under treatment *k*,

$$\hat{\mathbf{b}}_k^\alpha = \left( \sum_{i \in s_k} \frac{\mathbf{x}_i \mathbf{x}_i^t}{\pi_i^* \omega_i^2} \right)^{-1} \sum_{i \in s_k} \frac{\mathbf{x}_i y_{ik}^\alpha}{\pi_i^* \omega_i^2} \tag{3.4}$$

the Horvitz-Thompson estimator for $\mathbf{b}_k^\alpha$ based on the observations obtained in $s_k$ and

$$\hat{\overline{\mathbf{X}}}_k = \frac{1}{N} \sum_{i \in s_k} \frac{\mathbf{x}_i}{\pi_i^*} \tag{3.5}$$

the Horvitz-Thompson estimator of $\overline{\mathbf{X}}$ based on the sampling units in $s_k$. Now the generalized regression estimator for $\overline{Y}_k^\alpha$ based on the observations obtained in subsample $s_k$ is given by

$$\hat{\overline{Y}}_{k;greg}^\alpha = \hat{\overline{Y}}_k^\alpha + \hat{\mathbf{b}}_k^{\alpha t} (\overline{\mathbf{X}} - \hat{\overline{\mathbf{X}}}_k). \tag{3.6}$$

The generalized regression estimator can be approximated by a first order Taylor linearization about $(\overline{Y}_k, \mathbf{b}_k, \overline{\mathbf{X}})$, where $\mathbf{b}_k = E_\alpha(\mathbf{b}_k^\alpha)$ and is given by

$$\hat{\overline{Y}}_{k;greg}^\alpha \approx \hat{\overline{Y}}_k^\alpha + \mathbf{b}_k^t (\overline{\mathbf{X}} - \hat{\overline{\mathbf{X}}}_k). \tag{3.7}$$

Note that $\hat{\overline{Y}}_{k;greg}^\alpha$ is an approximately design-unbiased estimator for $\overline{Y}_k^\alpha$. Since $E_\alpha \overline{Y}_k^\alpha = \overline{Y}_k$, $\hat{\overline{Y}}_{k;greg}^\alpha$ is also an approximately unbiased estimator for $\overline{Y}_k$. The vector $\hat{\overline{\mathbf{Y}}}_{\mathbf{GREG}}^\alpha = (\hat{\overline{Y}}_{1;greg}^\alpha, \ldots, \hat{\overline{Y}}_{K;greg}^\alpha)^t$ is an approximately unbiased estimator for $\overline{\mathbf{Y}}$.

### 3.4 Variance estimation of contrasts between subsample means

Let $\mathbf{V}$ denote the covariance matrix of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{GREG}}$. Since the subsamples are drawn from a finite population without replacement, the subsample estimates are dependent. As a result there is nonzero design covariance between the elements of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{GREG}}$. An estimator for $\mathbf{V}$ requires vectors $\mathbf{y}^{\alpha}_{i}$ containing the observations of all $K$ treatments from each sampling unit. In the experimental designs under consideration, however, each sampling unit is assigned to only one of the $K$ treatments and thus only one of the components of $\mathbf{y}^{\alpha}_{i}$, for $i \in s$ is actually observed. Therefore, a design-unbiased estimator for $\mathbf{V}$ cannot be obtained. Van den Brakel and Binder (2000) try to overcome this problem by imputing the unobserved components. The usefulness of their results, however, depends on the correctness of the imputation model. The problem of the missing observation can also be circumvented by deriving an approximately design-unbiased estimator for the covariance matrix of the *K-1* contrasts $\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{GREG}}$, denoted as $\mathbf{CVC}^{t}$. Let $E_s$ and $E_e$ denote the expectation with respect to the sampling design and the experimental design, respectively. Equivalently $Cov_s$ and $Cov_e$ denote the covariance with respect to the sampling design and the experimental design. Consider the following variance decomposition:

$$\mathbf{CVC}^{t} = Cov_{\alpha}E_{s}E_{e}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{GREG}}) + E_{\alpha}Cov_{s}E_{e}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{GREG}}) + E_{\alpha}E_{s}Cov_{e}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{GREG}}) \quad (3.8)$$

Under the condition that there exists a constant vector $\mathbf{a}$ of order *H*, such that $\mathbf{a}^{t}\mathbf{x}_{i} = 1$ for all $i \in U$, i.e. that at least the size of the finite population is used as auxiliary information in the weighting scheme, Van den Brakel (2001) proves under the assumptions of measurement error model (3.1) that

$$\begin{aligned}
Cov_{\alpha}E_{s}E_{e}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{GREG}}) &= \frac{1}{N^2}\sum_{i\in U}\mathbf{C}\boldsymbol{\Sigma}_{i}\mathbf{C}^{t} \\
E_{\alpha}Cov_{s}E_{e}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{GREG}}) &= \frac{1}{N^2}\sum_{i\in U}\left(\frac{1}{\pi_i}-1\right)\mathbf{C}\boldsymbol{\Sigma}_{i}\mathbf{C}^{t} \qquad (3.9)\\
E_{\alpha}E_{s}Cov_{e}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{GREG}}) &= E_{\alpha}E_{s}(\mathbf{CDC}^{t}) - \frac{1}{N^2}\sum_{i\in U}\frac{\mathbf{C}\boldsymbol{\Sigma}_{i}\mathbf{C}^{t}}{\pi_i}
\end{aligned}$$

where $\boldsymbol{\Sigma}_{i} = Var_{\alpha}(\boldsymbol{\varepsilon}^{\alpha}_{i})$ denotes the covariance matrix of the measurement errors, and $\mathbf{D}$ denotes a $K \times K$ diagonal matrix with elements

$$d_k = \frac{1}{n_k}\frac{1}{(n_{+}-1)}\sum_{i\in s}\left(\frac{n_{+}(y^{\alpha}_{ik}-\mathbf{b}^{t}_{k}\mathbf{x}_{i})}{N\pi_i} - \frac{1}{n_{+}}\sum_{i'\in s}\frac{n_{+}(y^{\alpha}_{i'k}-\mathbf{b}^{t}_{k}\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2 \equiv \frac{S^2_k}{n_k} \qquad (3.10)$$

in the case of a CRD and

$$d_k = \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{(n_{b+}-1)} \sum_{i \in s_b} \left( \frac{n_{b+}(y_{ik}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_{b+}} \sum_{i' \in s_b} \frac{n_{b+}(y_{i'k}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2 \equiv \sum_{b=1}^{B} \frac{S_{bk}^2}{n_{bk}} \tag{3.11}$$

in the case of an RBD. Inserting the three components in (3.9) into (3.8) gives $\mathbf{CVC}^t = E_{\alpha} E_s \mathbf{CDC}^t$. Conditionally on $\alpha$ and *s*, we can directly derive an approximately design-unbiased estimator for $\mathbf{D}$. Therefore, $\mathbf{CVC}^t$ can conveniently stated implicitly as the expectation over the measurement error model and the sampling design. Expressions for $E_{\alpha} E_s d_k$ are derived in Van den Brakel (2001) under different sampling designs for CRD's and RBD's. Given the realization of $\alpha$ and *s*, the allocation of the sampling units to subsample $s_k$ can be considered as simple random sampling without replacement from *s* in the case of a CRD and simple random sampling without replacement from $s_b$ in the case of an RBD. Consequently a design-unbiased estimator for $d_k$ is given by

$$\hat{d}_k = \frac{1}{n_k} \frac{1}{(n_k - 1)} \sum_{i \in s_k} \left( \frac{n_{+}(y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_k} \sum_{i' \in s_k} \frac{n_{+}(y_{i'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2 \equiv \frac{\hat{S}_k^2}{n_k} \tag{3.12}$$

in the case of a CRD and

$$\hat{d}_k = \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{(n_{bk}-1)} \sum_{i \in s_{bk}} \left( \frac{n_{b+}(y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_{bk}} \sum_{i' \in s_{bk}} \frac{n_{b+}(y_{i'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2 \equiv \sum_{b=1}^{B} \frac{\hat{S}_{bk}^2}{n_{bk}} \tag{3.13}$$

in the case of an RBD, where $s_{bk}$ denote the $n_{bk}$ sampling units of block *b* assigned to treatment *k*. As a result, an approximately design-unbiased estimator for $\mathbf{CVC}^t$ is given by $\mathbf{C\hat{D}C}^t$, where the diagonal elements of $\hat{\mathbf{D}}$ are defined by (3.12) or (3.13).

Results for the extended Horvitz-Thompson estimator follows as a special case from the results obtained for the generalized regression estimator with the common mean model as weighting scheme (Särndal et al., 1992, section 7.4), i.e. $(x_i) = 1$ and $\omega_i^2 = \omega^2$. The common mean model only uses the population total as auxiliary information and thus satisfies the condition that there exists a constant *H*-vector such that $\mathbf{a}^t \mathbf{x}_i = 1$ for all $i \in U$. Under this weighting scheme it follows that

$$\hat{\bar{Y}}_{k;greg}^{\alpha} = \left( \sum_{i \in s_k} \frac{1}{\pi_i^*} \right)^{-1} \left( \sum_{i \in s_k} \frac{y_{ik}^{\alpha}}{\pi_i^*} \right) \equiv \tilde{Y}_k^{\alpha}, \tag{3.14}$$

and $\hat{\mathbf{b}}_k^{\alpha} = \tilde{Y}_k^{\alpha}$. An approximately design-unbiased estimator for the covariance matrix of the contrasts between the subsample means is given by (3.12) or (3.13) for a CRD or an RBD respectively, where $\hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_i = \tilde{Y}_k^{\alpha}$.

The variance estimation procedure proposed in this section is derived under general complex sampling designs, which also allows for the dependency between the subsample estimates. Therefore it is remarkable that no second order inclusion probabilities are required in the variance estimators. After detailed consideration, this is a consequence of 1) the assumption of constant treatment effects, 2) the application of a weighting scheme that satisfies the condition $\mathbf{a}^t\mathbf{x}_i = 1$ for all $i \in U$, 3) the assumption that measurement errors $\boldsymbol{\varepsilon}_i^{\alpha}$ are independent, 4) the specific randomization mechanism of CRD's and RBD's, and 5) variances of contrasts between subsample means are calculated. Consequently, the variance estimators have the structure as if the K subsamples have been drawn independently from each other where the sampling units are drawn with unequal inclusion probabilities with replacement. See Van den Brakel (2001) for a more detailed discussion about this result.

### 3.5 The Wald test

To test hypothesis (3.2), the design-based estimators for the subsample means and the covariance matrix of the contrasts between the subsample means give rise to the following design-based Wald statistic:

$$W = \hat{\bar{\mathbf{Y}}}_{\mathbf{GREG}}^{\alpha^t}\mathbf{C}^t(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t)^{-1}\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{GREG}}^{\alpha}.$$

Due to the diagonal structure of $\hat{\mathbf{D}}$, this Wald statistic can be simplified to

$$W = \sum_{k=1}^{K}\frac{\hat{\bar{Y}}_{k;greg}^{\alpha^2}}{\hat{d}_k} - \left(\sum_{k=1}^{K}\frac{1}{\hat{d}_k}\right)^{-1}\left(\sum_{k=1}^{K}\frac{\hat{\bar{Y}}_{k;greg}^{\alpha}}{\hat{d}_k}\right)^2. \tag{3.15}$$

If $s$ is drawn by means of simple random sampling without replacement and the experimental design is a CRD, then the conditions given by Lehmann (1975, appendix 8) can be applied to show that $\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{GREG}}^{\alpha} \to N(\mathbf{C}\boldsymbol{\beta}, \mathbf{CVC}^t)$. Under general complex sampling designs, in the analysis of survey data, it is generally assumed that a limit theorem holds so that $\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{GREG}}^{\alpha} \to N(\mathbf{C}\boldsymbol{\beta}, \mathbf{CVC}^t)$. Then it follows under the null hypothesis, that $W$ is asymptotically chi-squared distributed with $K$-$1$ degrees of freedom, to calculate $p$-values or critical regions for $W$.

### 3.6 Pooled variance estimators

In the case of an RBD, the variance estimation procedure might be improved by pooling the estimators for the population variances $S_{bk}^2$, which are implicitly defined in (3.11), within each block. The pooled estimator for the population variance in block $b$ is given by

$$\hat{S}_b = \frac{1}{n_{b+}-1}\sum_{k=1}^{K}\sum_{i\in s_{bk}}\left(\frac{n_{b+}(y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_{b+}}\sum_{k=1}^{K}\sum_{i'\in s_{bk}}\frac{n_{b+}(y_{i'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2, \tag{3.16}$$

or alternatively

$$\hat{S}_b = \frac{1}{n_{b+} - K} \sum_{k=1}^{K} \sum_{i \in s_{bk}} \left( \frac{n_{b+}(y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_{bk}} \sum_{i' \in s_{bk}} \frac{n_{b+}(y_{i'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2 . \qquad (3.17)$$

In some specific situations, the derived Wald statistic coincides with the *F*-statistic, known from the more traditional model-based analysis procedures. Consider an RBD embedded in a self-weighted sampling design where the allocation of the sampling units to the treatments is proportional over the blocks, i.e. $\pi_i = n_{++} / N$ and $n_{bk} / n_{b+} = n_{+k} / n_{++}$ for all *b*. Let $\bar{y}_k^{\alpha} = (1/n_{+k}) \sum_{i \in s_k} y_{ik}^{\alpha}$, $\bar{y}_b^{\alpha} = (1/n_{b+}) \sum_{i \in s_b} y_{ik}^{\alpha}$, and $\bar{y}^{\alpha} = (1/n_{++}) \sum_{i \in s} y_{ik}^{\alpha}$. If $n_{b+} \approx n_{b+} - 1$, then it follows for the extended Horvitz-Thompson estimator and the pooled variance estimator (3.16) that

$$\hat{d}_k = \frac{1}{n_{+k}} \frac{1}{n_{++}} \sum_{b=1}^{B} \sum_{k=1}^{K} \sum_{i \in s_{bk}} (y_{ik}^{\alpha} - \bar{y}_k^{\alpha} - \bar{y}_b^{\alpha} + \bar{y}^{\alpha})^2 \equiv \frac{\hat{d}}{n_{+k}} . \qquad (3.18)$$

Denote $\bar{y}_{bk}^{\alpha} = (1/n_{bk}) \sum_{i \in s_{bk}} y_{ik}^{\alpha}$. Under the pooled variance estimator (3.17) it follows that

$$\hat{d}_k = \frac{1}{n_{+k}} \frac{1}{n_{++}} \sum_{b=1}^{B} \sum_{k=1}^{K} \sum_{i \in s_{bk}} (y_{ik}^{\alpha} - \bar{y}_{bk}^{\alpha})^2 \equiv \frac{\hat{d}}{n_{+k}} . \qquad (3.19)$$

If these pooled variance estimators are substituted into the Wald statistic (3.15), then it follows that

$$W = \frac{1}{\hat{d}} \left( \sum_{k=1}^{K} n_{+k} (\bar{y}_k^{\alpha})^2 - n_{++} (\bar{y}^{\alpha^2}) \right).$$

It follows that $W/(K-1)$ with $\hat{d}$ defined in (3.18) corresponds to the *F*-statistic of an ANOVA for the two-way layout without interactions. If $\hat{d}$ defined in (3.19) is applied, then $W/(K-1)$ corresponds to the *F*-statistic of an ANOVA for the two-way layout with interactions. (See e.g. Scheffé, 1959, ch. 4.)

For a CRD, a pooled variance estimator follows as a special case from (3.13) or (3.14) by taking *b*= 1, $n_{b+} = n_+$, and $n_{bk} = n_k$. Under both estimators, $W/(K-1)$ corresponds to the *F*-statistic of an ANOVA for the one-way layout. (See e.g. Scheffé, 1959, ch. 3.) If $n_{++} \to \infty$, then the *F*-statistic of an ANOVA for the one- and two-way layout tend to $\chi^2_{[K-1]}/(K-1)$, so the Wald statistic and the *F*-statistic have the same limit distribution.

### 3.7 Advantages of RBD's

The main advantage of RBD's is the elimination of the variation between the blocks in the analysis of treatment effects. The variance reduction achieved with blocking by embedded experiments follows if the variance components (3.10) and (3.11) are compared. This suggests to use sampling structures like strata, PSU's, clusters, interviewers and the like as block variables in an RBD (Fienberg and Tanur, 1987, 1988). Consider for example an experiment embedded in a stratified sampling design. It follows from the variance expression (3.10) and (3.11) that the efficiency of a stratified sampling design is nullified in the case of a CRD, while it is preserved under an RBD where strata are used as block variables. Furthermore, an RBD with strata as block variables ensures that each stratum is sufficiently represented within each subsample. In the case of experiments embedded in two-stage sampling designs or cluster samples, it might be efficient to use PSU's or clusters as block variables, since in most practical situations, the sampling units from the same PSU or cluster have a higher degree of homogeneity compared with sampling units from different PSU's or clusters. If the data collection is conducted by means of CATI or CAPI, then interviewers are potential block variables. An RBD with interviewers as block variables eliminates 1) the variation in the observations due to fixed or random interviewer effects specified in the measurement error model (3.1) and 2) the variation of the target parameter of the sampling units between interviewers. The second component might be substantial in surveys where data are collected by means of CAPI, for then interviewers work in separated relatively small regions. Generally, response rates between interviewers might differ substantially. This implies that blocking on interviewers makes it better possible to preserve the orthogonality properties of the experimental design, e.g. a proportional allocation of sampling units to treatments over the blocks, which improves the power of the experiment (see Van den Brakel and Van Berkel, 2002). The main disadvantage of using interviewers as block variables is that such designs might complicate the data collection considerably, since each interviewer must collect data under each of the $K$ treatments.

## 4. Embedded experiments with different randomization levels

In this section we consider the analysis of embedded experiments where the randomization levels of the experimental units and the sampling units are different. Consider for example an experiment embedded in a two-stage sampling design, where PSU's are households and secondary sampling units (SSU's) are persons. According to the experimental designs considered in the preceding section, the SSU's are randomized over the treatments by means of a CRD or an RBD. In many situations, however, it might be infeasible to apply different treatments within the same PSU (household) from a practical point of view. In such situations PSU's are randomized over the treatments and consequently the experimental units do not coincide with the ultimate sampling units of the sampling design.

Consider a finite population $U$ that consists of $M$ PSU's. The $j$-th PSU consists of $N_j$ SSU's. Let $N = \sum_{j=1}^{M} N_j$ denote the population size. To test hypothesis (3.2), a two-stage sample $s$, drawn from $U$ is available. Let $\pi_j^I$ denote the first order inclusion probability of the $j$-th PSU in the first stage of the sampling design and $\pi_{i|j}^{II}$ the first order inclusion probability of the $i$-th SSU in the second stage given that the *j*-th PSU was selected in the first stage. In the case of a CRD, the sample of PSU's is randomized over the *K* treatments. Let $m_k$ denote the number of PSU's assigned to subsample $s_k$. Then $m_+ = \sum_{k=1}^{K} m_k$ denotes the total number of PSU's in *s*. The conditional probability that PSU *j* is assigned to treatment *k,* given the realization of the first stage equals $m_k / m_+$. In the case of an RBD, the PSU's are deterministically divided in B blocks $s_b$. The PSU's within each block are randomized over the *K* treatments. Interviewers or strata of the first stage design are potential block variables in this situation. Let $m_{bk}$ denote the number of PSU's assigned to treatment *k* in block *b*. Then $m_{b+} = \sum_{k=1}^{K} m_{bk}$ denote the number of PSU's in block *b*, $m_{+k} = \sum_{b=1}^{B} m_{bk}$ the number of PSU's in subsample $s_k$ and $m_{++} = \sum_{b=1}^{B} \sum_{k=1}^{K} m_{bk}$ the total number of PSU's in *s*. The conditional probability that PSU *j* is assigned to treatment *k,* given the realization of the first stage and that PSU $j \in s_b$, equals $m_{bk} / m_{b+}$. Considering each subsample as a realization of a two-phase sample (see section 3.3) it follows that the first order inclusion probabilities of the *j*-th PSU in the first stage of $s_k$ equals $\pi_j^{*I} = (m_k / m_+)\pi_j$ in the case of a CRD or $\pi_j^{*I} = (m_{bk} / m_{b+})\pi_j$ in the case of an RBD. The first order inclusion probability of the *i*-th SSU in subsample $s_k$ is given by $\pi_i^* = \pi_j^{*I} \pi_{i|j}^{II}$. Let $n_j$ denote the number of SSU's drawn from each of the *j* PSU's in the second stage and let $y_{ijk}^{\alpha}$ denote the observation obtained from the *i*-th SSU, drawn from the *j*-th PSU assigned to the *k*-th treatment. The Horvitz-Thompson estimator for $\overline{Y}_k$ based on the observations in $s_k$ is given by

$$\hat{\overline{Y}}_k^{\alpha} = \frac{1}{N} \sum_{j \in s_k} \sum_{i=1}^{n_j} \frac{y_{ijk}^{\alpha}}{\pi_j^{*I} \pi_{i|j}^{II}} \equiv \frac{1}{N} \sum_{j \in s_k} \frac{\hat{y}_{jk}^{\alpha}}{\pi_j^{*I}}, \tag{4.1}$$

where $\hat{y}_{jk}^{\alpha}$ denotes the Horvitz-Thompson estimator for the population total of the *j*-th PSU assigned to the *k*-th treatment. The generalized regression estimator for $\overline{Y}_k$ based on the observations in $s_k$ is given by expression (3.6), where the Horvitz-Thompson estimator for the regression coefficients $\hat{\mathbf{b}}_k^{\alpha}$ and the population means of the auxiliary information $\hat{\overline{\mathbf{X}}}_k$ are defined by (3.4) and (3.5), respectively using the first order inclusion probabilities derived in this section.

The covariance matrix of the contrasts between the first order Taylor linearization of the subsample means is again given by $\mathbf{CVC}^t = E_{\alpha} E_s \mathbf{CDC}^t$. In the case of an RBD, the diagonal elements of $\mathbf{D}$ are given by

$$d_k = \sum_{b=1}^{B} \frac{1}{m_{bk}} \frac{1}{m_{b+} - 1} \sum_{j \in s_b} \left( \frac{m_{b+}(\hat{y}_{jk}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t} \hat{\mathbf{x}}_j)}{N\pi_j^I} - \frac{1}{m_{b+}} \sum_{j' \in s_b} \frac{m_{b+}(\hat{y}_{j'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t} \hat{\mathbf{x}}_{j'})}{N\pi_{j'}^I} \right)^2, \tag{4.2}$$

where

$$\hat{y}_{jk}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t}\hat{\mathbf{x}}_j = \sum_{i=1}^{n_j} \frac{y_{ijk}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t}\mathbf{x}_{ij}}{\pi_{i|j}^{II}} . \tag{4.3}$$

Here $\mathbf{x}_{ij}$ denotes the vector with auxiliary information of the *i*-th SSU drawn from the *j*-th PSU. An approximately design-unbiased estimator for the covariance matrix of the contrasts between the subsample means is given by $\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t$ . In the case of an RBD, the diagonal elements of $\hat{\mathbf{D}}$ are given by

$$\hat{d}_k = \sum_{b=1}^{B} \frac{1}{m_{bk}} \frac{1}{m_{bk}-1} \sum_{j \in s_{bk}} \left( \frac{m_{b+}(\hat{y}_{jk}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t}\hat{\mathbf{x}}_j)}{N\pi_j^I} - \frac{1}{m_{bk}} \sum_{j' \in s_{bk}} \frac{m_{b+}(\hat{y}_{j'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t}\hat{\mathbf{x}}_{j'})}{N\pi_{j'}^I} \right)^2 , \tag{4.4}$$

where $\hat{y}_{jk}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t}\hat{\mathbf{x}}_j$ is given by (4.3). In the case of a CRD an expression for the diagonal elements of $\mathbf{D}$ and $\hat{\mathbf{D}}$ are given by (4.2), (4.3) and (4.4) by taking $B$=1, $m_{bk} = m_k$ , and $m_{b+} = m_+$ .

An expression for the extended Horvitz-Thompson estimator is defined by (3.14), using the first order inclusion probabilities derived in this section. An expression for the variance components $\hat{d}_k$ is obtained by (4.3) where

$$\hat{y}_{jk}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha t}\hat{\mathbf{x}}_j = \sum_{i=1}^{n_j} \frac{y_{ijk}^{\alpha} - \tilde{Y}_k^{\alpha}}{\pi_{i|j}^{II}} = \hat{y}_{jk}^{\alpha} - \tilde{Y}_k^{\alpha}\hat{N}_j .$$

To improve the variance estimation procedure, the pooled variance estimators proposed in section 3.6 can be applied in equivalent way. Hypothesis (3.2) can be tested with Wald statistic (3.15), using the subsample estimates and the variance components derived in this section.

An experiment embedded in a two-stage sampling design can be designed as an RBD with PSU's as block variables and SSU's are the experimental units, or as an experiment where PSU's are the experimental units. If the variation between the PSU's is large and the variation between the SSU's within PSU's small, then an RBD where PSU's are block variables and SSU's are experimental units will be preferable. If the variation between PSU's is small and the variation between SSU's within PSU's is large, then under both experimental designs there will be a large variance component arise from the variation between SSU's within PSU's. In this situation an RBD with PSU's as block variables and SSU's as experimental units has the advantage that more degrees of freedom are available in the variance estimation procedure. If, however, the PSU's are completely observed, i.e. the sampling design is a cluster sample, then it might be efficient to apply an experimental design where PSU's are experimental units. Using PSU's as experimental units in this situation avoids the introduction of a large variance component from the variation between SSU's within PSU's, since PSU's are completely observed under one of the $K$ treatments. As emphasized in the introduction of this section, an experimental design where PSU's are the experimental units will generally be considered for practical reasons if the application of different treatments within a PSU is not feasible.

## 5. Testing hypotheses about ratios of population totals

The analysis procedure developed in the preceding sections can be applied to test hypotheses about estimates of population means or totals. In many sample surveys, however, target parameters are defined as the ratio of two population totals. Therefore, the methodology from the preceding sections will be extended to the analysis of ratios. Let $R_k = Y_k / Z_k$ denote the ratio of two population totals observed under treatment *k*. Then $\mathbf{R} = (R_1, \ldots, R_K)^t$ denotes the *K* dimensional vector with ratios observed under the different treatments of the experiment. The purpose of the experiment is to test the following hypothesis:

$$\begin{aligned} &H_0 : \mathbf{CR} = \mathbf{0} \\ &H_1 : \mathbf{CR} \neq \mathbf{0} \end{aligned} . \tag{5.1}$$

Let $y_{ikl}^{\alpha}$ and $z_{ikl}^{\alpha}$ denote the observations obtained from experimental unit *i*, assigned to interviewer *l* and treatment *k* on the $\alpha$-th occasion. Both variables are assumed to be a realization of measurement error model (3.1). The generalized regression estimator for $R_k$, is given by $\hat{R}_{k;greg}^{\alpha} = \hat{Y}_{k;greg}^{\alpha} / \hat{Z}_{k;greg}^{\alpha}$, where $\hat{Y}_{k;greg}^{\alpha}$ and $\hat{Z}_{k;greg}^{\alpha}$ are the generalized regression estimators of the population totals $Y_k$ and $Z_k$ based on the observations obtained in subsample $s_k$. Finally $\hat{\mathbf{R}}_{\mathbf{GREG}}^{\alpha} = (\hat{R}_{1;greg}^{\alpha}, \ldots, \hat{R}_{K;greg}^{\alpha})^t$ denotes the generalized regression estimator for **R**.

Let $\mathbf{V}$ denote the covariance matrix of $\hat{\mathbf{R}}_{\mathbf{GREG}}^{\alpha}$. To derive an expression for the covariance matrix of the *K*-1 contrasts between $\hat{\mathbf{R}}_{\mathbf{GREG}}^{\alpha}$, a linear approximation of $\hat{R}_{k;greg}^{\alpha}$ must be obtained first. The first order Taylor linearization of $\hat{R}_{k;greg}^{\alpha}$ about the point $R_k$ is given by

$$\hat{R}_{k;greg}^{\alpha} \approx R_k + \frac{1}{Z_k} (\hat{Y}_{k;greg}^{\alpha} - R_k \hat{Z}_{k;greg}^{\alpha}) . \tag{5.2}$$

Subsequently $\hat{Y}_{k;greg}^{\alpha}$ and $\hat{Z}_{k;greg}^{\alpha}$ in (5.2) are linearized with a first order Taylor approximation about $(Y_k, \mathbf{b}_k, \mathbf{X})$ and $(Z_k, \mathbf{d}_k, \mathbf{X})$ respectively, where $\mathbf{d}_k$ denotes the *H*-dimensional vector with the regression coefficients of the regression function of $z_{ikl}$ on $\mathbf{x}_i$. Then it follows that

$$\begin{aligned} \hat{R}_{k;greg}^{\alpha} &\approx R_k + \frac{1}{Z_k} (\hat{Y}_k^{\alpha} + \mathbf{b}_k^t (\mathbf{X} - \hat{\mathbf{X}}_k) - R_k (\hat{Z}_k^{\alpha} + \mathbf{d}_k^t (\mathbf{X} - \hat{\mathbf{X}}_k))) \\ &\equiv R_k + \hat{E}_k^{\alpha} + \frac{1}{Z_k} (\mathbf{b}_k^t \mathbf{X} - R_k \mathbf{d}_k^t \mathbf{X}) \end{aligned} ,$$

where

$$\hat{E}_k^{\alpha} = \frac{1}{Z_k} \sum_{i \in s_k} \frac{y_{ik}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_i - R_k (z_{ik}^{\alpha} - \mathbf{d}_k^t \mathbf{x}_i)}{\pi_i^*} .$$

The covariance matrix of the contrasts of $\hat{\mathbf{R}}_{\mathbf{GREG}}^{\alpha}$ is approximated with the covariance matrix of the first order Taylor linearization of the contrasts of $\hat{\mathbf{R}}_{\mathbf{GREG}}^{\alpha}$, which is obtained by that of $\hat{\mathbf{E}}^{\alpha} = (\hat{E}_1^{\alpha}, \ldots, \hat{E}_K^{\alpha})^t$. The same analysis used for the derivation of the covariance matrix of the contrasts between the subsample means in section 3, can be applied to show that $\mathbf{CVC}^t = E_\alpha E_s \mathbf{CDC}^t$. In the case of an RBD, $\mathbf{D}$ is a diagonal matrix with elements

$$d_k = \frac{1}{Z_k^2} \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_{b+} - 1} \sum_{i \in s_b} \left( \frac{n_{b+} e_{ik}^\alpha}{\pi_i} - \frac{1}{n_{b+}} \sum_{i' \in s_b} \frac{n_{b+} e_{i'k}^\alpha}{\pi_{i'}} \right)^2 , \tag{5.3}$$

where

$$e_{ik}^\alpha = y_{ik}^\alpha - \mathbf{b}_k^t \mathbf{x}_i - R_k (z_{ik}^\alpha - \mathbf{d}_k^t \mathbf{x}_i) .$$

An estimator for $\mathbf{CVC}^t$ is given by $\mathbf{C\hat{D}C}^t$ where the diagonal elements of $\hat{\mathbf{D}}$ are given by

$$\hat{d}_k = \frac{1}{\hat{Z}_{k;greg}^{\alpha^2}} \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_{bk} - 1} \sum_{i \in s_{bk}} \left( \frac{n_{b+} \hat{e}_{ik}^\alpha}{\pi_i} - \frac{1}{n_{bk}} \sum_{i' \in s_{bk}} \frac{n_{b+} \hat{e}_{i'k}^\alpha}{\pi_{i'}} \right)^2 \tag{5.4}$$

where

$$\hat{e}_{ik}^\alpha = y_{ik}^\alpha - \hat{\mathbf{b}}_k^{\alpha^t} \mathbf{x}_i - \hat{R}_{k;greg}^\alpha (z_{ik}^\alpha - \hat{\mathbf{d}}_k^{\alpha^t} \mathbf{x}_i) .$$

For experiments embedded in two- or multi-stage sampling designs, where the PSU's correspond to the experimental units (section 4), an expression for the diagonal elements of $\mathbf{D}$ under an RBD is given by

$$d_k = \frac{1}{Z_k^2} \sum_{b=1}^{B} \frac{1}{m_{bk}} \frac{1}{m_{b+} - 1} \sum_{j \in s_b} \left( \frac{m_{b+} \hat{e}_{jk}^\alpha}{\pi_j^I} - \frac{1}{m_{b+}} \sum_{j' \in s_b} \frac{m_{b+} \hat{e}_{j'k}^\alpha}{\pi_{j'}^I} \right)^2 , \tag{5.5}$$

where

$$\hat{e}_{jk}^\alpha = \sum_{i=1}^{n_j} \frac{y_{ijk}^\alpha - \mathbf{b}_k^t \mathbf{x}_{ij} - R_k (z_{ijk}^\alpha - \mathbf{d}_k^t \mathbf{x}_{ij})}{\pi_{i|j}^{II}} .$$

An expression for the diagonal elements of $\hat{\mathbf{D}}$ under an RBD is given by

$$\hat{d}_k = \frac{1}{\hat{Z}_{k;greg}^{\alpha^2}} \sum_{b=1}^{B} \frac{1}{m_{bk}} \frac{1}{m_{bk} - 1} \sum_{j \in s_{bk}} \left( \frac{m_{b+} \hat{\hat{e}}_{jk}^\alpha}{\pi_j^I} - \frac{1}{m_{bk}} \sum_{i' \in s_{bk}} \frac{m_{b+} \hat{\hat{e}}_{j'k}^\alpha}{\pi_{j'}^I} \right)^2 , \tag{5.6}$$

where

$$\hat{\hat{e}}_{jk}^\alpha = \sum_{i=1}^{n_j} \frac{y_{ijk}^\alpha - \hat{\mathbf{b}}_k^{\alpha^t} \mathbf{x}_{ij} - \hat{R}_{k;greg}^\alpha (z_{ijk}^\alpha - \hat{\mathbf{d}}_k^{\alpha^t} \mathbf{x}_{ij})}{\pi_{i|j}^{II}} .$$

An expression for the diagonal elements $d_k$ and $\hat{d}_k$ under a CRD follows as a special case from (5.3) and (5.4) or (5.5) and (5.6), respectively by taking $B = 1$, $n_{bk} = n_k$, and $n_{b+} = n_+$, or $m_{bk} = m_k$, and $m_{b+} = m_+$. Expressions for the extended Horvitz-Thompson estimator follow directly as a special case from the generalized regression estimator with the common mean model as a weighting scheme. The pooled variance estimators proposed in section 3.6 can be applied in an equivalent way for the variance estimators for contrasts between ratios. Finally hypothesis (5.1) can be tested with the Wald statistic defined in (3.15) where $\hat{Y}^{\alpha}_{k;greg}$ is replaced by $\hat{R}^{\alpha}_{k;greg}$ and the appropriate expressions for the variance components $\hat{d}_k$ are applied.

## 6. Discussion

To test hypotheses about estimates of finite population parameters of a sample survey, observed under different survey approaches or treatments, a design-based Wald statistic for the analysis of CRD's or RBD's embedded in generally complex sampling designs is derived. Parameter and variance estimators are based on the Horvitz-Thompson estimator or the generalized regression estimator, which account for the randomization mechanism of the sampling design, the experimental design and the estimation procedure of the regular survey in the analysis of the experiment. The application of a design-based analysis procedure in combination with the random selection of experimental units from a finite population by means of random sampling, enables us to generalize the results observed in the specific sample of the experiment to the entire survey population.

At first sight, an analysis procedure that accounts for the sampling design might be considered as an alternative to test hypotheses about the regression coefficients of a linear model or linear combinations of the subsample estimates, which reflect the treatment effects of the experiment. The superimposition of the experimental design on the sampling design, however, determines which specific features of the sampling design are nullified or preserved. For example, the effect of stratified sampling or two-stage sampling on the variance of the treatment effects is nullified under a CRD. Therefore, the application of such an alternative analysis procedure can still lead to misleading results, since the superimposition of the experimental design on the sampling design is ignored by the variance of the treatment effects. Furthermore the interpretation of the results is more complicated if the treatment effects are specified as the regression coefficients in a linear model, which are not directly linked with the finite population parameters as they are defined in the sample survey.

Since the subsamples are drawn without replacement from a finite population according to a general complex sampling design, a rather complicated expression for the covariance matrix with nonzero off-diagonal entries is expected. Although we allow for general complex sampling designs, as well as the dependency between the subsample estimates, the derived estimator for this covariance matrix has the structure as if the subsamples are drawn independently from each other where the sample units are selected with unequal probabilities with replacement. No second order inclusion probabilities are required, which simplifies the analysis considerably. As a result a Wald statistic, derived from a design-based perspective under general complex sampling designs, is obtained that still has the attractive relative simple structure of standard model-based analysis procedures.

The analysis procedure proposed in this paper is currently implemented in Statistics Netherlands' software package Bascula in order to facilitate the application of these methods. The calculations, however, can also be performed indirectly with software packages that facilitate survey estimation procedures like Stata or Sudaan. Therefore it should be recognized that the variance components in (3.13) and (4.4) can be calculated with the analytical variance estimator for a multi-stage sampling design where the first stage is drawn with replacement and subsequently taking the experimental units and the block variables as the PSU's and the stratum variables, respectively.

## References


Bethlehem, J.G. (1988), "Reduction of Nonresponse Bias through Regression Estimation", *Journal of Official Statistics*, 4, pp. 251-260.

Biemer, P.B. and S.L. Stokes (1991), "Approaches to the Modeling of Measurement Errors", in P.B. Biemer et al. (eds.) *Measurement Errors in Surveys,* New York: Wiley, pp. 487-516.

Cochran, W. (1977), "*Sampling Techniques*", New York: Wiley.

Fienberg, S.E. and J.M. Tanur (1987), "Experimental and Sampling Structures: Parallels Diverging and Meeting", *International Statistical Review*, 55, pp. 75-96.

Fienberg, S.E. and J.M. Tanur (1988), "From the inside out and the outside in: Combining experimental and sampling structures", *The Canadian Journal of Statistics*, 16, pp. 135-151.

Lehmann, E.L. (1975), "*Nonparametrics: Statistical Methods Based on Ranks*", New York: McGraw-hill.

Montgomery, D.C. (1997), "*Design and Analysis of Experiments*", New York: Wiley.

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

Scheffé, H. (1959), "*The Analysis of Variance*", New York: Wiley.

Van den Brakel, J.A. (2001), "*Design and Analysis of Experiments Embedded in Complex Sample Surveys*" Ph.D. Thesis, Erasmus University Rotterdam.

Van den Brakel, J.A. (2002), "Effects of Interviewers' Workload on the Dutch Labour Force Survey", *Research in Official Statistics*, 5, pp. 87-106.

Van den Brakel, J.A. and D. Binder (2000), "Variance Estimation for Experiments Embedded in Complex Sampling Schemes", *Proceedings of the Joint*

*Statistical Meetings, Survey Research Methods Section, American Statistical Association,* pp. 805-810.

Van den Brakel, J.A. and R.H. Renssen (1998), "Design and Analysis of Experiments Embedded in Sample Surveys", *Journal of Official Statistics*, 14, pp. 277-295.

Van den Brakel, J.A. and C.A.M. van Berkel (2002), "A Design-Based Analysis Procedure for Two-treatment Experiments Embedded in Sample Surveys", *Journal of Official Statistics*, 18, pp. 217-231.