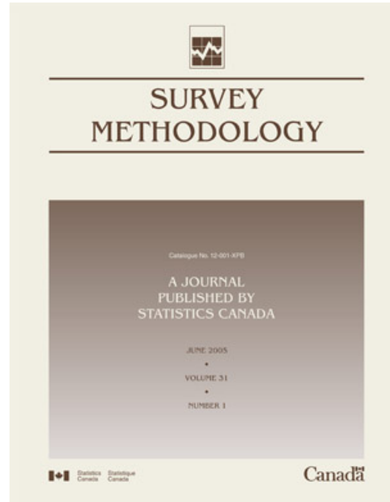
Catalogue no. 12-001-XIE

# Survey Methodology

June 2006




Statistics Canada Statistique Canada

Canada

## How to obtain more information

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

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

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

## Information to access the product

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

## Standards of service to the public

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

Statistics Canada

Business Survey Methods Division

# Survey Methodology

June 2006

Published by authority of the Minister responsible for Statistics Canada



July 2006

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

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (n$^{o}$ 12-001-XIF au catalogue).

---

**Note of appreciation**

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

Vol. 32, No. 1, pp. 115-120
Statistics Canada, Catalogue No. 12-001

# Design Effects for Multiple Design Samples

**Siegfried Gabler, Sabine Häder and Peter Lynn** [1]

Abstract

In some situations the sample design of a survey is rather complex, consisting of fundamentally different designs in different domains. The design effect for estimates based upon the total sample is a weighted sum of the domain-specific design effects. We derive these weights under an appropriate model and illustrate their use with data from the European Social Survey (ESS).

Key Words: Stratification; Clustering; Variance component model; Intraclass correlation coefficient; Selection probabilities.

## 1. Introduction

In survey research complex sample designs are often applied. These designs have features such as stratification, clustering and/or unequal inclusion probabilities, that lead to "design effects". The design effect is a measure that shows the effect of the design on the variance of an estimate. Design-based it is defined as follows (see Lohr 1999, page 239):

$$deff\,(plan, statistic) = \frac{\mathrm{V}(\text{estimate from sampling plan})}{\mathrm{V}\begin{pmatrix}\text{estimate from an srs with same number} \\ \text{of observation units}\end{pmatrix}}$$

where srs indicates a simple random sample. The use of clustering and/or unequal inclusion probabilities typically leads to design effects greater than 1.0; in other words the variance of an estimate is increased compared to the variance of the estimate from a simple random sample with the same number of observations. The consideration of design effects is very important when deciding upon the sample size of a survey in advance. For example, if a comparative survey with different countries is planned it is very useful to have estimates of the design effects for the different countries. Then it is possible to choose the net sample sizes in a way that the precision of the estimates will be approximately equal. For this, for a certain degree of precision the sample size that would be needed under srs (effective sample size) has to be multiplied by the predicted design effect.

The European Social Survey (ESS, see www.europeansocialsurvey.com) is a survey program where design effects are taken into consideration for calculating net sample sizes − aiming at the same effective sample size for each country ($n_{\text{eff}} = 1{,}500$). 22 countries participated in the first round of the ESS, only three of them with unclustered, equal probability designs (srs): Denmark, Finland and Sweden. For all other countries there was the need to predict the design effect in advance of the study. For this, a model based approach (see Gabler, Häder and Lahiri 1999) can be used which distinguishes between a design effect due to unequal inclusion probabilities (term 1) and a design effect due to clustering (term 2):

$$deff = m \frac{\sum_{i=1}^{I} m_i w_i^2}{\left(\sum_{i=1}^{I} m_i w_i\right)^2} \times [1 + (b^* - 1)\rho] = deff_p \times deff_c \qquad (1)$$

where $m_i$ are respondents in the $i^{\text{th}}$ selection probability class, each receiving a weight of $w_i$, $\rho$ is the intracluster correlation coefficient and

$$b^* = \frac{\sum_{c=1}^{C} \left(\sum_{j=1}^{b_c} w_{cj}\right)^2}{\sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj}^2}$$

where $b_c$ is the number of observations in cluster $c (c = 1, \ldots, C)$ and $w_{cj}$ is the design weight for sample element $j$ in cluster $c$. (This is of course a simplification that assumes no association between $y$ and $w_i$ or between $w_i$ and $b^*$ and ignores any effects of stratification, that will tend to be beneficial and modest. See Lynn, Gabler, Häder and Laaksonen (2007, forthcoming) and Park and Lee (2004) for discussion of the sensitivity of *deff* predictions to these assumptions; see Lynn and Gabler (2005) for discussion of alternative ways to predict $deff_c$).

In some countries the applied designs were even more complicated, consisting of fundamentally different designs in each of two independent domains. In the UK, *e.g.*, the design was a mixture of a clustered design with unequal inclusion probabilities (in Great Britain) and an unclustered

1. Siegfried Gabler and Sabine Häder, Zentrum für Umfragen, Methoden und Analysen (ZUMA), Postfach 12 21 55, 68072 Mannheim, Germany. E-mail: gabler@zuma-mannheim.de; Peter Lynn, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom. E-mail: plynn@essex.ac.uk.

sample (in Northern Ireland). In Poland, simple random samples were selected in one domain (cities and large towns), while a two-stage clustered design was applied in the second domain (all other areas). In Germany, a clustered equal-probability sample was selected in each domain (West Germany including West Berlin; East Germany), but the sampling fractions differed between the domains.

The question arose how to predict design effects for these dual design samples. As we show below, it is not simply a convex combination of the design effects for the different domains – apart from in some special cases. A general solution for multiple design samples will be presented in section 2, with illustrations of the application of this solution to prediction of design effects prior to field work (section 3) and to estimation of design effects post-field work (section 4). Section 5 concludes with discussion.

## 2. Design Effects for Multiple Design Samples

Let $\{C_1, \ldots, C_K\}$ be a partition of the clusters into $K$ domains. Then $C\bar{b} = \sum_{c=1}^{C} b_c = \sum_{k=1}^{K} \sum_{c \in C_k} b_c = \sum_{k=1}^{K} m_k = m$, where $m_k = \sum_{c \in C_k} b_c$ is the number of observations in the $k^{\text{th}}$ domain of clusters. Let $y_{cj}$ be the observation for sample element $j$ in cluster $c (c = 1, \ldots, C; \; j = 1, \ldots, b_c)$. The usual design-based estimator for the population mean is

$$\bar{y}_w = \frac{\sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj} y_{cj}}{\sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj}} = \sum_{k=1}^{K} \frac{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj}}{\sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj}} \bar{y}_w^{(k)}$$

where

$$\bar{y}_w^{(k)} = \frac{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj} \, y_{cj}}{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj}}.$$

We assume the following model M1:

$$\left.\begin{aligned} E(y_{cj}) &= \mu \\ \mathrm{Var}(y_{cj}) &= \sigma^2 \end{aligned}\right\} \quad \text{for} \quad c = 1, \ldots, C; j = 1, \ldots, b_c$$
$$\mathrm{Cov}(y_{cj}, y_{c'j'}) = \begin{cases} \rho_k \sigma^2 & \text{if } c = c' \in C_k; j \neq j' \\ 0 & \text{otherwise} \end{cases} \quad k = 1, \ldots, K. \tag{2}$$

Model M1 is appropriate to account for the cluster effect with different kinds of clusters and generalises an earlier approach (see, *e.g.*, Gabler *et al.* 1999). More general models can be found in Rao and Kleffe (1988, page 62). We define the (model) design effect as $deff = \mathrm{Var}_{M1}(\bar{y}_w) / \mathrm{Var}_{M2}(\bar{y})$, where $\mathrm{Var}_{M1}(\bar{y}_w)$ is the variance of $\bar{y}_w$ under model M1 and $\mathrm{Var}_{M2}(\bar{y})$ is the variance of the overall sample mean $\bar{y}$, defined as $\sum_{c=1}^{C} \sum_{j=1}^{b_c} y_{cj} / m$, computed under the following model M2:

$$\left.\begin{aligned} E(y_{cj}) &= \mu \\ \mathrm{Var}(y_{cj}) &= \sigma^2 \end{aligned}\right\} \quad \text{for} \quad c = 1, \ldots, C; j = 1, \ldots, b_c$$
$$\mathrm{Cov}(y_{cj}, y_{c'j'}) = 0 \quad \text{for all} \quad (c, j) \neq (c', j'). \tag{3}$$

Note that model M2 is appropriate under simple random sampling and provides the usual expression, $\mathrm{Var}_{M2}(\bar{y}) = \sigma^2 / m$.

Quite analogous to Gabler *et al.* (1999) we note that

$$\mathrm{Var}_{M1}\left( \sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj} \, y_{cj} \right) = \sigma^2 \sum_{k=1}^{K} \sum_{c \in C_k} \left\{ \sum_{j=1}^{b_c} w_{cj}^2 + \rho_k \sum_{j \neq j'}^{b_c} w_{cj} w_{cj'} \right\}. \tag{4}$$

Thus

$$deff = \sum_{k=1}^{K} \left( \frac{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj}}{\sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj}} \right)^2 \frac{m}{m_k} deff_k \tag{5}$$

where

$$deff_k = m_k \frac{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj}^2}{\left( \sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj} \right)^2} \times [1 + (b_k^* - 1)\rho_k] = deff_{pk} \times deff_{ck},$$

and

$$b_k^* = \frac{\sum_{c \in C_k} \left( \sum_{j=1}^{b_c} w_{cj} \right)^2}{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj}^2}.$$

It can be seen that *deff* is not a convex combination of the specific $\{deff_k\}$ except in some special cases. We consider here four realistic scenarios, each representing a simplification of the general case. Only in two of these scenarios (scenarios 1 and 4) does the combination become convex:

*Scenario 1*: *Equal weights for all units*

If $w_{cj} = 1$ for all $c, j$, then expression (5) simplifies to:

$$deff = \sum_{k=1}^{K} \frac{m_k}{m} deff_k. \tag{6}$$

*Scenario 2*: *Equal weights within each domain*

If $w_{cj} = w_k$ for all $c \in C_k, j$, then expression (5) becomes:

$$deff = \sum_{k=1}^{K} \left( \frac{m_k w_k}{\sum_{k=1}^{K} m_k w_k} \right)^2 \frac{m}{m_k} deff_k. \tag{7}$$

*Scenario 3*: *Weighted sample size proportional to domain population size*

If

$$\frac{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj}}{\sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj}} = \frac{N_k}{N},$$

where $N_k$ is population size in domain $k; N = \sum_{k=1}^{K} N_k$, then expression (5) becomes:

$$deff = \sum_{k=1}^{K} \left( \frac{N_k}{N} \right)^2 \frac{m}{m_k} deff_k. \tag{8}$$

*Scenario 4*: *Unweighted sample size proportional to domain population size*

If

$$\frac{m}{m_k} = \frac{N}{N_k},$$

then expression (8) becomes:

$$deff = \sum_{k=1}^{K} \frac{N_k}{N} deff_k. \tag{9}$$

## 3. Application to Prediction of *Deff*

In round 1 of the ESS, the sample design was a combination of two different sample designs for 5 out of 22 countries: United Kingdom, Poland, Belgium, Norway and Germany. We can apply the general formula (5) for design effects for multiple design samples to each of these cases, where $K = 2$. In some cases, we can equivalently use one of the simplified expressions (6) to (9). Here we illustrate how the formulae would be used in the prediction of design effects prior to fieldwork, for the purpose of establishing the required net (respondent) sample size to achieve a prescribed precision of estimation. In each case, the approach is to predict $\{deff_k\}$ using (1) for each $k$ and then use (5) to predict *deff*. To predict $\{deff_k\}$, the observed values of $\{w_{cj}\}$ from the ESS round 1 respondent sample are used to estimate, $b^*, m_i$ and $w_i$. In other words, these could be thought of as predictions for a future survey using the same design (*e.g.*, a future round of ESS). For illustration, we assume $\rho_k = 0.02 \,\forall k$ with a clustered design and $\rho_k = 0.00 \,\forall k$ with an unclustered design (0.02 is in fact the default value that was used for predicted design effects for clustered samples on the ESS in cases where estimates from previous surveys were not available). Our focus here is on the application of (5). For a more detailed description of the sample designs see Häder, Gabler, Laaksonen and Lynn (2003). We use three of the ESS countries – Poland, UK and Germany – as illustrations as these designs differ between the domains in different ways. The designs of Norway and Belgium were similar to that of Poland, with equal probabilities for all units but one domain clustered and one unclustered.

### 3.1 Poland

In Poland, the first domain covered the population living in towns of 100,000 inhabitants or more. Within this domain, a srs of persons was selected from the population register (PESEL data base) in each region, with slight variation between regions in the sampling fraction, reflecting anticipated differences in response rate. There were 42 towns in this domain and they accounted for about 31% of the target population.

The second domain corresponded to the rest of the population – people living in towns of 99,999 inhabitants or fewer and people living in rural areas. This part of the sample was stratified and clustered (158 clusters). The sampling of this second part was based on a two-stage design: PSUs were selected with probability proportional to size. The definition of a PSU was different for urban *vs.* rural areas. For urban areas, a PSU was equivalent to a town, whereas for rural areas, it was equivalent to a village. In the second stage, a cluster of 12 respondents was selected in each PSU by srs.

In the first domain, $\rho_1 = 0$ and $deff_{c1} = 1$. The modest variation in selection probabilities leads to $deff_{p1} = 1.005$ and, therefore, $deff_1 = deff_{c1} \cdot deff_{p1} = 1.005$. In the second domain, the design effect due to clustering is anticipated to be $deff_{c2} = 1.18$ (based on a prediction of $b^* = 10.07$) and $deff_{p2} = 1.01$ which results in $deff_2 = deff_{c2} \cdot deff_{p2} = 1.19$. Substituting these values of $deff_k$ in (5) leads to a prediction of $deff = 1.17$.

The design for Poland differs only slightly from scenario 2 and it can be seen that in this case the simpler expression, (7), provides a reasonable prediction if we approximate the weights as follows. Domain 1 contains 37.3% of the gross sample and 31% of the target population. Thus

$$w_1 = \frac{N_1 / N}{n_1 / n} = \frac{0.310}{0.373} = 0.831$$

and

$$w_2 = \frac{N_2 / N}{n_2 / n} = \frac{0.690}{0.627} = 1.100,$$

respectively, where $n_k$ is selected sample size in domain $k; \sum_{k=1}^{K} n_k$.

Now, we can apply expression (7) to find the predicted design effect for estimates for Poland: $deff = (0.194 \cdot 1.005) + (0.821 \cdot 1.19) = 1.17.$

### 3.2. United Kingdom

In the UK, the ESS sample design differed between Great Britain (England, Wales, Scotland) and Northern Ireland. In Great Britain a stratified three-stage design with unequal probabilities was applied. At the first stage 162 small areas known as "postcode sectors" were selected systematically with probability proportional to the number of addresses in the sector, after implicit stratification by region and population density. At stage 2, 24 addresses were selected in each sector, leading to an equal-probability sample of addresses. At the third stage, one person aged 15+ was selected at the selected address using a Kish grid.

For Northern Ireland a simple random sample of 125 addresses was drawn from the Valuation and Land Agency's list of domestic properties. One person aged 15+ was selected at the selected address using a Kish grid. Thus, the UK sample is clustered in one domain but not in the other. In both domains, there are unequal selection probabilities.

In Great Britain we predicted $deff_{c1} = 1.20$ (based on a prediction of $b^* = 11.11$) and $deff_{p1} = 1.22$, so $deff_1 =$ 1.46. In Northern Ireland we have predictions of $deff_{c2} = 1$ (by definition) and $deff_{p2} = 1.27$, so $deff_2 = 1.27$. From expression (5), $deff = 0.978 \cdot 1.46 + 0.023 \cdot 1.27 = 1.460$. It should also be noted that the selected sample sizes in the two domains were chosen to result in net sample sizes that would be approximately in proportion to the population sizes. In other words, the simplification of scenario 4 approximately holds. If we use expression (9), we get $deff = N_1/N \; deff_1 + N_2/N \; deff_2 = 0.97 \cdot 1.46 + 0.03 \cdot 1.27 = 1.457$, demonstrating that this provides a reasonable approximation to (5) in this case.

### 3.3. Germany

In Germany independent samples were selected in two domains, West Germany incl. West Berlin, and East Germany incl. East Berlin. In both domains, the sample was clustered and equal-probability, but a higher sampling fraction was used in East Germany.

At the first stage 100 communities (clusters) for West Germany, and 50 for East Germany were selected with probability proportional to the population size of the community (aged 15 years or older). The number of communities selected from each stratum was determined by a controlled rounding procedure. The number of sample points was 108 in the West, and 55 in the East (some larger communities have more than one sample point). At the second stage in each sample point there was drawn an equal number of individuals by a systematic random selection process. This was done using the local registers of residents' registration offices.

Since the sampling design is self-weighting for both East and West Germany, but with disproportional allocation, scenario 2 applies and we can use expression (7), where

$$w_1 = w_{\text{EAST}} = \frac{N_{\text{EAST}}}{N} \frac{n}{n_{\text{EAST}}} = 0.567$$

and

$$w_2 = w_{\text{WEST}} = \frac{N_{\text{WEST}}}{N} \frac{n}{n_{\text{WEST}}} = 1.257.$$

(we note that common practice on some surveys is to scale the weights so that they sum to population sizes. This would make no difference to the application here as expression (5) involves only ratios of sums of weights).

The design effect due to clustering for each domain was predicted as $deff_{c1} = 1.39$ and $deff_{c2} = 1.35$, respectively (*via* predictions of $b^* = 20.56$ and 18.65 respectively), so from (7) we have

$$deff = 0.120 \cdot 1.39 + 0.991 \cdot 1.35 = 1.51.$$

It should be noted that in this case any convex combination of the domain-specific design effects will lead to a prediction of $deff$ between 1.35 and 1.39. For example, (6) would give $deff = 1.36$. This fails to take into account the differences in selection probabilities *between* the domains. With this particular design – where the *only* difference in design between domains is the difference in selection probabilities – $deff$ might alternatively be predicted by taking the convex combination and multiplying it by the prediction of $deff_p$ from the first term in expression (1), *viz.* $deff = 1.36 \cdot 1.09 = 1.49$. But this method is equivalent only in the special case where $\{deff_k\}$ are equal – and approximately equivalent in this case, where the variation is small.

## 4. Application to Estimation of *Deff*

Here we illustrate the use of expression (5) in the estimation of design effects post-fieldwork. We present estimates for 5 demographic/behavioural variables and a set of 24 attitude measures from round 1 of the European Social Survey, for the same three countries as in section 3. For comparison, we present also the estimates that would be obtained using the simpler expressions (6), (8) and (9). It can be seen that the estimates of $deff$ differ greatly between variables. This is to be expected, reflecting variation in the association of $y$ with clusters and with selection probabilities. But here we are more interested in differences between estimation methods for the same variable.

For Germany, we see that estimators (6) and (9), which ignore variation in weights and in sampling rates between the two domains respectively, under-estimates *deff* for all variables. Estimator (8), which assumes only equal response rates in each domain, produces estimates very similar to (5). For Poland, all three simplified estimators under-estimate *deff*, though (6) perhaps performs marginally better than the other two. For UK, we observe the remarkable result that all four estimators produce almost identical estimates for every variable. The assumption in (9) (and therefore also that in (8)) holds for UK and while weights are by no means equal, the distribution of weights is very similar in each domain. It can be noted that (6) holds under a weaker assumption that

$$\frac{\sum_{c \in C_k} \sum_{j=1}^{b_c} w_{cj}}{\sum_{c=1}^{C} \sum_{j=1}^{b_c} w_{cj}} = \frac{m_k}{m},$$

*i.e.*, that the share of the weights in each stratum equals the share of sample units. It is striking that these relationships between the estimators are consistent across all the variables considered.

## 5. Discussion and Conclusion

Expression (5) provides an appropriate means of combining design effects for domains with fundamentally different designs. It can be applied in estimation by estimating *deff*s in the usual way for each domain and then combining them using knowledge of the weight and domain membership of sample units. Use of (5) in the prediction of *deff*s before a survey is carried out is only slightly more demanding, requiring prediction of the share of the weights in the responding sample in each domain in addition to a method of predicting design-specific *deff*s.

**Table 1**
Estimates of *Deff* for Means Under 4 Estimators for 3 Countries

| | DE | | | | GB | | | | PL | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimator: | (5) | (6) | (8) | (9) | (5) | (6) | (8) | (9) | (5) | (6) | (8) | (9) |
| Demographic/behavioural | | | | | | | | | | | | |
| Persons in household | 1.87 | 1.85 | 1.87 | 1.74 | 1.66 | 1.66 | 1.66 | 1.66 | 1.51 | 1.43 | 1.41 | 1.42 |
| Years of education | 3.25 | 2.80 | 3.25 | 2.88 | 2.81 | 2.79 | 2.80 | 2.79 | 1.77 | 1.66 | 1.63 | 1.64 |
| Net household income | 2.46 | 2.15 | 2.46 | 2.19 | 2.82 | 2.80 | 2.80 | 2.80 | 2.16 | 2.00 | 1.95 | 1.98 |
| Time watching TV | 2.08 | 1.86 | 2.08 | 1.87 | 2.04 | 2.03 | 2.03 | 2.03 | 1.31 | 1.26 | 1.25 | 1.25 |
| Time reading newspaper | 1.79 | 1.62 | 1.79 | 1.61 | 1.35 | 1.35 | 1.35 | 1.35 | 1.73 | 1.63 | 1.60 | 1.61 |
| Attitude measures | | | | | | | | | | | | |
| Discriminated by race | 1.16 | 1.03 | 1.16 | 1.04 | 1.92 | 1.92 | 1.92 | 1.92 | 1.02 | 1.01 | 1.01 | 1.01 |
| Discriminated by religion | 1.22 | 1.05 | 1.22 | 1.08 | 1.26 | 1.26 | 1.26 | 1.26 | 1.07 | 1.05 | 1.05 | 1.05 |
| General happiness | 2.56 | 2.11 | 2.55 | 2.23 | 1.56 | 1.55 | 1.56 | 1.55 | 1.49 | 1.42 | 1.40 | 1.41 |
| Trust in others | 2.20 | 1.96 | 2.20 | 1.98 | 1.85 | 1.84 | 1.84 | 1.84 | 1.66 | 1.57 | 1.54 | 1.55 |
| Trust in Euro Parliament | 1.83 | 1.59 | 1.83 | 1.62 | 1.50 | 1.50 | 1.50 | 1.50 | 1.43 | 1.37 | 1.35 | 1.36 |
| Trust in legal system | 2.07 | 1.72 | 2.07 | 1.81 | 1.37 | 1.37 | 1.37 | 1.37 | 1.42 | 1.36 | 1.34 | 1.35 |
| Trust in police | 1.92 | 1.63 | 1.92 | 1.69 | 1.24 | 1.24 | 1.24 | 1.24 | 1.24 | 1.20 | 1.19 | 1.19 |
| Trust in politicians | 1.75 | 1.62 | 1.75 | 1.59 | 1.38 | 1.38 | 1.38 | 1.38 | 1.63 | 1.54 | 1.51 | 1.53 |
| Trust in parliament | 1.64 | 1.48 | 1.64 | 1.48 | 1.45 | 1.45 | 1.45 | 1.45 | 1.13 | 1.10 | 1.10 | 1.10 |
| Left-right scale | 1.70 | 1.65 | 1.70 | 1.58 | 1.48 | 1.47 | 1.48 | 1.48 | 1.31 | 1.26 | 1.25 | 1.25 |
| Satisfaction with life | 2.06 | 1.74 | 2.06 | 1.81 | 1.68 | 1.67 | 1.67 | 1.67 | 1.30 | 1.25 | 1.24 | 1.25 |
| Satisfaction with education system | 3.03 | 2.89 | 3.03 | 2.79 | 1.37 | 1.37 | 1.37 | 1.37 | 1.40 | 1.34 | 1.32 | 1.33 |
| Satisfaction with health system | 3.76 | 3.21 | 3.76 | 3.32 | 1.65 | 1.64 | 1.64 | 1.64 | 1.65 | 1.56 | 1.53 | 1.54 |
| Religiosity | 1.94 | 1.75 | 1.94 | 1.75 | 1.57 | 1.56 | 1.56 | 1.56 | 1.73 | 1.63 | 1.60 | 1.61 |
| Attitudes to immigrants | 2.77 | 2.68 | 2.77 | 2.57 | 1.92 | 1.92 | 1.92 | 1.92 | 1.89 | 1.76 | 1.73 | 1.74 |
| Supports law against ethnic discrimination | 2.82 | 2.85 | 2.82 | 2.66 | 1.73 | 1.72 | 1.72 | 1.72 | 2.57 | 2.36 | 2.29 | 2.33 |
| Importance of family | 2.17 | 1.99 | 2.17 | 1.97 | 1.19 | 1.19 | 1.19 | 1.19 | 1.21 | 1.17 | 1.17 | 1.17 |
| Importance of friends | 2.31 | 2.09 | 2.31 | 2.08 | 1.34 | 1.34 | 1.34 | 1.34 | 1.54 | 1.46 | 1.44 | 1.45 |
| Importance of work | 2.20 | 2.16 | 2.20 | 2.05 | 1.90 | 1.89 | 1.89 | 1.89 | 1.69 | 1.59 | 1.57 | 1.58 |
| Support people worse off | 2.70 | 2.47 | 2.70 | 2.45 | 1.35 | 1.35 | 1.35 | 1.35 | 1.78 | 1.67 | 1.64 | 1.66 |
| Always obey law | 2.43 | 2.21 | 2.43 | 2.20 | 1.53 | 1.52 | 1.52 | 1.52 | 2.11 | 1.96 | 1.91 | 1.93 |
| Political activism | 3.26 | 2.83 | 3.26 | 2.89 | 1.94 | 1.94 | 1.94 | 1.94 | 2.16 | 2.00 | 1.96 | 1.98 |
| Liberalism | 2.28 | 2.18 | 2.28 | 2.10 | 1.78 | 1.77 | 1.78 | 1.78 | 1.75 | 1.64 | 1.61 | 1.63 |
| Participation in groups | 3.75 | 3.04 | 3.75 | 3.24 | 2.26 | 2.25 | 2.25 | 2.25 | 1.82 | 1.71 | 1.68 | 1.69 |

We have shown in section 4 above that use of alternative, simpler, methods of combining the domain-specific *deff*s does not always result in good estimates. In particular, the use of a convex combination will tend to result in an under-estimation, the extent of which depends on the extent of departure from the assumptions underlying the simplified expressions. In our empirical illustration, departures were modest, but it is easy to imagine designs with greater variation between domains in mean selection probabilities or in the distribution of design weights. We would therefore recommend that estimators (6)–(9) are used only if the assumptions genuinely hold, or if the sample design data necessary to calculate (5) is not available, in which case the analyst should at least make arbitrary allowance for under-estimation based on his or her knowledge of the design.

An important issue that is outside the scope of this article is how to deal with non-response when predicting or estimating design effects for multiple design samples. The expressions throughout section 2 of this article refer to the number of observations, *i.e.*, respondent sample units, in each domain, $m_k$, and the calculations in sections 3 and 4 are based on predicted numbers of observations and actual numbers of observations respectively. But the natural interpretation of the differences between the four scenarios in section 2 may be in terms of sample design, where the weights are design weights. Thus, scenario 2, for example, would refer to a design that is *epsem* within domains, but where the sampling fraction is permitted to differ between domains. However, in most realistic applications non-response will occur and may well be differential both between and within domains. This is often reflected in an adjustment to the design weight. Thus, the simplification of scenario 2 would only apply if the non-response adjustment were constant within domains, in addition to the design being *epsem* within domains.

Scenario 3, if interpreted with respect to design alone, should always hold for any well-specified design in which the domains form explicit strata. Expression (8) is therefore equivalent to expression (5) in the absence of non-response. In the presence of non-response, scenario 3 requires that the (design-weighted) response rates are equal in each domain. Similarly, scenario 4 requires that the net inclusion rate (the product of coverage rate, sampling fraction and response rate) is equal in each domain, whereas a design interpretation would not consider the response rate component.

Appropriate ways to incorporate non-response adjustment into design effect estimation and, in particular, how that might effect estimation for multiple design samples, would appear to be an area worthy of further research.

## Acknowledgement

The third author is grateful to ZUMA for a guest professorship which provided the time and stimulation to write this paper and for the support of the UK Longitudinal Studies Centre at the University of Essex, which is funded by grant number H562255004 of the UK Economic and Social Research Council.

## References

Gabler, S., Häder, S. and Lahiri, P. (1999). A model based justification of Kish's formula for design effects for weighting and clustering. *Survey Methodology*, 25, 105-106.

Häder, S., Gabler, S., Laaksonen, S. and Lynn, P. (2003). The sample. Chapter 2 in *ESS 2002/2003: Technical Report*. http://www.europeansocialsurvey.com.

Lohr, S.L. (1999). *Sampling: Design and Analysis.* Pacific Grove: Duxbury Press.

Lynn, P., and Gabler, S. (2005). Approximations to $b^*$ in the prediction of design effects due to clustering. *Survey Methodology*, 31, 101-104.

Lynn, P., Gabler, S., Häder, S. and Laaksonen, S. (2007, forthcoming). Methods for achieving equivalence of samples in cross-national surveys. *Journal of Official Statistics,* accepted.

Park, I., and Lee, H. (2004). Design effects for the weighted mean and total estimators under complex survey sampling. *Survey Methodology*, 30, 183-193.

Rao, C.R., and Kleffe, J. (1988). *Estimation of Variance Components and Applications*. Amsterdam: North-Holland.