EuroCohort Work Package No. 8

07/03/2019

Deliverable 8.1
Sample Design: Criteria and Procedures

Peter Lynn, University of Essex, UK

Version

Deliverable Date	**December 2018**		
Submitted Date	**29th March 2019**		
Dissemination level	**Confidential**		
WP Leaders	**University of Essex**		
Document history			
Version	Date	Comments	**Modified by**
1.0	7th Mar 2019		**PL**
1.1	29th Mar 2019		**GP**
1.2	3rd Apr 2019		**PL**

Abstract

This report sets out the principles that will govern sample design for EuroCohort, the criteria that will be used for sampling each cohort, and the procedures that will be followed to develop and approve each sample design. Also included is a template sample design description form to be completed for each country for each cohort.

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under the Grant Agreement no 777449.

EuroCohort Sample Design: Criteria and Procedures

1. Introduction and Principles

For *Eurocohort* to provide a robust and trusted basis for policy-related analysis, it is essential that sample design, like other aspects of the survey methodology, should follow best practice and should be executed to the highest possible standards. We consider probability sampling to be the only approach that satisfies these criteria. While non-probability methods have received considerable attention in recent years[i], and have even been used for some policy-related research purposes, robust and widely-accepted inference methods for non-probability samples do not yet exist, nor is their arrival imminent. For that reason, non-probability sampling is not appropriate for scientific surveys such as *Eurocohort* that require population inference. Moreover, it will be important to avoid serious coverage errors. To that end, in each country and for each cohort, a sampling frame/method should be used that provides the best possible coverage of the target population. Finally, it will be important to ensure that the samples provide a basis for analysis and estimation of sufficient precision to serve *Eurocohort's* purposes. To that end, each sample design should provide a prescribed level of statistical precision. However, it is clear that available frames and procedures vary between countries and that other important potential influences on sample design, such as population density and the homogeneity of schools or localities, also vary. It will therefore be impossible for the sample design to be the same in each country. And indeed, this is not necessary. We espouse the general approach of Kish[ii]:

> *Sample designs may be chosen flexibly and there is no need for similarity of sample designs. Flexibility of choice is particularly advisable for multinational comparisons, because the sampling resources differ greatly between countries. All this flexibility assumes probability selection methods: known probabilities of selection for all population elements.*

We propose that *Eurocohort* should strive to use the best possible random sampling practice in each participating country. The *Eurocohort* sample designs should enable comparative analysis with useful and estimable precision. In summary, the key principles are:

- Sampling frames/methods that provides the best possible coverage of *Eurocohort* target populations, to avoid coverage errors;
- Probability-based sampling, to ensure a valid basis for statistical inference;
- Sample designs that provide a prescribed level of statistical precision, to ensure the value of the data;
- The most appropriate sample design for each country and each cohort, to recognise national variation while adhering to the objectives.

2. Criteria for Sampling a Cohort

2.1 Population Coverage

We suggest the following definitions of the target populations. For illustration, we assume that the first wave of fieldwork with Cohort 1 will take place starting in September 2022. All dates in what follows can be amended proportionally should a different start date transpire:

Infant cohort (C1): all persons born between 01-09-2013 and 31-08-2014 and resident within the country at age 8 years 3 months, regardless of nationality, citizenship or language.

Birth cohort (C2): all persons born between 01-09-2023 and 31-08-2024 and resident within the country at age 3 months, regardless of nationality, citizenship or language.

This includes children whose families are temporarily away from the country for less than 6 months (e.g. on holiday or working abroad). It excludes children whose families have been, or will be, away from the country for 6 months or more, and temporary visitors (staying in the country for less than 6 months). Note also that for the infant cohort, the target population is not restricted to children currently attending (mainstream) school. The sample design in each country should achieve complete, or near-complete, coverage of the target population.

2.2 Probability Sampling

The sample is to be selected by strict random probability methods at all stages. This means that every member of the *Eurocohort* target population in a country should have a larger than zero probability of being selected into the sample and that this probability should be known for each child actually selected. The probability of selection for each sampled child must be recorded and retained in a file of sample design data.

Quota sampling is not permitted in any part of the sampling procedure, nor is substitution of non-responding, non-contactable or non-accessible sample members.

2.3 Statistical Precision

A target minimum level of statistical precision will be prescribed for each initial national cohort sample (i.e. the sample participating at wave 1). This level will not vary between countries or cohorts, as this in turn guarantees a minimum level of precision for comparisons of countries. In practice, the statistical precision of any survey estimate is determined by several factors. Key ones are:

1) Sample size;
2) Distribution of selection probabilities, and their association with the survey variable(s) upon which the estimate is based;
3) Sample clustering, and the association of the clusters with the survey variables;
4) Sample stratification, and the association of the strata with the survey variables;
5) Population variance of the survey variables.

Once a wave of data collection is completed, precision can be estimated empirically, provided that indicators of selection probabilities, clusters and strata are available (and weighting variables, if weighting is applied). Precision can, and does, vary between different estimates based on the same sample.

But at the sample design stage, precision must be predicted based on some assumptions, and in a way that provides a single standardised prediction for a sample design (not a separate one for each possible estimate). We propose the use of simple heuristics to do this, similar in some respects to those used on other major cross-national surveys in Europe, such as ESS[iii] and SHARE[iv]. Specifically, we propose specification of a minimum effective sample size (neff) that should be achieved at wave 1. This is the size of a simple random sample (i.e. one in which factors (2), (3) and (4) do not apply) that would provide the same precision as the actual design under consideration.

Attrition rates will differ over a 25 year period for both cohorts and between countries. In setting the initial target sample sizes for each cohort it is necessary to anticipate likely levels of attrition in such a way as to ensure that the later data collection waves retain sufficient respondents for inferential statistical analysis. We propose that this wave 1 minimum effective sample size should be set at 10,000 for the birth cohort and 8,000 for the infant cohort, with a waiver for the smallest countries, for whom the effective sample size will be not smaller than 5,000 for the birth cohort and 4,000 for the infant cohort.

For sample planning purposes, the sample size that would be equivalent to the required effective sample size can be predicted as set out in Annex B. This prediction requires estimation of the distribution of selection probabilities in the sample, the mean number of respondents in each cluster if the sample is clustered, and the relative homogeneity of clusters. A further step of estimation will then be required to convert this target net sample size (number of participants) into an appropriate gross sample size (selected sample size). This step requires estimation of the sample eligibility rate and the wave 1 response rate.

2.4 Sample Design

It is likely that a population register would be an ideal sampling frame to achieve the desired coverage. Birth registrations or health service registrations may form an acceptable frame for the birth cohort, but for the infant cohort it is likely that a sample from birth registrations would have to be supplemented from other sources with a sample of recent immigrants. School rolls or equivalent could provide good coverage for the infant cohort if supplemented with a sample of children not in school, wherever this is a non-negligible proportion.

Sample design may vary between countries and cohorts, to recognise national variation while adhering to common principles and parameters. Some countries will prefer to use clustered samples, where the sampled children are clustered within either geographical areas identified on the sampling frame or schools. This will reduce unit costs of data collection, particularly for the first wave, when face-to-face interviewing will be dominant, but it will also tend to reduce the precision of estimates, so a larger sample size will be needed (than would be needed if the sample were not clustered) to meet the EuroCohort precision requirements.

Proportionate stratified sampling will be encouraged, as this will tend to increase the precision of sample-based estimates. It is likely that relevant variables for stratification will often be available on the sampling frame (for example, gender, year and month of birth, and in some cases achievement/ability indicators, or socio-demographic indicators relating to parents).

3. Proposed Procedures for Sample Design Development

Procedures for the development of sample designs on cross-national surveys appear to be of two sorts. Some surveys provide written rules, which may be in the form of a technical manual (PISA, PIAAC, ISSP) or legislation (EU-SILC, EU-LFS). Country teams then work largely in isolation to develop their sample designs following the documentation. Other surveys (SHARE, ESS) adopt a more hands-on approach. In addition to written guidelines, members of the central survey team are available to support and advise national teams in the development of their designs and each national design has to be approved by the central team before the survey can proceed[v] (Lynn et al 2007). It is generally perceived that the latter model improves comparability and compliance.

EuroCohort will adopt the latter model. A small group of sampling experts will be responsible for producing EuroCohort sampling guidelines and for working with each national team to develop and approve the sample design for each cohort. The EuroCohort sampling experts will work collaboratively with the national teams, providing advice, helping with the necessary calculations (particularly the prediction of design effect components, with which national teams cannot be assumed to be familiar), and commenting on proposed design features. Key features of each sample design will be documented in a standard form (see draft at Annex A), which will serve both as a record of the approval process and, ultimately, as technical documentation for data users.

4. Sampling Guidelines

4.1 Sampling Frames

The quality of the sample will be higher, the more completely the sampling frame covers the target population. The choice of sampling frame will also constrain the extent to which it is possible to control variation in selection probabilities, and hence the likely value of $deff_p$, which will influence the number of participants required. Thus, the choice of sampling frame is of great importance.

Sampling frames for EuroCohort are likely to include the following types:

1. Population registers. Many countries have national population registers. These registers invariably include the date of birth of each person and can therefore be used to identify, and sample, persons belonging to the age cohorts relevant to EuroCohort.
2. Registers of births. Most countries have a register of births. This may be a viable option as a sampling frame if there is no population register available. However, there is typically a variable time-lag between a birth and entry on the register, so this may introduce a constraint on the possible timing of sample selection and wave 1 fieldwork (the same may be true of some population registers).
3. Health service registers. In most countries, persons are registered for access to health services. Such registers, like population registers, invariably include date of birth and therefore could serve as sampling frames for EuroCohort.
4. School rolls. Where none of the above registers exist, or are accessible, or where it is deemed desirable to cluster the sample within schools and/or collect data in-school, school rolls may be appropriate sampling frames.
5. Education registers. Some countries have centralised registers of pupils in education. If access can be gained to such registers for sampling purposes, this may avoid the need to gain the co-operation of individual schools in sample selection, as the sample can be selected centrally.

Population registers and birth registers are both likely to provide good coverage of the target population. School rolls will exclude any children who are not in school (e.g. home-schooled or excluded. Education registers may, in some countries, additionally exclude children not attending state schools. A separate methods of sampling may be needed in order to include children attending private schools or other minor categories of schools that are not administered by the state.

All types of registers permit the use of equal-probability sampling, leading to $deff_p$ = 1.0, which minimises the number of interviews needed in order to meet the effective sample size requirement. This is true regardless of whether or not the sample is clustered. School rolls, on the other hand, rarely provide equal-probability sampling. This is because of an inevitable delay between the production of roll data used to select schools at the first stage and the selection of pupils within schools at the second stage. During this period, school rolls may increase or decrease in size. In some countries, national data on school rolls often refers to the previous year rather than the current one, and in some cases such data are available only for school years rather than birth years. Nevertheless, differences between size measures used to assign probabilities to schools and those encountered at the second stage of sampling are likely to be modest and therefore so are any consequent precision losses.

If data collection can be organised in-school this is likely to result in relatively high participation rates. This could be an argument for using school rolls or education registers (if they indicate current school) as a sampling frame, in order to enable sample clustering within schools. However, given the likely need for wave 1 data collection to be in-home, this may not be a relevant consideration. Countries wishing to use a geographically-clustered sample would likely be able to do this with any of the sampling frame types mentioned here. However, geographical indicators are less likely to be up-to-date with birth registers than with other types of registers, as these generally indicate either at the time of the birth or on the day on which the birth was registered.

Access to potential sampling frames for EuroCohort purposes may need to be negotiated. Time and resources should be allowed for this.

If none of the types of sampling frames mentioned above are available, countries may resort to screening a general population sample of households to locate children within the target age group. However, this is likely to be a very costly way of reaching a sample of appropriate size as the number of addresses to be screened will need to be many times larger than the target sample size of children (possibly up to 100 times larger). It is therefore expected that this method, though acceptable in principle, will not be feasible. If screening methods are used, the addresses/ households to screen must be selected with strict probability methods. Random walk methods are unacceptable as these do not strictly control selection probabilities. Area sampling, followed either by screening of all addresses in each selected area or by probability sampling from a list constructed by complete enumeration of addresses in the selected area, could be acceptable.

Sampling Frames: Key Points

Population registers, birth registers, health service registers, education registers and school rolls are all potential options, though not all options are available in all countries;

Population, birth registers, and health service registers will tend to have good coverage;

Any of the potential frames is likely to include information that could be used to provide effective geographical clustering of the sample;

School rolls are likely to have an advantage only if data collection can be carried out in-school (which is unlikely to be the case if face-to-face interviewing is required).

4.2 Multi-stage Sampling

Multi-stage sample designs are used either to increase the cost-efficiency of the design (as they result in a sample which is *clustered*, usually within relatively small geographical areas, such that each sample cluster forms the workload for one interviewer) or because the constraints on available sampling frames leave no choice (for example, if school rolls were the only available sampling frame). Examples include:

2-stage. First stage small geographical areas; second stage persons

3-stage. First stage geographical areas; second stage smaller geographical areas; third stage persons

2-stage. First stage schools; second stage persons

3-stage. First stage schools; second stage classes; third stage persons

It should be noted that cost savings in data collection due to clustered designs generally accrue only with face-to-face interviewing. Furthermore, in longitudinal surveys the savings tend to reduce over time as some sample members move home (the sample "de-clusters"). As EuroCohort will be a mixed-mode longitudinal study, potential cost savings from a clustered design may be limited. Longer-term costs should be considered and balanced against the precision loss associated with clustered designs (see below).

Key aspects of multi-stage designs are the following:

- The overall selection probability of each person is the product of the conditional selection probabilities at each stage of the sample design. Careful control of the relationship between these probabilities is therefore important;
- The predicted design effect due to clustering ($deff_c$) depends on two features of the sample design: the relative homogeneity of the first-stage units (*primary sampling units*, *PSUs*), and the number of interviews conducted in each.

With respect to the control of probabilities, an efficient design is one in which PSUs are selected with probability proportional to the number of target persons in the PSU (children in the relevant age cohort), and subsequently the same number of persons is selected in each sampled PSU. If there are good reasons for wanting to vary the sample size of persons per PSU between two or more strata (for example, a smaller sample size per PSU in rural areas than in urban areas or in smaller schools than in larger schools), then the first-stage probabilities should be modified to compensate (for example, a larger probability for PSUs in rural areas or for smaller schools).

Regarding $deff_c$, the following points should be noted:

Relatively heterogeneous PSUs are desirable (greater precision and therefore fewer interviews required). Typically, larger geographical areas are more heterogeneous than smaller areas. Also, schools may be more heterogenous than classes (this is particularly true if classes are formed based on ability, but is to a lesser extent always true due to the shared experiences of pupils within the same class). Thus, if possible, larger rather than smaller areas should be used, and schools rather than classes (i.e. sample pupils across all classes in each school, rather than sampling one or two classes). Even an increase from a mean PSU size of, say, 100 children in the target age range to 200 is likely to be worthwhile, so it is worth considering whether smaller geographical units could be combined to create larger units prior to sample selection;

Smaller *sample* size per PSU is desirable. Thus, to the extent possible, the number of sampled PSUs should be maximised and the number of sampled persons per PSU minimised.

Multistage Sampling: Key Points

Larger geographical areas are preferred to smaller areas as PSUs; Schools are preferred to classes;

A larger number of sampled PSUs is preferred to a smaller number;

Large variability in the size of PSUs (within strata) is undesirable;

If possible, PSUs should be sampled with probability proportional to the number of persons in the PSU, and a fixed number of persons then selected in each PSU.

4.3 Stratification

Proportionate stratified sampling can improve the precision of sample estimates. If, for example, strata are regions, this ensures that the sample distribution by region matches the population distribution: there is no random sampling variation in respect of region.

Stratification can be either explicit or implicit. With *explicit stratification*, the units on the sampling frame are sorted into distinct strata and a sample is selected independently from each stratum. With *implicit stratification* the units on the sampling frame are sorted (ranked) in a meaningful order and a systematic sample (every n^{th}) is then selected from the sorted list. Either method is effective at improving precision. More important is the choice of variables to define the stratification.

Stratification is more effective the more strongly associated the stratification variables are with the survey variables (i.e. the EuroCohort questionnaire variables). Individual-level variables therefore tend to be more beneficial than regions or characteristics of small areas (e.g. population density or local unemployment rate) or characteristics of schools, other things being equal. However, other things are not always equal. It may be that the individual-level variables available from a register, such as birth month and gender, are less strongly predictive of survey measures than characteristics of schools such as examination pass rates or the proportion of children from families supported by social welfare. This will depend on the available information and the local circumstances (e.g. the extent to which examination pass rates vary between schools).

When using any kind of register, it is likely that useful stratification can be achieved using the information on the register. This remains true whether the design is a one-stage sample of persons or a multi-stage sample where the units selected at the first stage are geographical areas. The latter could, for example, be stratified by aggregate characteristics of the persons in the area. When selecting schools as PSUs, characteristics of the schools can be used to provide proportionate stratification.

Stratification: Key Points

Proportionate stratification is beneficial and is preferred to simple random sampling;

Stratification can be either explicit or implicit;

Choose stratification variables that are correlated with the survey variables;

With multi-stage sampling, PSUs can usually be stratified by geography or by geographically-defined variables;

Persons within PSUs should be selected by systematic random sampling with implicit stratification, in preference to simple random sampling.

4.4 Predicting $deff_p$

To predict $deff_p$ requires a prediction of the distribution of (relative) overall probabilities of selection for survey respondents. Most EuroChort sample designs are likely to be equal-probability designs, in which case $deff_p = 1$:

- Single-stage equal-probability sample of babies/children;
- PSUs selected with probability proportional to size; fixed number of babies/children selected from each PSU.

Possible sources of variation in selection probabilities include:

- Oversampling regions or subpopulations of particular analytical interest. For example, if 20% of the population lives in Region A and are given twice the selection probability of persons in Region B, then persons in Region A will have a relative weight of 0.5. And persons in Region A will constitute one-third of the sample (because (20%*2P)/((20%*2P)+(80%*P)) = 1/3, where P is the probability of selection in Region B). Thus:

$$deff_p = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2}$$

$$= n\frac{(0.333n\times0.5^2)+(0.667n\times1^2)}{\left((0.333n\times0.5)+(0.667n\times1)\right)^2} \quad = \frac{0.750}{0.833^2} \quad = 1.08$$

- Selecting schools with probability proportional to an approximate size measure, which does not correspond perfectly with the current size measure, which is identified only at the next sampling stage. The design effect will be modest if the two size measures are highly correlated (perhaps in the range 1.01 to 1.05), but will be larger the lower the correlation.

4.5 Predicting *deff*$_c$

For single-stage, unclustered, samples, $deff_c = 1$. However, for multi-stage (clustered) samples, it is necessary to predict the design effect due to clustering. To do this, we need predictions of both the mean number of interviews per PSU, $\bar{b}$, and the relative homogeneity of persons living within the PSU, ρ. The predicted value of $\bar{b}$ is simply the ratio of the total number of achieved interviews to the number of sample PSUs. However, the required number of achieved interviews is determined by the prediction of *deff*, so the problem is circular and must be solved iteratively.

The intra-cluster correlation coefficient, ρ, will in practice vary between survey variables and estimates. However, to determine the required sample size only one value can be used. Based on general impressions of ρ from other surveys, we suggest using a value of 0.02 when PSUs are defined by geography, 0.03 when PSUs are schools and 0.04 if PSUs are classes (not recommended).

The following table shows how predictions of $deff_c$ depend on the predicted values of $\bar{b}$ and ρ:

$deff_c$	$\bar{b} = 5$	$\bar{b} = 10$	$\bar{b} = 15$	$\bar{b} = 20$	$\bar{b} = 25$
$\rho = 0.02$	1.08	1.18	1.28	1.38	1.48
$\rho = 0.03$	1.12	1.29	1.42	1.57	1.72
$\rho = 0.04$	1.16	1.36	1.56	1.76	1.96

It can be seen that the design effect increases quite rapidly as both $\bar{b}$ and ρ increase; *deff*$_c$ can become considerable if $\bar{b}$ exceeds 12.

4.6 Estimating the Required Number of Interviews

Suppose that PSUs are geographical areas ($\rho = 0.02$) and that the proposed sample design has $\bar{b} = 20$. Then, the predicted value of *deff*$_c$ (from the table above) is 1.38. If the design is equal-probability, then *deff*$_p$ = 1.00, so the overall *deff* can be estimated as *deff* = 1.38 x 1.00 = 1.38. Then, the minimum required number of interviews at wave 1 would be *n* = 10,000 x *deff* = 13,800. But this number could be reduced if we change the sample design to have smaller sample sizes per cluster. Reducing $\bar{b}$ to 10 would reduce *deff*$_c$ to 1.18 and hence the minimum number of interviews would reduce to 11,800. Achieving this reduction in $\bar{b}$ would involve increasing the number of sample PSUs from 690 to 1,180. These two designs provide equivalent precision. The choice between these designs and others of equivalent precision should depend on the associated field costs. The preferred sample design should be the one that maximises precision for a fixed budget or minimises the budget required to deliver a fixed precision. However, it should be noted that the cost reduction due to sample clustering will be much greater at wave 1 than at subsequent waves, when much data collection will be online. The longer-term costs should therefore be considered.

4.7 Calculating the Required Sample Size

The steps in calculating the minimum required gross (initial) sample size are:

i. Predict $deff_p$;
ii. Predict $deff_c$ and hence $deff = deff_p \times deff_c$;
iii. Calculate the minimum required number of interviews, m: $m = n_{eff} \times deff$
iv. Predict the response rate, *rr*, and the ineligibility rate, *ri*. The predicted response rate should be realistic but should be the highest achievable based on good survey practice. The ineligibility rate, *ri*, indicates the proportion of selected babies/children that will turn out to be ineligible for the survey (for example, persons who have died or moved abroad). This is expected to be close to 0 in most cases.
v. Calculate the minimum required gross (initial) sample size, n: $n \geq m/\left(rr \times (1 - ri)\right)$.

An example of this calculation:

i. An equal-probability design, so, $deff_p = 1.00$.
ii. Geographical areas as PSUs ($\rho = 0.02$) and $\bar{b} = 20$, so the predicted value of $deff_c$ (as in section 4.6 above) is 1.38.
iii. $m = 10{,}000 \times 1.38 = 13{,}800$.
iv. Response rate is predicted to be 65%. 0.05% of sampled persons are expected to have died or moved abroad by the time of wave 1 fieldwork.
v. $n \geq 13{,}800/(0.65 \times 0.9995) = 21{,}241$

Annex A: Draft National Sample Design Description Form

EuroCohort Sample Design Description

Country:		<country>
National Co-ordinator:		<name and email address>
Survey Institute:		<institute name>
Sampling Expert:		<name> (<email address)
Date:		*<date>*
Status:	☐	In development
	☐	Approved
	☐	Final (post-field work)

1.1 Target Population

Number of residents in the age cohort:	<number>
Source and reference date:	<details>

1.2 Population Coverage

<Describe here any population subgroups not covered by the sample design. Include an estimate of the proportion of the total population that each subgroup accounts for>

2. Summary of the Sample Design

<Provide an overview of the sample design in one or two paragraphs. Outline the sampling frame, the source of any other data used in the design, the stratification to be used, and the clustering to be used (number and nature of primary sampling units), if any.>

3. Sample Design Details

First Sampling Stage

Sample units: *< State the units to be selected, e.g. municipalities, postal sectors, schools, addresses, persons, etc >*

Sampling frame: *< Describe the sampling frame of these first-stage units >*

Sample size: *< Number of units to be selected >*

Stratification:

☐ Explicit

☐ Implicit

☐ None

Strata: *< If explicit or implicit stratification used, describe how the units are stratified prior to selection. If the stratification is explicit, state how many strata there are and how they are defined >*

Allocation to strata: *< If the stratification is explicit, describe how the number of units to select from each stratum is determined (if applicable)>*

Selection algorithm: *< Describe how it is determined which units to select (in each stratum). For example, simple random sampling, systematic sampling; with equal probabilities or with probability proportional to size; etc >*

Second Sampling Stage

Sample units: *<State the units to be selected, e.g. classes, persons, etc>*

Sampling frame: *<Describe the sampling frame of these second-stage units>*

Sample size: *<Number of units to be selected>*

Stratification:

☐ Explicit

☐ Implicit

☐ None

Strata: *<If explicit or implicit stratification used, describe how the units are stratified prior to selection. If the stratification is explicit, state how many strata there are and how they are defined>*

Allocation to strata:	*< If the stratification is explicit, describe how the number of units to select from each stratum is determined (if applicable)>*
Selection algorithm:	*<Describe how it is determined which units to select (in each stratum). For example, simple random sampling, systematic sampling; with equal probabilities or with probability proportional to size; etc>*

4. Planning the Sample Size

Parameters of the Planned Gross Sample Size

Achieved interviews per cluster $(\bar{b})$	Intraclass Correlation Coefficient (ρ)	Design Effect due to Selection Probabilities (Deff_p)	Response Rate (rr)	Ineligible Rate (ri)	Effective Sample Size (n_{eff})
< >	< >	< >	< >	< >	< >

Design Effect

Deff_c	=	$1 + (\bar{b} - 1) \times \rho$	=	< >
Deff_p	=			< >
Deff	=	$\mathrm{Deff}_p \times \mathrm{Deff}_c$	=	< >

Gross Sample Size

Min. n_{net}	=	$\mathrm{Deff} \times n_{eff}$	=	< >
Target n_{net}	=			< >
n_{gross}	=	$\dfrac{n_{net}}{\mathrm{rr} \times (1 - \mathrm{ri})}$	=	< >

Remarks

<Any further comments or explanations about the sample design, including assumptions about ineligibility rates and response rates>

5. Sample Design Data File (SDDF)

< The list of variables is to be edited, to provide details as appropriate >

Variable	**Description**
idno	Personal identifier
p1	Probability of selection at first stage of sampling
p2	Conditional probability of selection at second stage of sampling
p3	Conditional probability of selection at third stage of sampling
stratex1	Indicator of explicit stratum at first stage of sampling
stratim1	Order of selection of PSU
stratim2	Order of selection of person within PSU
strtval1	Value of the first variable used to stratify PSUs
strtval2	Value of the second variable used to stratify PSUs
psu	PSU identifier
outcome	Final outcome
frame1	Auxiliary variable from sampling frame: < >
frame2	Auxiliary variable from sampling frame: < >
frame3	Auxiliary variable from sampling frame: < >

Probabilities of Selection

< Define the values of the probability indicators that will be included in the SDDF. For example, for probability proportional to size selection of schools as PSUs, using school roll in the relevant age group count as the size measure, we might have $P1_i = n1\frac{N_i}{N}$, where $n1$ is the number of schools to be sampled, N_i is the school roll for the i^{th} school and N is the total school roll for all schools on the frame; etc. >

Annex B: Prediction of Effective Sample Size

The effective sample size , n_{eff}, can be predicted for any proposed sample design as follows:

$$\widehat{n_{eff}} = {}^{n}/_{\widehat{deff}}$$

where $\widehat{deff} = \widehat{deff}_P \times \widehat{deff}_C$ is the predicted design effect;

$\widehat{deff}_p = \frac{n \sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2}$ is the design effect due to variation in selection probabilities, and

$\widehat{deff}_c = 1 + \left(\bar{b} - 1\right)\rho$ is the design effect due to sample clustering;

w_i is the design weight for sample element i, which is the reciprocal of the selection probability;

$\bar{b}$ is the mean number of responding sample members per cluster;

ρ indicates the relatively homogeneity of clusters (intra-cluster correlation coefficient).

Note that the clusters are likely to be either schools or administrative geographical areas in the case of C1, and geographical areas for C2. In some countries, an unclustered sample may be feasible, in which case $\widehat{deff}_c = 1$.

Endnotes

---

[i] Research and discussion of inference from non-probability samples has appeared in special issues of the *International Journal of Epidemiology* (volume 42, 2013) and *Longitudinal and Life Course Studies* (volume 6, 2015). This will also be the topic of a forthcoming special issue of *Survey Research Methods*, based on papers presented at a conference in Paris in March 2017 organised by ESRA and ELIPSS. A workshop on the topic was also held in Washington DC in September 2017, organised by the National Institute of Statistical Sciences. A large part of the *Public Opinion Quarterly* 2017 special issue on the future of surveys was devoted to a symposium on probability and nonprobability surveys, consisting of two major papers and nine short commentary articles. Interest in probability sampling has even spread recently to other fields of science: an influential paper published in *Nature Communications* in October 2017 demonstrated that neuroimaging research can be misleading when based on non-representative samples. This provoked considerable debate about the use of probability sampling of populations for research in human biology.

[ii] Kish, L. (1994), "Multipopulation survey designs: five types with seven shared aspects," *International Statistical Review* 62, 167-186.

[iii] https://www.europeansocialsurvey.org/methodology/ess_methodology/sampling.html

[iv] Klevmarken, A.N., B. Swensson, and P. Hesselius (2005) The SHARE Sampling Procedures and Calibrated Design Weights. In *The Survey of Health, Ageing and Retirement in Europe – Methodology,* eds. A. Börsch-Supan, and H. Jürges. Mannheim: MEA.

[v] Lynn, P., Gabler, S., Häder, S. & Laaksonen, S. (2007) "Methods for achieving equivalence of samples in cross-national surveys". *Journal of Official Statistics*, 23(1): 107-124